An introduction to Psychometric Theory
Theory of Data, Issues in Scaling

William Revelle

Department of Psychology
Northwestern University
Evanston, Illinois USA

NORTHWESTERN
UNIVERSITY

April, 2025

Outline

Data = Model + Residual

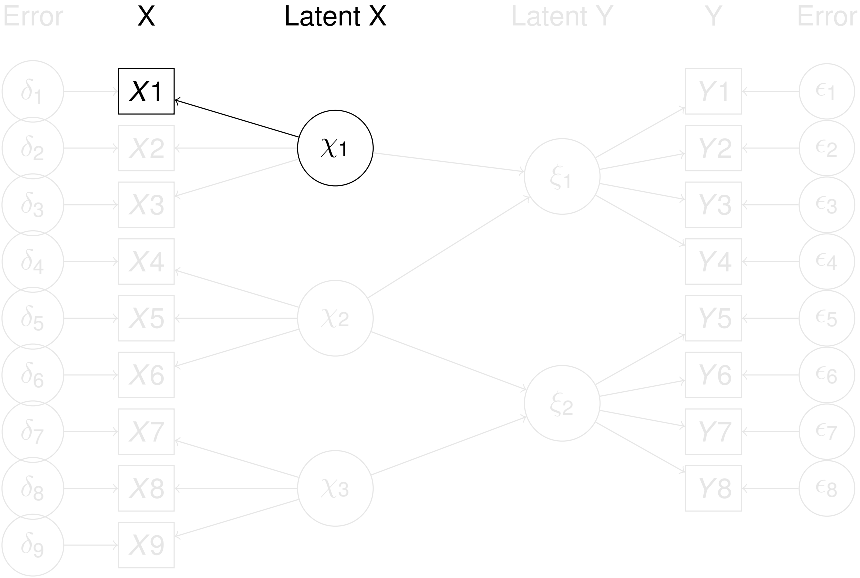
- The fundamental equations of statistics are that
  - Data = Model + Residual
  - Residual = Data - Model
- The problem is to specify the model and then evaluate the fit of the model to the data as compared to other models
  - Fit = f(Data, Residual)
  - Typically: $Fit = f(1 - \frac{Residual^2}{Data^2})$
  - $Fit = f(1 - \frac{(Data-Model)^2}{Data^2})$
- Even for something as simple as the mean is a model of the data. The residual left over after we remove the mean is the variance.
- This is a course in developing, evaluating, and comparing models of data.

Psychometrics as model estimation and model fitting

We will explore a number of models

1. Modeling the process of data collection and of scaling
   - $X = f(\theta)$
   - How to measure X, properties of the function f.
2. Correlation and Regression
   - $Y = \beta X$
   - $R_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$
3. Factor Analysis and Principal Components Analysis
   - $R = FF' + U^2 \quad R = CC'$
4. Reliability $\rho_{xx} = \frac{\sigma^2_\theta}{\sigma^2_X}$
5. Item Response Theory
   - $p(X|\theta, \delta) = f(\theta - \delta)$
6. Structural Equation Modeling
   - $\rho_{yy} Y = \beta \rho_{xx} X$

A theory of data and fundamentals of scaling



Consider the following numbers, what do they represent?

Table: Numbers without context are meaningless. What do these number represent? Which of these numbers represent the same thing?

2.718281828459045 08	3.14159265358979316
24	86,400
37	98.7
365.25	365.25636305
31,557,600	31,558,150
3,412.1416	.4046856422
299,792,458	6.022141 * 10^{23}
42	X

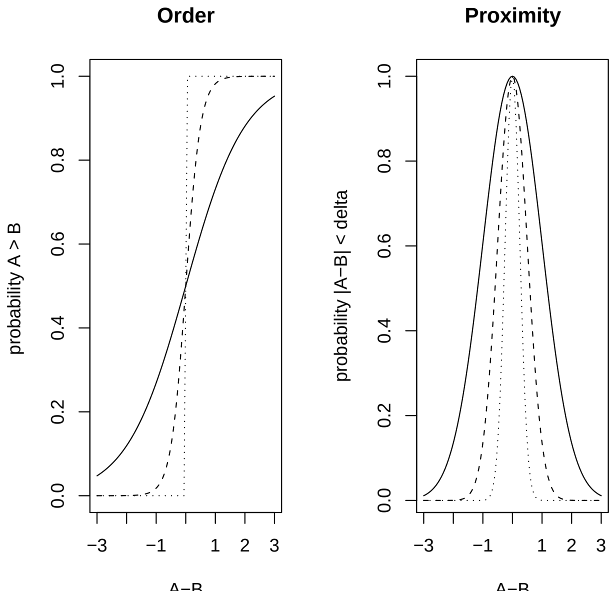
See also Thorndike (1904) for an amazing introduction to the problem of numbers and measurement.

Clyde Coombs and the Theory of Data

1. O = the set of objects
   - $O = \{o_i, o_j \dots o_n\}$
2. S = the set of Individuals
   - $S = \{s_i, s_j \dots s_n\}$
3. Two comparison operations
   - order ($x > y$)
   - proximity ($|x - y| < \epsilon$
4. Two types of comparisons
   - Single dyads
     - (s_i, s_j) (s_i, o_j) (o_i, o_j)
   - Pairs of dyads
     - $(s_i, s_j)(s_k, s_l)$ $(s_i, o_j)(s_k, o_l)$ $(o_i, o_j)(o_k, o_l)$

Coombs (1964)

2 types of comparisons: Monotone ordering and single peak proximity



Theory of Data and types of measures

Table: The theory of data provides a 3 x 2 x 2 taxonomy for various types of measures

Elements of Dyad	Number of Dyads	Comparison	Name
People x People	1	Order	Tournament rankings
People x People	1	Proximity	Social Networks
Objects x Objects	1	Order	Scaling
Objects x Objects	1	Proximity	Similarities
People x Objects	1	Order	Ability Measurement
People x Objects	1	Proximity	Attitude Measurement
People x People	2	Order	Tournament rankings
People x People	2	Proximity	Social Networks
Objects x Objects	2	Order	Scaling
Objects x Objects	2	Proximity	Multidimensional scaling
People x Objects	2	Order	Ability Comparisons
People x Objects	2	Proximity	Preferential Choice
People x Objects x Objects	2	Proximity	Individual Differences in Multidimensional Scaling

Tournaments to order people (or teams)

1. Goal is to order the players by outcome to predict future outcomes
2. Complete Round Robin comparisons
   - Everyone plays everyone
   - Requires $N * (N-1)/2$ matches
   - How do you scale the results?
3. Partial Tournaments – Seeding and group play
   - World Cup
   - NCAA basketball
   - Is the winner really the best?
   - Can you predict other matches

Simulating a hypothetical chess game

```
> set.seed(42)
```

1. Set the random seed to get the same results

```
> p <- seq(-1.5, 1.5, 0.2)
> n <- length(p)
```

2. Generate a sequence of latent values

```
> pdif <- -p %+% t( p )
```

3. Find the matrix sum of a column vector and row vector

```
> prob <- 1/(1 + exp(pdif))
```

4. Convert to probabilities (using a logit model)

```
> match <- matrix(rbinom(n*n,1,prob),n,n)
> tournament <- t(upper.tri(match) * (1-match))
          + upper.tri(match)*match
> colnames (tournament) <- rownames(match) <- paste("P",1:n,sep="")
```

5. Convert probabilities to outcomes

```
> diag(tournament) <- NA
> tournament
```

6. Show the results

A hypothetical chess tournament

Table: Simulated wins and losses for 16 chess players. Entries reflect row beating column. Thus, P1 wins 4 matches, while P 16 wins 14 matches.

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16
P1	NA	1	0	1	1	0	0	0	0	0	1	0	0	0	0	0
P2	0	NA	1	1	0	0	0	0	0	0	0	0	0	0	0	0
P3	1	0	NA	0	0	0	1	0	1	0	0	0	0	0	0	0
P4	0	0	1	NA	1	1	0	0	0	1	0	0	0	0	0	0
P5	0	1	1	0	NA	1	1	0	0	0	0	1	0	0	0	0
P6	1	1	1	0	0	NA	0	0	1	0	0	1	0	0	0	0
P7	1	1	0	1	0	1	NA	1	1	1	0	0	0	0	0	0
P8	1	1	1	1	1	1	0	NA	1	0	0	0	1	1	0	0
P9	1	1	0	1	1	0	0	0	NA	1	0	1	1	0	0	0
P10	1	1	1	0	1	1	0	1	0	NA	0	1	0	0	0	1
P11	0	1	1	1	1	1	1	1	1	1	NA	1	1	0	1	0
P12	1	1	1	1	0	0	1	1	0	0	0	NA	0	1	1	0
P13	1	1	1	1	1	1	1	0	0	1	0	1	NA	0	0	0
P14	1	1	1	1	1	1	1	0	1	1	1	0	1	NA	1	0
P15	1	1	1	1	1	1	1	1	1	1	0	0	1	0	NA	0
P16	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	NA

The problem: How to scale the players

1. We want to assign numeric values to each player
2. What is best way to map from the values to the data?
3. How well do these values recreate the data?
4. Although players ranks can vary infinitely, pairwise competitions always are between 0 and 1
5. What kind of ranking can we use, what kind of choice model?

Multiple ways of ordering the results

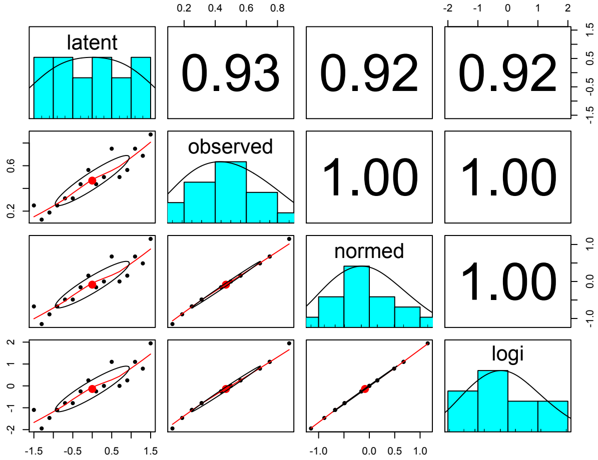
```
> score <- rowMeans(tournament, na.rm = TRUE)
> qscore <- qnorm(score) #convert means to normal deviates
> logi <- logit(score)    #convert means to logit units
> chess.df <- data.frame(latent = p, observed = score,
                         normed = qscore, logi)

> chess.df  #show the data

> pairs.panels(chess.df) #plot the results in a SPLOM
```

	latent	observed	normed	logi
P1	-1.5	0.2666667	-0.62292572	-1.0116009
P2	-1.3	0.1333333	-1.11077162	-1.8718022
P3	-1.1	0.2000000	-0.84162123	-1.3862944
P4	-0.9	0.2666667	-0.62292572	-1.0116009
P5	-0.7	0.3333333	-0.43072730	-0.6931472
P6	-0.5	0.3333333	-0.43072730	-0.6931472
P7	-0.3	0.4666667	-0.08365173	-0.1335314
P8	-0.1	0.6000000	0.25334710	0.4054651
P9	0.1	0.4666667	-0.08365173	-0.1335314
P10	0.3	0.5333333	0.08365173	0.1335314
P11	0.5	0.8000000	0.84162123	1.3862944
P12	0.7	0.5333333	0.08365173	0.1335314
P13	0.9	0.6000000	0.25334710	0.4054651
P14	1.1	0.8000000	0.84162123	1.3862944
P15	1.3	0.7333333	0.62292572	1.0116009
P16	1.5	0.9333333	1.50108595	2.6390573

1. Find the mean for each row
2. Express these as normal deviates
3. Express means as logit units
4. Organize all three and the original latent score into a data frame
5. Show the results
6. Graph the results

Note: these numbers have been revised because they were not correcting for the NAs

All three methods match the latent pretty well



So why bother with normal or logitistic modeling, why not just use total score?

