Chapter 2
A Theory of Data

At first glance, it would seem that we have an infinite way to collect data. Measuring the diameter of the earth by finding the distance to the horizon, measuring the height of waves produced by a nuclear blast by nailing (empty) beer cans to a palm tree, or finding Avogadro's number by dropping oil into water are techniques that do not require great sophistication in the theory of measurement. In psychology we can use self report, peer ratings, reaction times, psychophysiological measures such as the Electric Encephelagram (EEG), the basal level of Skin Conductance (SC), or the Galvanic Skin Response (GSR). We can measure the number of voxels showing activation greater than some threshold in a functional Magnetic Resonance Image (fMRI), or we can measure life time risk of cancer, length of life, risk of mortality, etc. Indeed, the basic forms of data we can collect probably are unlimited. But in fact, it is possible to organize these disparate forms of data in terms of an abstract organization in terms of what is being measured and in comparison to what.

2.1 A Theory of Data: Objects, People, and Comparisons

Consider the following numbers and try to assign meaning to them: (2.718282), (3.1415929), (24), (37), (98.7), (365.256363051), (3,413), (86,400), (31,557,600), (299,792,458), and $(6.0221415 \pm 0.0000010) \times 10^{23}$ [1]. Because a number by itself is meaningless, all measures reflect a comparison between at least two elements. Without knowing the units of measure it is difficult to recognize that 37 °C and 98.7 °F represent the same average body tempature, that 24 (hours) and 86,400 (seconds) both represent one day, and that 365.256 (days) is the same length of time as 31,557,600 (seconds).

The comparison can be one of order (which is more) or one of proximity (are the numbers the same or almost the same?). Given a set of Objects (O) and a set of People (P) Clyde Coombs organized the types of measures one can take in a 2 x 2 x 2 table of the possible combinations of three distinct dimensions (Coombs, 1964). The elements may be drawn from the set of People (P), the set of Objects (O), or the Cartesian Cross Products of the sets of

[1] e, *pi*, hours in day, average body temperature in °C, average body temperature in °F, days in a sidereal year, BTUs/KWH, seconds in a day, seconds in a year, speed of light in meters/sec, number of atoms in a mole (Avogadro's Number).

People (P x P), Objects (OxO), or People by Objects (PxO). Furthermore, we can compare the results of these comparisons by forming either single dyads or pairs of dyads.[2]

That is, if we let O_j refer to the jth object and P_i to the ith person, then we can compare O_j to O_k, P_i to P_k, P_i to O_j , and so on.

1. The type of comparison made can either be one of *order* (is $P_i < P_k$) or of *distance* (if $\delta = |P_i - P_k|$, then is $\delta < X$)?
2. The elements being compared may be People, Objects, or People x Objects.
3. The number of dyads may be either one or more.

2.1.1 Modeling the comparison process

The two types of comparisons considered by Coombs were an order relationship and a proximity relationship. Given three objects on a line, we say that if A is greater than B if A- B > 0. Similarly, B is greater than C if B-C > 0. Without error, if A, B and C are on a line, if A > B and B > C, then A > C. With error we say that

$$p(A > B|A,B) = f(A - B). \tag{2.1}$$

Alternatively, A is close to B if the absolute difference between them, is less than some threshold, δ.

$$p(|A - B| < \delta|A,B,\delta) = f(|A - B|,\delta). \tag{2.2}$$

This distinction may be seen graphically by considering the probability of being greater as a function of the distance A -B (Fig 2.1) or the absolute difference between A and B.[3]

By using these three dimensions, it is possible to categorize the kind of data that we collect (Table 2.1).

2.2 Models and model fitting

For all of the following examples of estimating scale values it is important to ask how well do the estimated scale values recreate the data from which they were derived. Good scale values for the objects or for the people should provide a better fit to the data than do bad scale values. That is, given a data matrix, D, with elements d_{ij}, we are trying to find model values, m_i and m_j such that some function, f, when applied to the model values, best recreates d_{ij}. For data that are expressed as probabilities of an outcome, the model should provide a rule for comparing multiple scale values that are not necessarily bounded 0-1 with output values that are bounded 0-1. That is, we are interested in a mapping function f such that for any values of m_i and m_j

$$0 \leq f(m_i, m_j) \leq 1 \tag{2.3}$$

[2] This taxonomy can be generalized if we consider a third component of measurement: when is the measurement taken. We will consider the implications of a three dimensional organization in terms of Cattell's Databox (Cattell, 1966a) in chapter **??**