1. How to predict wins and losses from the prior scores
   - What is the likelihood that player P16 will beat player 1 if they play again? We need some mapping function from scale to model of the data
2. $P(A > B) = f(A - B)$ But what is the function?
   - Must map unlimited A and B into 0-1 space
3. Several classic rules
   - Bradly - Terry - Luce Choice rule

$$p(A > B|A, B) = \frac{p(A)}{p(A) + p(B)}. \tag{1}$$

   - Thurston Normal deviation model

$$p(A > B|A, B) = pnorm(z_A - z_B) \tag{2}$$

   - Elo/Rasch logistic model where $logit_A = log(p_A/(1 - p_A))$

$$p(A > B|A, B) = \frac{1}{1 + e^{(logit_B - logit_A)}} \tag{3}$$

(Bradley and Terry, 1952; Luce, 1959, 1977; Joe, 1991; Elo, 1978; Elo, 1961)

How well do these various models fit the data?

1. Generate the model of wins and losses
   - Compare to the data
   - Find the residuals
   - Summarize these results
2. Can do it "by hand"
   - Take the scale model, model the data
   - Find residuals
   - Find a goodness of fit
3. Can use a *psych* function: `scaling.fits` to find the fit
   - Although we don't need to know how the function works, it is possible to find out by using just the function name
   - To find out how to call a function, ?function, e.g., ?scaling.fits
   - To run a function, just say function() e.g. scaling.fits(model, data)

Bradly - Terry - Luce model based upon scores

```
>round(score,4)
    P1     P2     P3     P4     P5     P6     P7     P8     P9    P10    P11    P12    P13
P14    P15    P16
0.2667 0.1333 0.2000 0.2667 0.3333 0.3333 0.4667 0.6000 0.4667 0.5333 0.8000 0.5333 0.6000 0.8000
> btl <- score/(score %+% t(score))
```

```
round(btl,2)
```

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16
P1	0.50	0.67	0.57	0.50	0.44	0.44	0.36	0.31	0.36	0.33	0.25	0.33	0.31	0.25	0.27	0.22
P2	0.33	0.50	0.40	0.33	0.29	0.29	0.22	0.18	0.22	0.20	0.14	0.20	0.18	0.14	0.15	0.12
P3	0.43	0.60	0.50	0.43	0.38	0.38	0.30	0.25	0.30	0.27	0.20	0.27	0.25	0.20	0.21	0.18
P4	0.50	0.67	0.57	0.50	0.44	0.44	0.36	0.31	0.36	0.33	0.25	0.33	0.31	0.25	0.27	0.22
P5	0.56	0.71	0.62	0.56	0.50	0.50	0.42	0.36	0.42	0.38	0.29	0.38	0.36	0.29	0.31	0.26
P6	0.56	0.71	0.62	0.56	0.50	0.50	0.42	0.36	0.42	0.38	0.29	0.38	0.36	0.29	0.31	0.26
P7	0.64	0.78	0.70	0.64	0.58	0.58	0.50	0.44	0.50	0.47	0.37	0.47	0.44	0.37	0.39	0.33
P8	0.69	0.82	0.75	0.69	0.64	0.64	0.56	0.50	0.56	0.53	0.43	0.53	0.50	0.43	0.45	0.39
P9	0.64	0.78	0.70	0.64	0.58	0.58	0.50	0.44	0.50	0.47	0.37	0.47	0.44	0.37	0.39	0.33
P10	0.67	0.80	0.73	0.67	0.62	0.62	0.53	0.47	0.53	0.50	0.40	0.50	0.47	0.40	0.42	0.36
P11	0.75	0.86	0.80	0.75	0.71	0.71	0.63	0.57	0.63	0.60	0.50	0.60	0.57	0.50	0.52	0.46
P12	0.67	0.80	0.73	0.67	0.62	0.62	0.53	0.47	0.53	0.50	0.40	0.50	0.47	0.40	0.42	0.36
P13	0.69	0.82	0.75	0.69	0.64	0.64	0.56	0.50	0.56	0.53	0.43	0.53	0.50	0.43	0.45	0.39
P14	0.75	0.86	0.80	0.75	0.71	0.71	0.63	0.57	0.63	0.60	0.50	0.60	0.57	0.50	0.52	0.46
P15	0.73	0.85	0.79	0.73	0.69	0.69	0.61	0.55	0.61	0.58	0.48	0.58	0.55	0.48	0.50	0.44
P16	0.78	0.88	0.82	0.78	0.74	0.74	0.67	0.61	0.67	0.64	0.54	0.64	0.61	0.54	0.56	0.50

(Bradley and Terry, 1952; Luce, 1959, 1977)

$$p(A > B|A, B) = \frac{p(A)}{p(A) + p(B)}. \tag{4}$$

BTL Residuals are data - model

```
> resid <- tournament - btl

> round(resid,2)
       P1    P2    P3    P4    P5    P6    P7    P8    P9   P10   P11   P12   P13   P14
P15   P16
P1     NA  0.33 -0.57  0.50  0.56 -0.44 -0.36 -0.31 -0.36 -0.33  0.75 -0.33 -0.31 -0.25 -0.27 -0.2
P2  -0.33    NA  0.60  0.67 -0.29 -0.29 -0.22 -0.18 -0.22 -0.20 -0.14 -0.20 -0.18 -0.14 -0.15 -0.1
P3   0.57 -0.60    NA -0.43 -0.38 -0.38  0.70 -0.25  0.70 -0.27 -0.20 -0.27 -0.25 -0.20 -0.21 -0.1
P4  -0.50 -0.67  0.43    NA  0.56  0.56 -0.36 -0.31 -0.36  0.67 -0.25 -0.33 -0.31 -0.25 -0.27 -0.2
P5  -0.56  0.29  0.38 -0.56    NA  0.50  0.58 -0.36 -0.42 -0.38 -0.29  0.62 -0.36 -0.29 -0.31 -0.2
P6   0.44  0.29  0.38 -0.56 -0.50    NA -0.42 -0.36  0.58 -0.38 -0.29  0.62 -0.36 -0.29 -0.31 -0.2
P7   0.36  0.22 -0.70  0.36 -0.58  0.42    NA  0.56  0.50  0.53 -0.37 -0.47 -0.44 -0.37 -0.39 -0.3
P8   0.31  0.18  0.25  0.31  0.36  0.36 -0.56    NA  0.44 -0.53 -0.43 -0.53  0.50  0.57 -0.45 -0.3
P9   0.36  0.22 -0.70  0.36  0.42 -0.58 -0.50 -0.44    NA  0.53 -0.37  0.53  0.56 -0.37 -0.39 -0.3
P10  0.33  0.20  0.27 -0.67  0.38  0.38 -0.53  0.53 -0.53    NA -0.40  0.50 -0.47 -0.40 -0.42
 0.64
P11 -0.75  0.14  0.20  0.25  0.29  0.29  0.37  0.43  0.37  0.40    NA  0.40  0.43 -0.50  0.48 -0.4
P12  0.33  0.20  0.27  0.33 -0.62 -0.62  0.47  0.53 -0.53 -0.50 -0.40    NA -0.47  0.60  0.58 -0.3
P13  0.31  0.18  0.25  0.31  0.36  0.36  0.44 -0.50 -0.56  0.47 -0.43  0.47    NA -0.43 -0.45 -0.3
P14  0.25  0.14  0.20  0.25  0.29  0.29  0.37 -0.57  0.37  0.40  0.50 -0.60  0.43    NA  0.48 -0.4
P15  0.27  0.15  0.21  0.27  0.31  0.31  0.39  0.45  0.39  0.42 -0.48 -0.58  0.45 -0.48
NA -0.44
P16  0.22  0.12  0.18  0.22  0.26  0.26  0.33  0.39  0.33 -0.64  0.46  0.36  0.39  0.46  0.44
NA
```

Find Goodness of Fit "by hand"

```
> btl <- score/(score %+% t(score))
> resid <- tournament - btl

> sum(resid^2,na.rm=TRUE)
[1] 41.78075
> sum(tournament^2,na.rm=TRUE)
[1] 120
> GF <- 1 - sum(resid^2,na.rm=TRUE)/sum(tournament^2,na.rm=TRUE)

> GF
[1] 0.651827
```

1. Find model
2. Find Residual = Model - Data
3. Goodness of Fit is $1 - Residual^2 / Data^2$

Use a simple function:

```
scaling.fits(score, tournament, test = "choice", rowwise = FALSE)
$GF
[1] 0.651827

$original
[1] 120

$resid
[1] 41.78075
```

Automate it by calling a function (scaling.fits) repeatedly for alternative models

These data may be analyzed using repeated calls to the `scaling.fits` function:

```
> tests <- c("choice","logit","normal")
>   fits <- matrix(NA, ncol = 3,nrow=4)
>  for (i in 1:4) {
+      for (j in 1:3) {
+          fits[i, j] <- scaling.fits(chess.df[i], data = tournament,
            test = tests[j],rowwise=FALSE)$GF[1]    } }
>  rownames(fits) <- c("latent","observed", "normed", "logistic")
>  colnames(fits) <- c("choice", "logistic", "normal")
>  round(fits, 2)
```

```
         choice logistic normal
latent     0.63     0.67   0.65         The generating data
observed   0.65     0.59   0.63         The observed data
normed     0.66     0.68   0.70          Normal transformed data
logistic   0.66     0.70   0.70         Logistic transformed data
```

Note how the scaled data fit the observed choices better than the actual observed orders fit.

Advanced: The scaling.fits function

```
> scaling.fits <-
function (model, data, test = "logit", digits = 2, rowwise = TRUE) {
    model <- as.matrix(model)
    data <- as.matrix(data)
    if (test == "choice") {
        model <- as.vector(model)
        if (min(model) <= 0)
            model <- model - min(model)
        prob = model/(model %+% t(model))
    }
    else {
        pdif <- model %+% -t(model)
        if (test == "logit") {
            prob <- 1/(1 + exp(-pdif))
        }
        else {
            if (test == "normal") {
                prob <- pnorm(pdif)
            }
        }
    }
    if (rowwise) {
        prob = 1 - prob
    }
    error <- data - prob
    sum.error2 <- sum(error^2, na.rm = TRUE)
    sum.data2 <- sum(data^2, na.rm = TRUE)
    gof <- 1 - sum.error2/sum.data2
    fit <- list(GF = gof, original = sum.data2, resid = sum.error2,
        residual = round(error, digits))
    return(fit)
}
```

Friendship as proximity

1. Chess or football provides a ranking based upon an ordering relationship ($p_i > p_j$).
2. Alternatively, friendship groups are based upon closeness ($|p_i - p_j| < \delta$)
   2.1 Do you know person j?
   2.2 Do you like person j? or as an alternative:
   2.3 Please list all your friends in this class (and is j included on the list)
   2.4 Would you be interested in having a date with person j?
   2.5 Would you like to have sex with person j?
   2.6 Would you marry person j?
3. Typically such data will be a rectangular matrix for there are asymmetries in closeness.

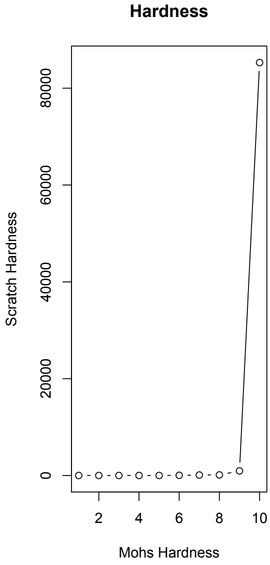
Moh's hardness scale provides rank orders of hardness

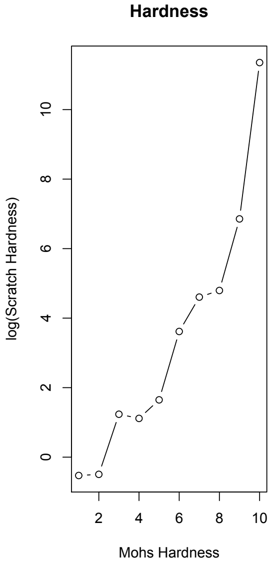
How hard is rock? The scratch test.

Table: Mohs' scale of mineral hardness. An object is said to be harder than X if it scratches X. Also included are measures of relative hardness using a sclerometer (for the hardest of the planes if there is a ansiotropy or variation between the planes) which shows the non-linearity of the Mohs scale (Burchard, 2004).

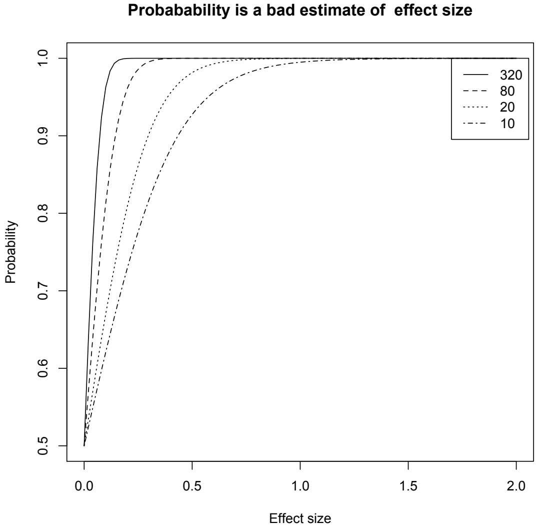
Mohs Hardness	Mineral	Scratch hardness
1	Talc	.59
2	Gypsum	.61
3	Calcite	3.44
4	Fluorite	3.05
5	Apaptite	5.2
6	Orthoclase Feldspar	37.2
7	Quartz	100
8	Topaz	121
9	Corundum	949
10	Diamond	85,300

Measuring Hardness – Scratch versus Mohs





Why we should report effect sizes rather than p: Another example of non-linearity



Effect Size is differences of means/ pooled within group standard deviation

$$d = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\sigma_1^2 n_1 + \sigma_2^2 n_2}/(n_1 + n_2)}$$

$$t = \frac{d}{\sqrt{1/n_1 + 1/n_2}}$$

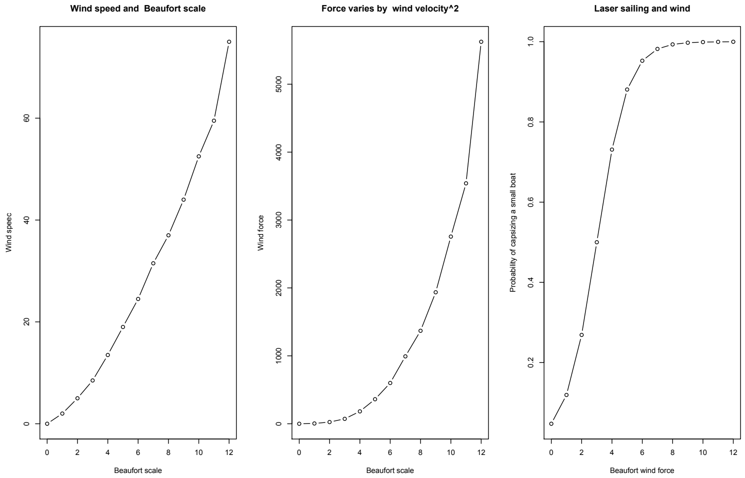
p(t) is non linear with t

Ordering based upon external measures

Table: The Beaufort scale of wind intensity is an early example of a scale with roughly equal units that is observationally based. Although the units are roughly in equal steps of wind speed in nautical miles/hour (knots), the force of the wind is not linear with this scale, but rather varies as the square of the velocity.

Force	Wind (Knots)	WMO Classification	Appearance of Wind Effects
0	Less than 1	Calm	Sea surface smooth and mirror-like
1	1-3	Light Air	Scaly ripples, no foam crests
2	4-6	Light Breeze	Small wavelets, crests glassy, no breaking
3	7-10	Gentle Breeze	Large wavelets, crests begin to break, scattered whitecaps
4	11-16	Moderate Breeze	Small waves 1-4 ft. becoming longer, numerous whitecaps
5	17-21	Fresh Breeze	Moderate waves 4-8 ft taking longer form, many whitecaps, some spray
6	22-27	Strong Breeze	Larger waves 8-13 ft, whitecaps common more spray
7	28-33	Near Gale	Sea heaps up, waves 13-20 ft, white foam streaks off breakers
8	34-40	Gale Moderately	High (13-20 ft) waves of greater length, edges of crests begin to break into spindrift, foam blown in streaks
9	41-47	Strong Gale	High waves (20 ft), sea begins to roll, dense streaks of foam, spray may reduce visibility
10	48-55	Storm	Very high waves (20-30 ft) with overhanging crests, sea white with densely blown foam, heavy rolling, lowered visibility
11	56-63	Violent Storm	Exceptionally high (30-45 ft) waves, foam patches cover sea, visibility more reduced
12	64+	Hurricane	Air filled with foam, waves over 45 ft, sea completely white with driving spray, visibility greatly reduced

The Beaufort scale is non-linear with force or probability of capsizing



Models of scaling objects

1. Assume each object ($a, b, ...z$) has a scale value ($A, B, ...Z$) with some noise for each measurement.
2. Probability of $A > B$ increases with difference between a and b
3. $P(A > B) = f(a - b)$
4. Can we find a function, f, such that equal differences in the latent variable (a, b, c) lead to equal differences in the observed variable?
5. Several alternatives
   - Direct scaling on some attribute dimension (simple but flawed)
   - Indirect scaling by paired comparisons (more complicated but probably better)

Scaling of Objects: O x O comparisons

1. Typical object scaling is concerned with order or location of objects
2. Subjects are assumed to be random replicates of each other, differing only as a source of noise
3. Absolute scaling techniques
   - Grant Proposals: 1 to 5
   - “On a scale from 1 to 10" this [object] is a X?
   - If A is 1 and B is 10, then what is C?
   - College rankings based upon selectivity
   - College rankings based upon "yield"
   - Zagat ratings of restaurants
   - A - F grading of papers

Absolute scaling: difficulties

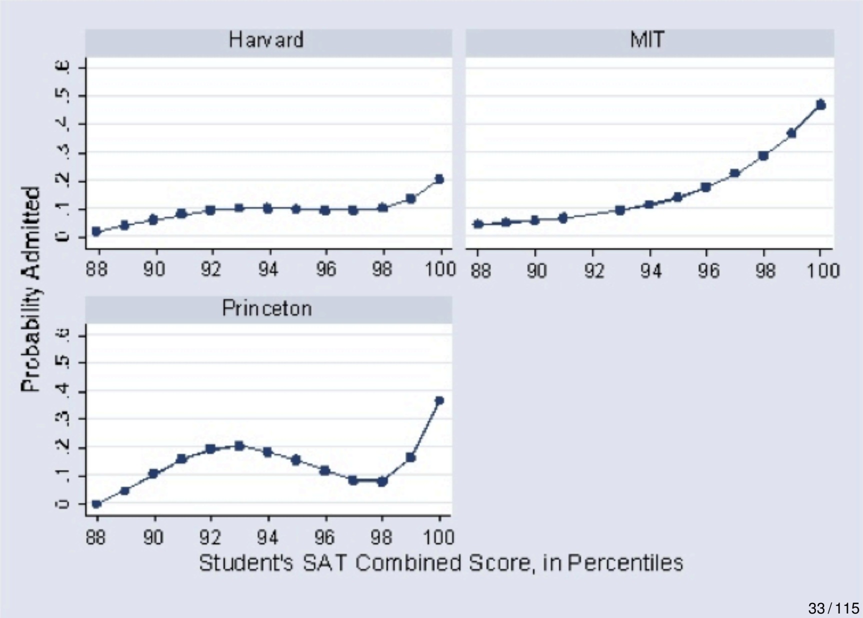
1. "On a scale from 1 to 10" this [object] is a X?
    - sensitive to context effects
    - what if a new object appears?
    - Need unbounded scale
2. If A is 1 and B is 10, then what is C?
    - results will depend upon A, B

Absolute scaling: artifacts

1. College rankings based upon selectivity
   - accept/applied
   - encourage less able to apply
2. College rankings based upon “yield"
   - matriculate/accepted
   - early admissions guarantee matriculation
   - don’t accept students who will not attend
3. Proposed solution: college choice as a tournament
   - Consider all schools that accept a student
   - Which school does he/she choose?

Avery et al. (2013)

A revealed preference ordering Avery et al. (2013)



A revealed preference ordering Avery et al. (2013)

TABLE III

A REVEALED PREFERENCE RANKING OF COLLEGES BASED ON MATRICULATION DECISIONS

Rank Based on Matriculation (with Covariates)	College Name	Theta	Implied Prob. of "Winning" vs. College Listed… 1 Row Below	Implied Prob. of "Winning" vs. College Listed… 10 Rows Below	Rank Based on Matriculation (no Covariates)
1	Harvard University	9.13	0.59	0.93	1
2	Caltech	8.77	0.56	0.92	3
3	Yale University	8.52	0.59	0.92	2
4	MIT	8.16	0.51	0.89	5
5	Stanford University	8.11	0.52	0.90	4
6	Princeton University	8.02	0.73	0.90	6
7	Brown University	7.01	0.56	0.78	7
8	Columbia University	6.77	0.54	0.73	8
9	Amherst College	6.61	0.51	0.71	9
10	Dartmouth	6.57	0.52	0.72	10
11	Wellesley College	6.51	0.53	0.71	12
12	University of Pennsylvania	6.39	0.56	0.71	11

Weber-Fechner Law and non-linearity of scales

1. Early studies of psychophysics by Weber (1834b,a) and subsequently Fechner (1860) demonstrated that the human perceptual system does not perceive stimulus intensity as a linear function of the physical input.
2. The basic paradigm was to compare one weight with another that differed by amount Δ, e.g., compare a 10 gram weight with an 11, 12, and 13 gram weight, or a 10 kg weight with a 11, 12, or 13 kg weight.
3. What was the Δ that was just detectable? The finding was that the perceived intensity follows a logarithmic function.
4. Examining the magnitude of the "*just noticeable difference*" or *JND*, Weber (1834b) found that

$$JND = \frac{\Delta Intensity}{Intensity} = constant. \qquad (5)$$

Weber-Fechner Law and non-linearity of scales

1. An example of a logarithmic scale of intensity is the decibel measure of sound intensity.
2. Sound Pressure Level expressed in decibels (dB) of the root mean square observed sound pressure, P_o (in Pascals) is

$$L_p = 20Log_{10}\frac{P_o}{P_{ref}} \tag{6}$$

3. where the reference pressure, P_{ref}, in the air is $20\mu Pa$.
4. Just to make this confusing, the reference pressure for sound measured in the ocean is $1\mu Pa$. This means that sound intensities in the ocean are expressed in units that are 20 dB higher than those units used on land.

The Just Noticeable Difference in Person and risk perception

1. Although typically thought of as just relevant for the perceptual experiences of physical stimuli, Ozer (1993) suggested that the JND is useful in personality assessment as a way of understanding the accuracy and inter judge agreement of judgments about other people.
2. In addition, Sinn (2003) has argued that the logarithmic nature of the *Weber-Fechner Law* is of evolutionary significance for preference for risk and cites Bernoulli (1738) as suggesting that our general utility function is logarithmic.
3. The whole of Prospect Theory (Kahneman and Tversky, 1979; Kahneman, 2011) is based upon this non-linearity of utilities: Better to skip lunch than be someone's dinner.

Money and non linearity

... the utility resulting from any small increase in wealth will be inversely proportionate to the quantity of goods already possessed .... if ... one has a fortune worth a hundred thousand ducats and another one a fortune worth same number of semi-ducats and if the former receives from it a yearly income of five thousand ducats while the latter obtains the same number of semi-ducats, it is quite clear that to the former a ducat has exactly the same significance as a semi-ducat to the latter (Bernoulli, 1738, p 25).

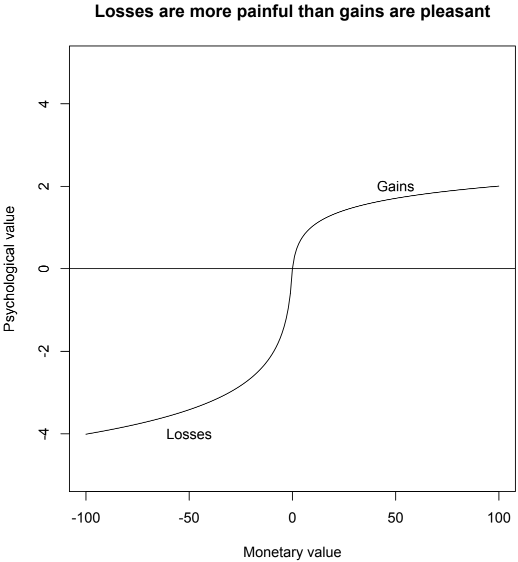
Implies a log function for utility.

Econs and Humans

1. Simple expected value theory $\Longrightarrow$ value = probability of event x value of event
2. Bernouli theory of expected utility came to dominate choice theory and is fundamental to economics
3. Studied by comparing gambles and showing utility is non linear with value
   - Would you rather have $80 or a 80% chance of $100 + 20% of $10?
   - expected value is 80 versus .8 * 100 + .2 * 10 = 82
4. Bernouli value (from Kahneman, 2011)

Wealth (millions)	1	2	3	4	5	6	7	8	9	10
Utility units	10	30	48	60	70	78	84	90	96	100

Kahneman and Tversky: Prospect Theory



Kahneman and Tversky (1979) Better to skip lunch than be someone's dinner.

Thurstonian scaling: basic concept

1. Every object has a value
2. Rated strength of object is noisy with Gaussian noise
3. $P(A > B) = f(z_a - z_b)$
4. Assume equal variance for each item
5. Convert choice frequency to normal deviates
6. Scale is average normal deviates

Thurstone choice model

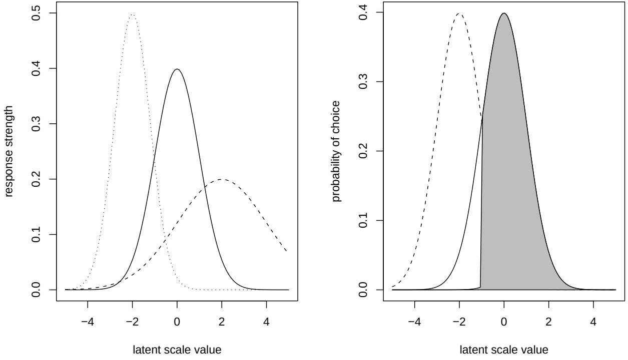


Figure: Thurstone's model of paired discrimination. Left panel: three items differ in their mean level as well as their variance. Right panel: choice between two items with equal variance reflects the relative strength of the two items. The shaded section represents choosing item 2 over item 1.

Thurstone's Vegetable data as an example of one dimensional scaling

Table: Consider the likelihood of liking a vegetable. Numbers reflect probability that the column is preferred to the row. Can we turn this into a scale?

The `veg` data set from the psych package in R

Variable	Turn	Cab	Beet	Asp	Car	Spin	S.Beans	Peas	Corn
Turn	0.50	0.82	0.77	0.81	0.88	0.89	0.90	0.89	0.93
Cab	0.18	0.50	0.60	0.72	0.74	0.74	0.81	0.84	0.86
Beet	0.23	0.40	0.50	0.56	0.74	0.68	0.84	0.80	0.82
Asp	0.19	0.28	0.44	0.50	0.56	0.59	0.68	0.60	0.73
Car	0.12	0.26	0.26	0.44	0.50	0.49	0.57	0.71	0.76
Spin	0.11	0.26	0.32	0.41	0.51	0.50	0.63	0.68	0.63
S.Beans	0.10	0.19	0.16	0.32	0.43	0.37	0.50	0.53	0.64
Peas	0.11	0.16	0.20	0.40	0.29	0.32	0.47	0.50	0.63
Corn	0.07	0.14	0.18	0.27	0.24	0.37	0.36	0.37	0.50

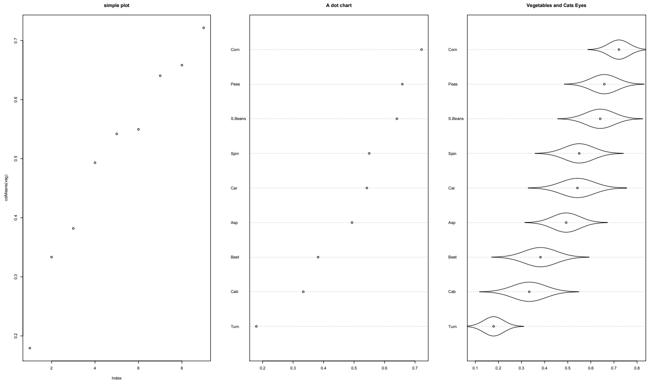
#show the data from the veg data set from the psych package
veg

(Guilford, 1954)

Some simple R

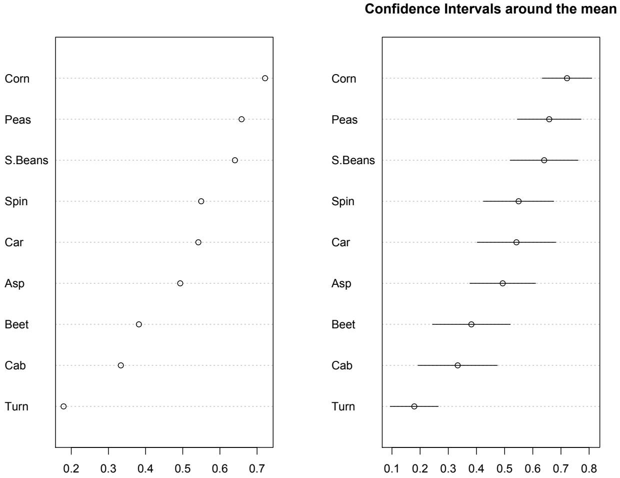
```
> veg      #shows the data

          Turn   Cab  Beet   Asp   Car  Spin S.Beans  Peas  Corn
Turn     0.500 0.818 0.770 0.811 0.878 0.892   0.899 0.892 0.926
Cab      0.182 0.500 0.601 0.723 0.743 0.736   0.811 0.845 0.858
Beet     0.230 0.399 0.500 0.561 0.736 0.676   0.845 0.797 0.818
Asp      0.189 0.277 0.439 0.500 0.561 0.588   0.676 0.601 0.730
Car      0.122 0.257 0.264 0.439 0.500 0.493   0.574 0.709 0.764
Spin     0.108 0.264 0.324 0.412 0.507 0.500   0.628 0.682 0.628
S.Beans  0.101 0.189 0.155 0.324 0.426 0.372   0.500 0.527 0.642
Peas     0.108 0.155 0.203 0.399 0.291 0.318   0.473 0.500 0.628
Corn     0.074 0.142 0.182 0.270 0.236 0.372   0.358 0.372 0.500
> colMeans(veg) #show the means (but too many decimals)
     Turn       Cab      Beet       Asp       Car      Spin   S.Beans
Peas      Corn
0.1793333 0.3334444 0.3820000 0.4932222 0.5420000 0.5496667 0.6404444 0.65
> round(colMeans(veg))     #round off, but not enough decimals
   Turn     Cab    Beet     Asp     Car    Spin S.Beans    Peas    Corn
      0       0       0       0       1       1       1       1
1
> round(colMeans(veg),2) #this looks pretty good
   Turn     Cab    Beet     Asp     Car    Spin S.Beans    Peas    Corn
   0.18    0.33    0.38    0.49    0.54    0.55    0.64    0.66    0.72
```

Three ways of plotting the data



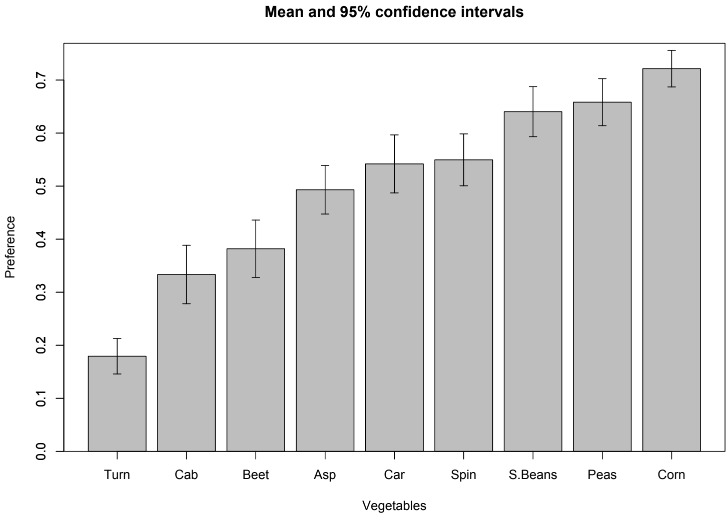

```
op<- par(mfrow=c(1,3))    #I want to draw three graphs
 plot(colMeans(veg),main="simple plot")          #the basic plot command
 dotchart(colMeans(veg),main="A dot chart")    #dot charts are more informative
error.dots(veg,eyes=TRUE,main="Vegetables and Cats Eyes")
 op <- par(mfrow=c(1,1))    #set the plotting back to a single graph
```

And yet more ways of plotting the data




```
op <- par(mfrow=c(1,2))    #I want to draw two graphs
dotchart(colMeans(veg))    #dot charts are more informative
error.dots(veg)         #add error bars to the plot
op <- par(mfrow=c(1,1))    #set the plotting back to a single graph
```

Alternatively, use the `error.bars` function from *psych*




```
error.bars(veg, bars=TRUE, ylab="Preference", xlab="Vegetables", main="Mean and 95% confidence interv
```

Naive scaling

```
> round(veg,2)
```

	Turn	Cab	Beet	Asp	Car	Spin	S.Beans	Peas	Corn
Turn	0.50	0.82	0.77	0.81	0.88	0.89	0.90	0.89	0.93
Cab	0.18	0.50	0.60	0.72	0.74	0.74	0.81	0.84	0.86
Beet	0.23	0.40	0.50	0.56	0.74	0.68	0.84	0.80	0.82
Asp	0.19	0.28	0.44	0.50	0.56	0.59	0.68	0.60	0.73
Car	0.12	0.26	0.26	0.44	0.50	0.49	0.57	0.71	0.76
Spin	0.11	0.26	0.32	0.41	0.51	0.50	0.63	0.68	0.63
S.Beans	0.10	0.19	0.16	0.32	0.43	0.37	0.50	0.53	0.64
Peas	0.11	0.16	0.20	0.40	0.29	0.32	0.47	0.50	0.63
Corn	0.07	0.14	0.18	0.27	0.24	0.37	0.36	0.37	0.50

```
> round(colMeans(veg),2)
   Turn     Cab    Beet     Asp     Car    Spin S.Beans
Peas   Corn
   0.18    0.33    0.38    0.49    0.54    0.55    0.64
0.66   0.72

> veg.t <- colMeans(veg) - mean(veg[,1])

> round(veg.t,2)

   Turn     Cab    Beet     Asp     Car    Spin S.Beans
Peas   Corn
   0.00    0.15    0.20    0.31    0.36    0.37    0.46
0.48   0.54
```

1. Show the data
2. Find the mean for each column. Round to 2 decimals
3. Subtract the mean for the first column from the means
4. But these are not really useful scale values.

Convert the vegetables data set to normal deviates

```
> z.veg <- qnorm(as.matrix(veg))
> round(z.veg,2)              #see table
```

	Turn	Cab	Beet	Asp	Car	Spin	S.Beans	Peas	Corn
Turn	0.00	0.91	0.74	0.88	1.17	1.24	1.28	1.24	1.45
Cab	−0.91	0.00	0.26	0.59	0.65	0.63	0.88	1.02	1.07
Beet	−0.74	−0.26	0.00	0.15	0.63	0.46	1.02	0.83	0.91
Asp	−0.88	−0.59	−0.15	0.00	0.15	0.22	0.46	0.26	0.61
Car	−1.17	−0.65	−0.63	−0.15	0.00	−0.02	0.19	0.55	0.72
Spin	−1.24	−0.63	−0.46	−0.22	0.02	0.00	0.33	0.47	0.33
S.Beans	−1.28	−0.88	−1.02	−0.46	−0.19	−0.33	0.00	0.07	0.36
Peas	−1.24	−1.02	−0.83	−0.26	−0.55	−0.47	−0.07	0.00	0.33
Corn	−1.45	−1.07	−0.91	−0.61	−0.72	−0.33	−0.36	−0.33	0.00

```
> scaled.veg <- colMeans(z.veg)
> round(scaled.veg,2)
   Turn     Cab    Beet     Asp     Car    Spin S.Beans
Peas    Corn
  -0.99   -0.47   -0.33   -0.01    0.13    0.16    0.41
0.46    0.64
> scaled <- scaled.veg - min(scaled.veg)
> round(scaled,2)
   Turn     Cab    Beet     Asp     Car    Spin S.Beans
Peas    Corn
   0.00    0.52    0.65    0.98    1.12    1.14    1.40
1.44    1.63
```

1. Convert to normal deviates using the norm function. But that only works on matrices, so we need to convert the data.frame into a matrix.
2. Display the data
3. Find the column means and show them
4. subtract the smallest value to form a positive scale.

Form the model based upon these scale values

```
> pdif <- - scaled %+% t(scaled)
>  colnames(pdif) <- rownames(pdif) <- colnames(z.veg)
>  round(pdif,2)
```

	Turn	Cab	Beet	Asp	Car	Spin	S.Beans	Peas	Corn
Turn	0.00	0.52	0.65	0.98	1.12	1.14	1.40	1.44	1.63
Cab	−0.52	0.00	0.13	0.46	0.60	0.62	0.88	0.92	1.11
Beet	−0.65	−0.13	0.00	0.33	0.46	0.49	0.75	0.79	0.98
Asp	−0.98	−0.46	−0.33	0.00	0.14	0.16	0.42	0.46	0.65
Car	−1.12	−0.60	−0.46	−0.14	0.00	0.03	0.28	0.33	0.51
Spin	−1.14	−0.62	−0.49	−0.16	−0.03	0.00	0.26	0.30	0.49
S.Beans	−1.40	−0.88	−0.75	−0.42	−0.28	−0.26	0.00	0.04	0.23
Peas	−1.44	−0.92	−0.79	−0.46	−0.33	−0.30	−0.04	0.00	0.19
Corn	−1.63	−1.11	−0.98	−0.65	−0.51	−0.49	−0.23	−0.19	0.00

```
> modeled <- pnorm(pdif)
> round(modeled,2)
```

	Turn	Cab	Beet	Asp	Car	Spin	S.Beans	Peas	Corn
Turn	0.50	0.70	0.74	0.84	0.87	0.87	0.92	0.93	0.95
Cab	0.30	0.50	0.55	0.68	0.72	0.73	0.81	0.82	0.87
Beet	0.26	0.45	0.50	0.63	0.68	0.69	0.77	0.79	0.84
Asp	0.16	0.32	0.37	0.50	0.55	0.57	0.66	0.68	0.74
Car	0.13	0.28	0.32	0.45	0.50	0.51	0.61	0.63	0.70
Spin	0.13	0.27	0.31	0.43	0.49	0.50	0.60	0.62	0.69
S.Beans	0.08	0.19	0.23	0.34	0.39	0.40	0.50	0.52	0.59
Peas	0.07	0.18	0.21	0.32	0.37	0.38	0.48	0.50	0.57
Corn	0.05	0.13	0.16	0.26	0.30	0.31	0.41	0.43	0.50

1. Subtract the column value from the row value using the `matrix.addition` function from *psych*.
2. Show the result
3. Convert the normal deviates into probabilities using the `norm` function.

Data = Model + Residual

1. What is the model?
    - Pref = Mean (preference)
    - $p(A > B) = f(A, B)$
    - what is f?
2. Possible functions
    - f = A - B (simple difference)
    - $\frac{A}{A+B}$ Luce choice rule
    - Thurstonian scaling
    - logistic scaling
3. Evaluating functions – Goodness of fit
    - Residual = Model - Data
    - Minimize residual
    - Minimize *residual*^{2}

Examine the residuals

```
> resid <- veg - modeled
> round(resid,2)
```

	Turn	Cab	Beet	Asp	Car	Spin	S.Beans	Peas	Corn
Turn	0.00	0.12	0.03	-0.03	0.01	0.02	-0.02	-0.03	-0.02
Cab	-0.12	0.00	0.05	0.05	0.02	0.00	0.00	0.02	-0.01
Beet	-0.03	-0.05	0.00	-0.07	0.06	-0.01	0.07	0.01	-0.02
Asp	0.03	-0.05	0.07	0.00	0.01	0.02	0.01	-0.08	-0.01
Car	-0.01	-0.02	-0.06	-0.01	0.00	-0.02	-0.04	0.08	0.07
Spin	-0.02	0.00	0.01	-0.02	0.02	0.00	0.03	0.06	-0.06
S.Beans	0.02	0.00	-0.07	-0.01	0.04	-0.03	0.00	0.01	0.05
Peas	0.03	-0.02	-0.01	0.08	-0.08	-0.06	-0.01	0.00	0.05
Corn	0.02	0.01	0.02	0.01	-0.07	0.06	-0.05	-0.05	0.00

```
> sum(resid)

[1] 3.816392e-16

> sum(resid^2)

[1] 0.1416574

> sum(resid^2)/sum(veg^2)

[1] 0.005697482
```

1. Subtract the model from the data to find the residuals
2. Sum the residuals (equal 0)
3. Sum the squared residuals
4. Compare this to the original data (badness of fit)
5. Convert to a goodness of fit

Consider alternative scaling models

	constant	equal	squared	reversed	raw	thurstone
Turn	0.5	1	1	9	0.00	0.00
Cab	0.5	2	4	8	0.15	0.52
Beet	0.5	3	9	7	0.20	0.65
Asp	0.5	4	16	6	0.31	0.98
Car	0.5	5	25	5	0.36	1.12
Spin	0.5	6	36	4	0.37	1.14
S.Beans	0.5	7	49	3	0.46	1.40
Peas	0.5	8	64	2	0.48	1.44
Corn	0.5	9	81	1	0.54	1.63

	choice	logistic	normal
Constant	0.81	0.81	0.81
Equal	0.99	0.88	0.81
Squared	0.98	0.74	0.74
Reversed	0.40	−0.27	−0.43
Raw	0.97	0.89	0.93
Thurstone	0.97	0.97	0.99

1. Constant says all items are equal
2. Equal implies the steps are all 1
3. Square the values of equal
4. Reverse the rank order!
5. Just the scale values based upon means
6. Thurstonian scaling

Thurstonian scaling as an example of model fitting

We don't really care all that much about vegetables, but we do care about the process of model fitting.

1. Examine the data
2. Specify a model
3. Estimate the model
4. Compare the model to the data
5. Repeat until satisfied or exhausted

Multidimensional Scaling: ($|o_i - o_j| < |o_k - o_l|$)

$$Distance_{xy} = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}. \tag{7}$$

Consider the `cities` data set of airline distances.

```
> cities
```

	ATL	BOS	ORD	DCA	DEN	LAX	MIA	JFK	SEA	SFO	MSY
ATL	0	934	585	542	1209	1942	605	751	2181	2139	424
BOS	934	0	853	392	1769	2601	1252	183	2492	2700	1356
ORD	585	853	0	598	918	1748	1187	720	1736	1857	830
DCA	542	392	598	0	1493	2305	922	209	2328	2442	964
DEN	1209	1769	918	1493	0	836	1723	1636	1023	951	1079
LAX	1942	2601	1748	2305	836	0	2345	2461	957	341	1679
MIA	605	1252	1187	922	1723	2345	0	1092	2733	2594	669
JFK	751	183	720	209	1636	2461	1092	0	2412	2577	1173
SEA	2181	2492	1736	2328	1023	957	2733	2412	0	681	2101
SFO	2139	2700	1857	2442	951	341	2594	2577	681	0	1925
MSY	424	1356	830	964	1079	1679	669	1173	2101	1925	0

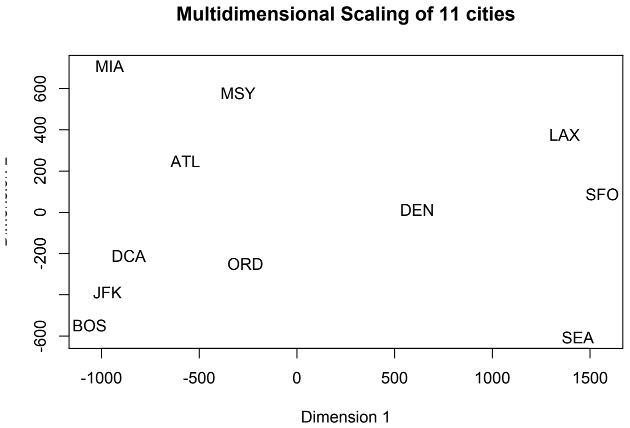
A two dimensional solution of the airline distances

```
> city.location <- cmdscale(cities, k=2)
> plot(city.location,type="n", xlab="Dimension␣1",
   ylab="Dimension␣2",main ="cmdscale(cities)")
> text(city.location,labels=names(cities))
> round(city.location,0)
```

	[,1]	[,2]
ATL	-571	248
BOS	-1061	-548
ORD	-264	-251
DCA	-861	-211
DEN	616	10
LAX	1370	376
MIA	-959	708
JFK	-970	-389
SEA	1438	-607
SFO	1563	88
MSY	-301	577

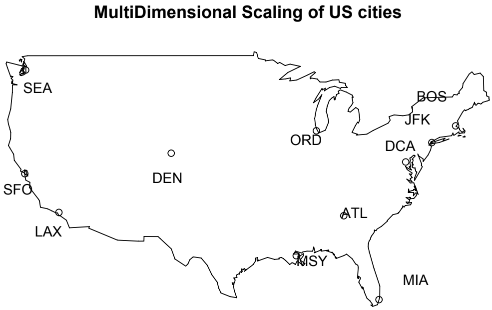
1. Use the `cmdscale` function to do multidimensional scaling, ask for a 2 dimensional solution
2. Plot the results (don't actually show the points)
3. Add the names of the cities
4. Show the numeric results

Original solution for 11 US cities. What is wrong with this figure? Axes of solutions are not necessarily directly interpretable.



Revised solution for 11 US cities after making `city.location <- -city.location` and adding a US map.

The correct locations of the cities are shown with circles. The MDS solution is the center of each label. The central cities (Chicago, Atlanta, and New Orleans are located very precisely, but Boston, New York and Washington, DC are north and west of their correct locations.



Preferential Choice: Unfolding Theory ($|s_i - oj| < |s_k - o_l|$)

1. “Do I like asparagus more than you like broccoli?" compares how far apart my ideal vegetable is to a particular vegetable (asparagus) with respect to how far your ideal vegetable is to another vegetable (broccoli).
2. More typical is the question of whether you like asparagus more than you like broccoli. This comparison is between your ideal point (on an attribute dimension) to two objects on that dimension.
3. Although the comparisons are ordinal, there is a surprising amount of metric information in the analysis.
4. This involves *unfolding* the individual preference orderings to find a joint scale of individuals and stimuli (Coombs, 1964, 1975).
5. Can now be done using multidimensional scaling of people and objects using proximity measures.

Some data collection

Please email me your answers to the these two set of questions

1. If you had complete choice in your life, how many children would you like to have?
2. If had complete choice, please rank order the number of children you would like to have
3. Part 2:
4. If you had complete choice in your life, how many children would you like to have? Call that A
5. If you could not have N children, would you rather have A - 1 or A + 1 (call that B)
6. If you could not have A or B children, would you rather have 1 + max(A,B) or min(A,B) - 1, call that C
7. If could not have A,B, or C children, would you rather have 1 + max(A,B,C) or min(A,B,C) - 1, call that D
8. If could not have A,B, C or Dchildren, would you rather have 1 + max(A,B,C,D) or min(A,B,C,D) - 1, call that E

Please email revelle@northwestern.edu your answers

Two different J scales

Table: Midpoint ordering gives some metric information. Left hand side: If the midpoint (2|3) comes after (to the right of) the midpoint (0|5) that implies that 3 is closer to 5 than 0 is to 2. Right hand side: The midpoint (2|3) comes before (0|5) and thus 2 is closer to 0 than 3 is to 5. Similarly, that 2|5 comes before 3|4 implies that 4 is closer to 5 than 2 is to 3.

<table>
<tr><td>0</td><td></td><td>1</td><td></td><td></td><td>2</td><td></td><td>3</td><td></td><td>4</td><td>5</td></tr>
<tr><td>0</td><td></td><td></td><td></td><td>0|5</td><td></td><td></td><td></td><td></td><td></td><td>5</td></tr>
<tr><td>0</td><td></td><td></td><td></td><td></td><td></td><td>2|3</td><td></td><td></td><td></td><td>5</td></tr>
<tr><td>0</td><td></td><td></td><td>1|2</td><td></td><td></td><td></td><td></td><td></td><td></td><td>5</td></tr>
<tr><td>0</td><td>0|1</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>4|5</td><td>5</td></tr>
<tr><td>0</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>3|4</td><td></td><td>5</td></tr>
<tr><td>0</td><td></td><td></td><td></td><td></td><td></td><td></td><td>2|5</td><td></td><td></td><td>5</td></tr>
</table>

Measuring Abilities and Attitudes

1. Abilities and most models of personality assume an order relationship
   - The comparison is between the person and an item.
   - $s_i > o_j$
   - A measurement mode without error is the Guttman scale where $prob(correct|\theta, \delta) = 1|\theta > \delta, \quad 0|\theta < \delta$
   - With error, a prototypical example is the Rasch scale where $prob(correct|\theta, \delta) = f(\theta - \delta)$
2. Attitudes (and some personality models) assume a single peak (non-monotone) ordering
   - People endorse attitudes that they are close to, and reject more extreme items.

The Bogardus Social Distance scale as a Guttman scale

Table: The *Bogardus Social Distance Scale* is one example of items that can be made to a *Guttman scale*

"According to my first feeling reactions I would willingly admit members of each race (as a class, and not the best I have known, nor the worst member) to one or more of the classifications under which I have placed a cross (x)."

1. Would exclude from my country
2. As visitors only to my country
3. Citizenship in my country
4. To employment in my occupation in my country
5. To my street as neighbors
6. To my club as personal chums
7. To close kinship by marriage

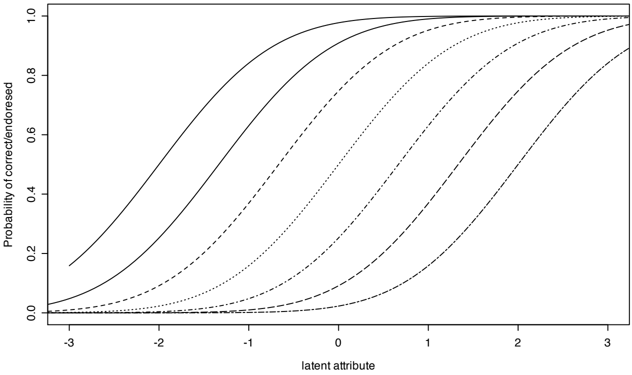
(Bogardus, 1925)

Creating a Guttman scale

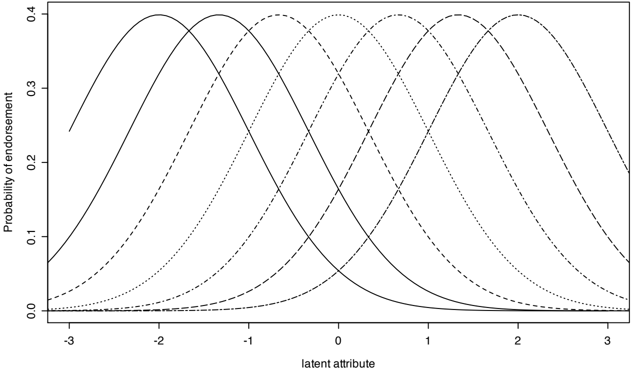
```
> guttman <- matrix(rep(0,56),nrow=8)
> for (i in 1:7) { for (j in 1:i) {guttman[i+1,j] <- 1}}
> rownames(guttman) <- paste("S",1:8,sep="")
> colnames(guttman) <- paste("O",1:7,sep="")
> guttman
   O1 O2 O3 O4 O5 O6 O7
S1  0  0  0  0  0  0  0
S2  1  0  0  0  0  0  0
S3  1  1  0  0  0  0  0
S4  1  1  1  0  0  0  0
S5  1  1  1  1  0  0  0
S6  1  1  1  1  1  0  0
S7  1  1  1  1  1  1  0
S8  1  1  1  1  1  1  1
> rowSums(guttman)
S1 S2 S3 S4 S5 S6 S7 S8
 0  1  2  3  4  5  6  7
```

1. Create a matrix of 0s
2. Add 1s below the diagonal
3. Give the rows and columns names
4. Show it
5. “score” it

A basic error model with parallel trace lines



Non-monotonic trace lines measure attitudes



Preferential choice as a comparison of an ideal point for the individual and the objects being chosen

1. The data are rank orders, but the analysis can produce quasi interval scales
2. Originally done by hand, now can be done using the *smacof* package (de Leeuw and Mair, 2009)

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.

Scale	Basic operations	Transformations	Invariant statistic	Examples
Nominal	equality $x_i = x_j$	Permutations	Counts Mode χ^2 and (ϕ) correlation	Detection Species classification Taxons
Ordinal	order $x_i > x_j$	Monotonic (homeomorphic) x' =f(x) f is monotonic	Median Percentiles Spearman correlations*	Mhos Hardness scale Beaufort Wind (intensity) Richter earthquake scale
Interval	differences $(x_i - x_j) > (x_k - x_l)$	Linear (Affine) x' = a + bx	Mean (μ) Standard Deviation (σ) Pearson correlation (r) Regression (β)	Temperature (°F, °C) Beaufort Wind (velocity)
Ratio	ratios $\frac{x_i}{x_j} > \frac{x_k}{x_l}$	Multiplication (Similiarity) x' = bx	Coefficient of variation ($\frac{\sigma}{\mu}$)	Length, mass, time Temperature (°K) Heating degree days

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 linear measure of

Graphical and tabular summaries of data

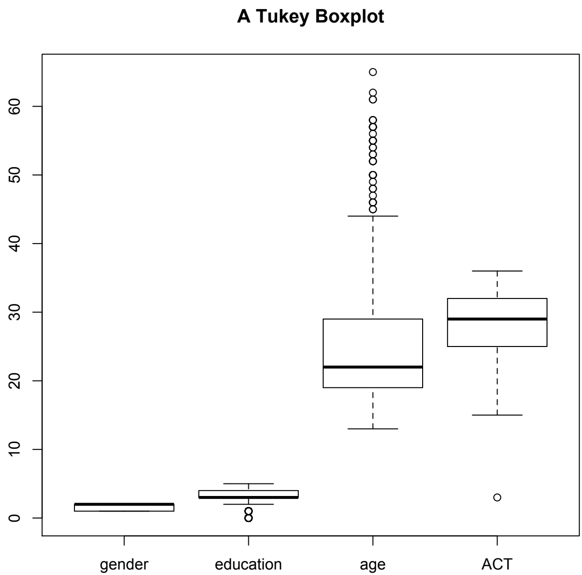
1. The Tukey 5 number summary shows the important characteristics of a set of numbers
   - Maximum
   - 75th percentile
   - Median (50th percentile)
   - 25th percentile
   - Minimum
2. Graphically, this is the box plot
   - Variations on the box plot include confidence intervals for the median

The summary command gives the Tukey 5 numbers

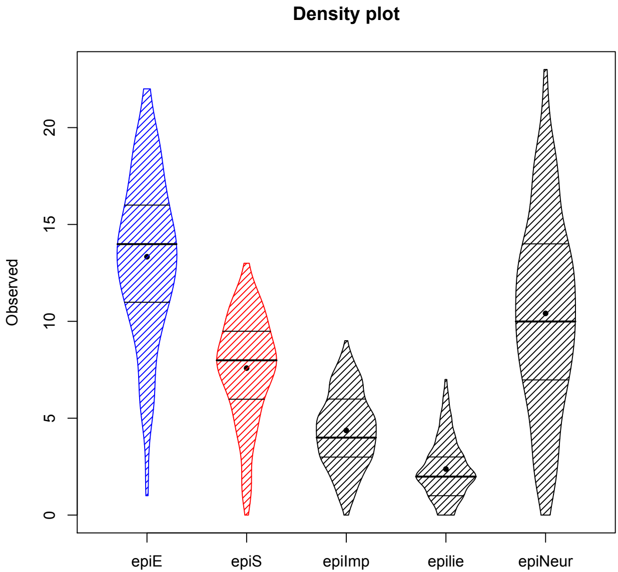
```
> summary(sat.act)
```

```
     gender        education          age             ACT             SATV            SATQ
 Min.   :1.000   Min.   :0.000   Min.   :13.00   Min.   : 3.00   Min.   :200.0   Min.   :200.
 1st Qu.:1.000   1st Qu.:3.000   1st Qu.:19.00   1st Qu.:25.00   1st Qu.:550.0   1st Qu.:530.
 Median :2.000   Median :3.000   Median :22.00   Median :29.00   Median :620.0   Median :620.
 Mean   :1.647   Mean   :3.164   Mean   :25.59   Mean   :28.55   Mean   :612.2   Mean   :610.
 3rd Qu.:2.000   3rd Qu.:4.000   3rd Qu.:29.00   3rd Qu.:32.00   3rd Qu.:700.0   3rd Qu.:700.
 Max.   :2.000   Max.   :5.000   Max.   :65.00   Max.   :36.00   Max.   :800.0   Max.   :800.
                                                                                 NA's   :13
```

A box plot of the first 4 sat.act variables



A violin or density plot of the first 5 epi.bfi variables



The `describe` function gives more descriptive statistics

```
> 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 $\bar{X} = X_. = (\sum_{i=1}^{N} X_i)/N$ `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. `mean(x,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. `winsor(x,trim=.2)`

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

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

Circular Mean $\bar{x}_{circular} = tan^{-1}\left(\frac{\sum cos(x)}{\sum sin(x)}\right)$ `circular.mean(x)` (where x is in radians)

circadian.mean `circular.mean(x)` (where x is in hours)

Circular statistics

For variables that vary geographically (e.g., wind direction, flying direction) or diurnally, seasonally (e.g. arousal, positive affect).

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.

Subject	Energetic Arousal	Positive Affect	Tense Arousal	Negative Affect
1	9	14	19	24
2	11	16	21	2
3	13	18	23	4
4	15	20	1	6
5	17	22	3	8
6	19	24	5	10
Arithmetic Mean	14	19	12	9
Circular Mean	14	19	24	5

Some hypothetical data stored in a data.frame

Participant	Name	Gender	θ	X	Y	Z
1	Bob	Male	1	12	2	1
2	Debby	Female	3	14	6	4
3	Alice	Female	7	18	14	64
4	Gina	Female	6	17	12	32
5	Eric	Male	4	15	8	8
6	Fred	Male	5	16	10	16
7	Chuck	Male	2	13	4	2

```
> 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
```

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 Rdirectly.

```
> 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.

```
> 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	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
3	3	Alice	Female	7	18	14	64

It was harder to see the perfect relationship between θ and X, Y, and Z with the original data.

Multiple estimates of the central tendency using the apply function

```
> apply(my.data[4:7],2,mean)
   theta          X          Y          Z

 4.00000  15.00000   8.00000  18.14286

  > apply(my.data[4:7],2,mean, trim=.2)

theta      X      Y      Z
  4.0   15.0    8.0   12.4

 > apply(my.data[4:7],2,winsor.mean, trim=.2)

  theta          X          Y          Z
 4.00000  15.00000   8.00000  12.91429

 > apply(my.data[4:7],2,harmonic.mean)

  theta           X           Y           Z
 2.699725  14.729687   5.399449   3.527559

 > apply(my.data[4:7],2,geometric.mean)
```

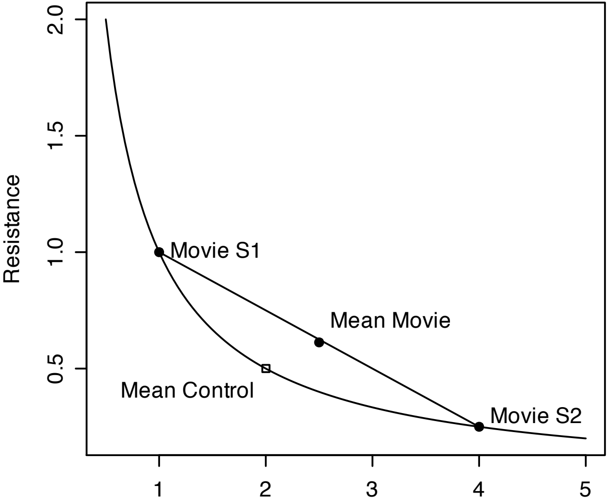
1. The basic mean is applied to columns 4 - 7
2. Then do this, but trim the top and bottom 20%
3. Now, don't trim, but winsorize
4. Compare with the harmonic mean
5. Compare with geometric mean.

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)

Condition	Subject	Skin Conductance	Skin Resistance
Pretest (Control)	1	2	.50
	2	2	.50
Average		2	.50
Posttest (Movie)	1	1	1.00
	2	4	.25
Average		2.5	.61

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 course. But what about the students?

Faculty Member	Freshman/ Sophmore	Junior	Senior	Graduate	Mean	Median
A	20	10	10	10	12.5	10
B	20	10	10	10	12.5	10
C	20	10	10	10	12.5	10
D	20	100	10	10	35.0	15
E	200	100	400	10	177.5	150
Total						
Mean	56	46	110	10	50.0	39
Median	20	10	10	10	12.5	10

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.

Class size	Number of classes	number of students
10	12	120
20	4	80
100	2	200
200	1	200
400	1	400

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) (1 * 1 + 19 * 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 (52 * 1 + 19 * 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

School	Pretest	Posttest	Change
Junior College	1	5	4
Non-selective university	5	27	22
Selective university	27	73	45

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

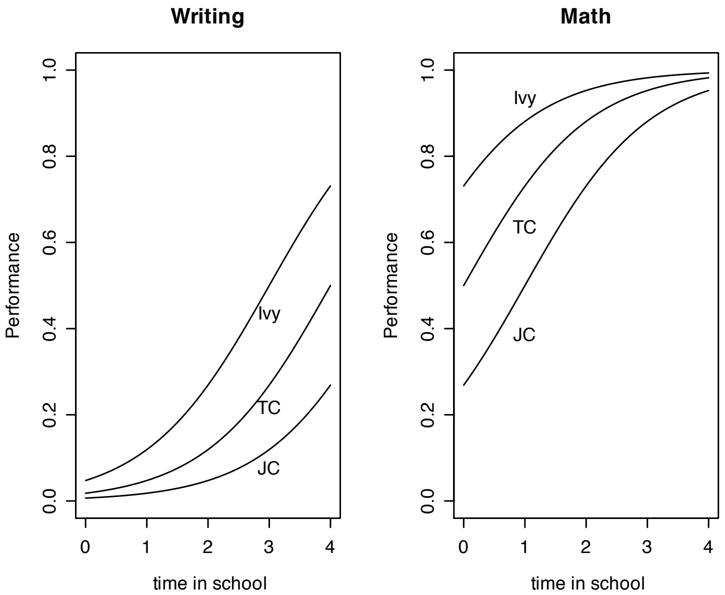
Another research team in motivational and educational psychology was interested in the effect that different teaching at various colleges and universities affect math performance. They used the same schools as the previous example with the same design.

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

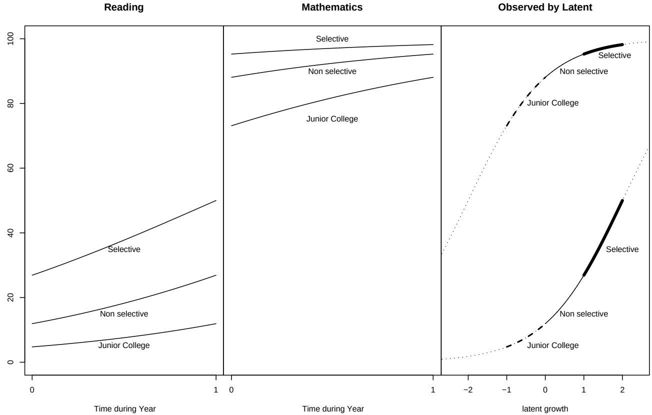
School	Pretest	Posttest	Change
Junior College	27	73	45
Non-selective university	73	95	22
Selective university	95	99	4

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?



The effect of scaling upon the latent variable - observed variable relationship

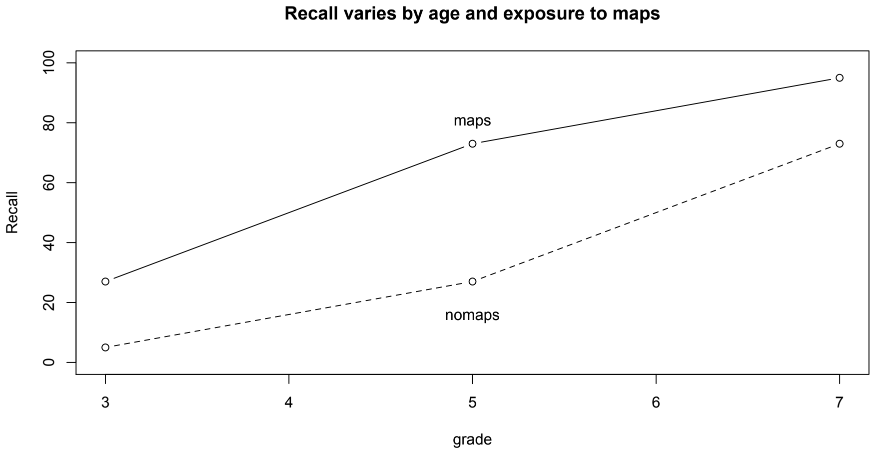


The problem of scaling is ubiquitous

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

	No Map	Maps	Effect	
3rd grade	5	27	22	Too young
5th grade	27	73	46	Critical period
7th grade	73	95	22	Too old

Map use is most effective at a particular developmental stage



R code for the prior figure

R code

```
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)
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(5,80,"maps")   #add line labels
> text(5,15,"nomaps")
```

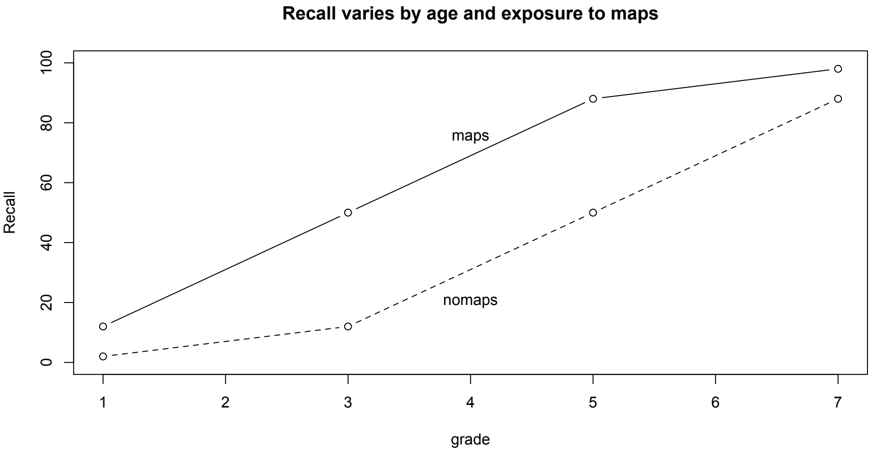
```
    grade nomaps maps
3rd     3      5   27
5th     5     27   73
7th     7     73   95
```

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

	No Map	Maps	Effect	
1st grade	2	12	10	Too young
3rd grade	12	50	38	
5th grade	50	88	38	Critical period
7th grade	88	98	10	Too old

A critical period in developmental?



R code for the prior figure

R code

```
mapuse <- matrix(c( 1,2,12,10,3,12,50,38,5,50,88,38,7,88,98,10),ncol=4
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")
```

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

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.
4. Ratio scales preserve absolute magnitude differences.

Ordinal scales

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.
5. Most data are normal towards the middle.
6. Most monotonic relationships are somewhat linear in the middle.

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

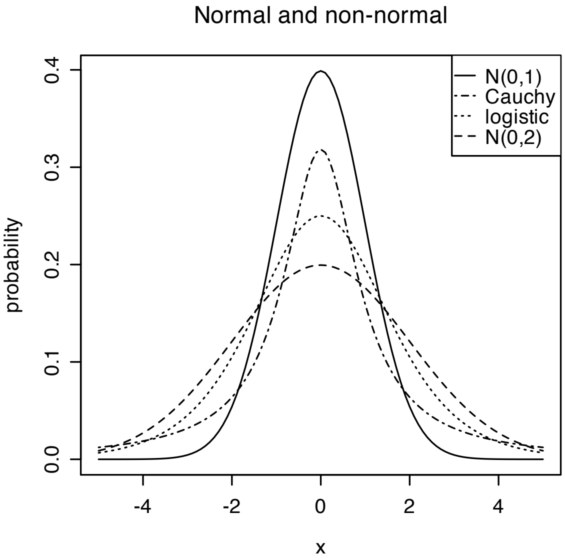
Variance (σ^2) is the Mean Square deviation (implies interval data)

Standard Deviation (σ) is the Root Mean Square deviation.

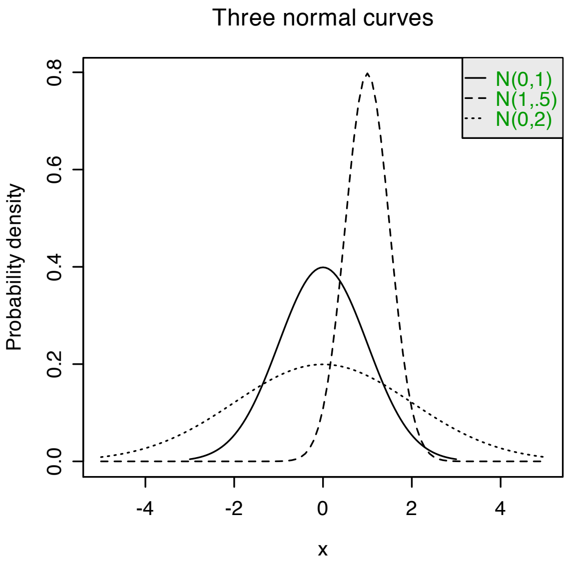
Coefficient of Variation $\frac{\sigma_x}{\mu_x}$

Average difference $\sigma_x\sqrt{2}$

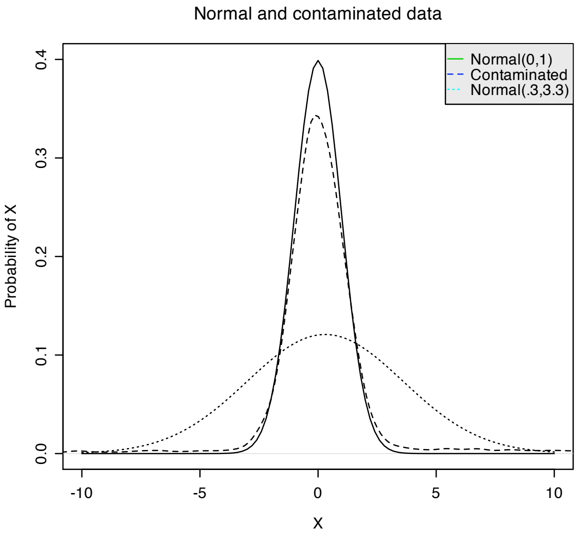
Normal and non-normal curves



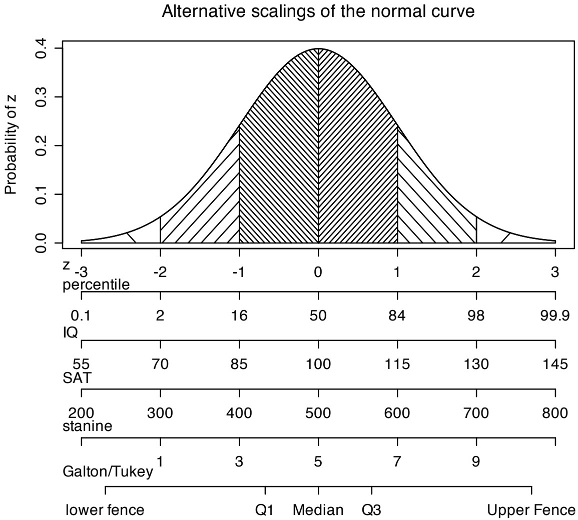
Three normal curves



Seriously contaminated data



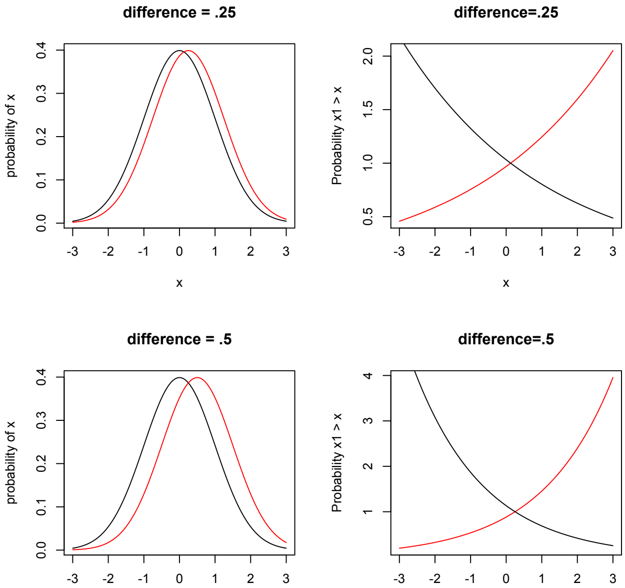
The normal curve and its frequent transforms



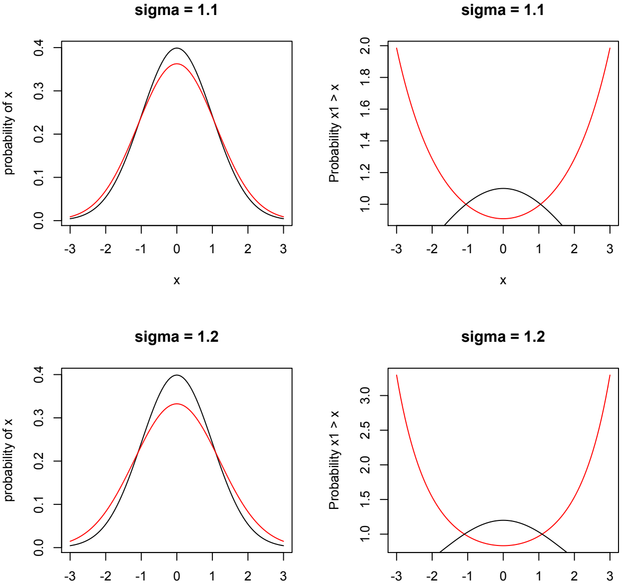
Decision making and the benefit of extreme selection ratios

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?
5. Climate change is a nice example, a 2° change in mean leads to a large increase in extreme events.

The effect of small mean differences at the tails of a distribution



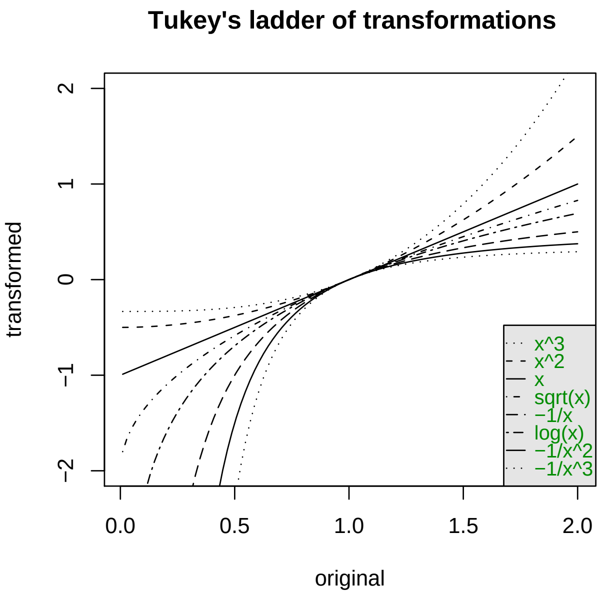
The effect of small differences in variance at the tails of a distribution



Tukey's ladder

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.

Transformation	effect	
x^3	emphasize large numbers	reduce negative skew
x^2	emphasize large numbers	reduce negative skew
x	the basic data	
$\sqrt{x}$	emphasize smaller numbers	reduce positive skew
-1/x	emphasize smaller numbers	reduce positive skew
log(x)	emphasize smaller numbers	reduce positive skew
$-1/x^2$	emphasize smaller numbers	reduce positive skew
$-1/x^3$	emphasize smaller numbers	reduce positive skew


Tukey's ladder of transformations
transformed
original
0.0
0.5
1.0
1.5
2.0
−2
−1
0
1
2
x^3
x^2
x
sqrt(x)
−1/x
log(x)
−1/x^2
−1/x^3

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. Probabilty 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 ($s = at^2 <=> t = \sqrt{\frac{s}{a}}$)
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(observed/reference)
9. pH of solutions is -log(concentration of hydrogen ions)

Avery, C. N., Glickman, M. E., Hoxby, C. M., and Metrick, A. (2013). A revealed preference ranking of U.S. colleges and universities. *The Quarterly Journal of Economics*, 128(1):425–467.

Bernoulli, D. (1954/1738). Exposition of a new theory on the measurement of risk ("Specimen theoriae novae de mensura sortis," Commentarii Academiae Scientiarum Imperialis Petropolitanae 5, St. Petersburg 175-92.) translated by Louise C. Sommer. *Econometrica*, 22(1):23–36.

Bogardus, E. S. (1925). Measuring social distances. *Journal of Applied Sociology*, pages 299–308.

Bradley, R. A. and Terry, M. E. (1952). Rank analysis of incomplete block designs: I. the method of paired comparisons. *Biometrika*, 39(3/4):324–345.

Burchard, U. (2004). The sclerometer and the determination of the hardness of minerals. *Mineralogical Record*, 35:109–120.

Coombs, C. (1964). *A Theory of Data*. John Wiley, New York.

Coombs, C. (1975). A note on the relation between the vector model and the unfolding model for preferences. *Psychometrika*, 40:115–116.

de Leeuw, J. and Mair, P. (2009). Multidimensional scaling using majorization: SMACOF in R. *Journal of Statistical Software*, 31(3):1–30.

Elo, A. (1978). *The Rating of Chessplayers, Past and Present*. Batsford, London.

Elo, A. (1961). New USCF rating system. *Chess Life*, 16:160–161.

Fechner, Gustav Theodor (H.E. Adler, T. (1966/1860). *Elemente der Psychophysik (Elements of psychophysics)*. Breitkopf & Hartel., Leipzig.

Guilford, J. P. (1954). *Psychometric Methods*. McGraw-Hill, New York, 2nd edition.

Joe, H. (1991). Rating systems based on paired comparison models. *Statistics & Probability Letters*, 11(4):343 – 347.

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47:263–291.

Luce, R. D. (1959). *Individual choice behavior*. Wiley, New York.

Luce, R. D. (1977). The choice axiom after twenty years. *Journal of Mathematical Psychology*, 15(3):215–233.

Ozer, D. J. (1993). Classical psychophysics and the assessment of agreement and accuracy in judgments of personality. *Journal of Personality*, 61(4):739–767.

Rossi, G. B. (2007). Measurability. *Measurement*, 40(6):545 – 562.

Sinn, H. W. (2003). Weber's law and the biological evolution of risk preferences: The selective dominance of the logarithmic utility function, 2002 geneva risk lecture. *Geneva Papers on Risk and Insurance Theory*, 28(2):87–100.

Stevens, S. (1946). On the theory of scales of measurement. *Science*, 103(2684):677–680.

Thorndike, E. L. (1904). *An introduction to theory of mental and social measurements*. The Science Press, New York.

Tukey, J. W. (1977). *Exploratory data analysis*. Addison-Wesley Pub. Co., Reading, Mass.

Weber, E. H. (1834b). *De pulsu, resorptione, auditu et tactu. Annotationes anatomicae et physiologicae*. Kohler, Leipzig.

Weber, E. H. (1948/1834a). Concerning touch, 1834. In Dennis, W., editor, *Readings in the history of psychology*, pages 155–156. Appleton-Century-Crofts Appleton-Century-Crofts Print, East Norwalk, CT.