[3] The R-code for this and subsequent figures is included in Appendix-H

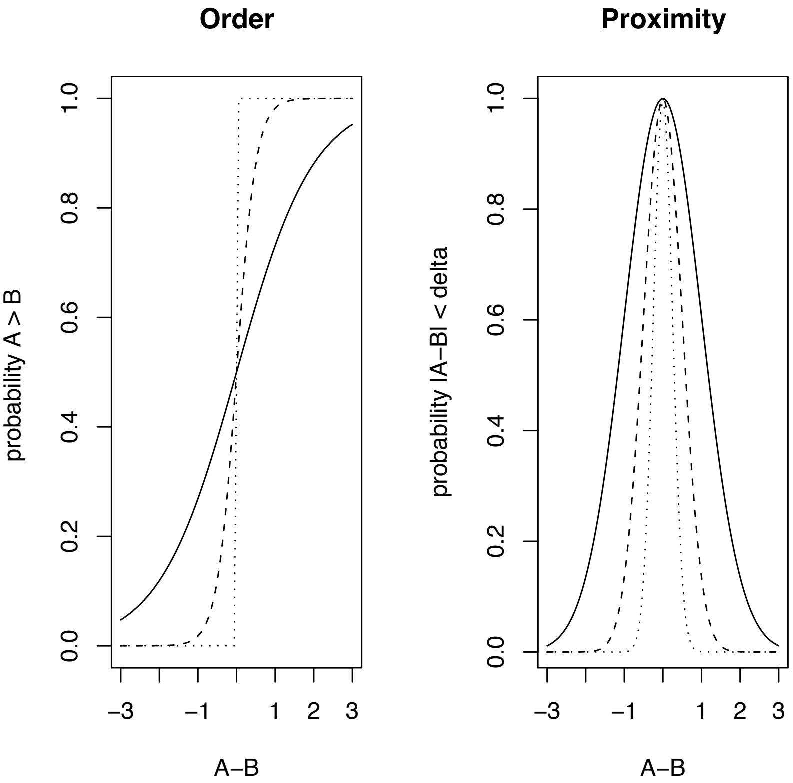

Fig. 2.1 Left panel: The probability of observing A > B as a function of the difference between A and B. The greater the signed difference, the greater the probability that A will be reported as greater than B. The three lines represent three different amounts of sensitivity to distance. Right Panel: The probability of observing A is the same as (close to) B as a function of the difference between A and B. The less the absolute difference, the greater the probability they will be reported as the same.

Although there are any number of such functions, there are at least two conventional ones that have such a property, one is the inverse normal transformation (where p values are mapped into the corresponding z values of a cumulative normal distribution), the other is the inverse logistic function (where p values are mapped onto the corresponding values of the logistic function). Both of these mappings satisfy the requirements of Equation 2.3 for any values of x and y.

Remembering Equations 1.1 and 1.2, we need to find scale values that minimize some function of the error. Applying $f(m_i,m_j)$ for all values of i and j produces the model matrix M. Let the error matrix E = D - M. Because average error will tend to be zero no matter how badly the model fits, median absolute error or average squared error are typical estimates of the amount of error. But such estimates are essentially "badness of fit" indices; goodness of fit indices tend to be 1 - badness. Both goodness or badness estimates should somehow reflect the size of the error with respect to the original data matrix. Thus, a generic estimate of goodness of fit becomes

Table 2.1 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	Chapter
People x People	1	Order	Tournament rankings	Theory of Data 2.4
People x People	1	Proximity	Social Networks	Theory of Data 2.5
Objects x Objects	1	Order	Scaling	Thurstonian scaling 2.6.2
Objects x Objects	1	Proximity	Similarities	Multidimensional scaling 2.6.2
People x Objects	1	Order	Ability Measurement	Test Theory 6, 7.1
People x Objects	1	Proximity	Attitude Measurement	Attitudes 7.3.2
People x People	2	Order	Tournament rankings	
People x People	2	Proximity	Social Networks	Theory of Data 2.5
Objects x Objects	2	Order	Scaling	Theory of Data 2.7
Objects x Objects	2	Proximity	Multidimensional scaling Individual Differences in MDS	Theory of Data 2.7
People x Objects	2	Order	Ability Comparisons	
People x Objects	2	Proximity	Preferential Choice	Unfolding Theory 2.8

$$GF = f(Data, Model) \tag{2.4}$$

Variations on this generic goodness of fit estimate include Ordinary Least Squares Estimates such as

$$GF = \frac{(Data - Model)^2}{Data^2} \tag{2.5}$$

as well as measures of median absolute deviation from the median, or many variations on Maximum Likelihood Estimates of χ^2.

2.3 A brief diversion: Functions in R

The examples in the rest of this (and subsequent) chapter(s) are created and analyzed using small example snippets of R code. For the reader interested just in psychometrics, these snippets can be ignored and the text, tables, and figures should suffice. However, reading the brief pieces of code and trying to run them line by line or section by section will help the reader learn how to use R. Even if you choose not to run the R-code while reading the text, by reading the R, some familiarity with R syntax will be gained.

As discussed in much more detail in Appendix A, R is a `function` driven language. Almost all operations invoke a `function`, usually by passing in some values, and then taking the output of the `function` as an object to be used in a later `function`. All functions have the form of function(parameter list). Some parameter lists are empty. Most functions have names that are directly understandable, at least by context. To see how a function works, entering the name of the function without the parentheses will usually show the function, although some functions are invisible or hidden in namespaces. A list of all functions used in the text is in Appendix B. Table B.1 briefly describes all the functions used in this chapter.

Programming in R can be done by creating new `functions` made up of other `functions`. **Packages** are merely combinations of these new functions that are found to be useful. The

psych package is a collection of functions that have proven to be useful in doing psychometrics. The functions within the package can be examined by entering the function name without the parentheses. For simplicity of reading the text, if more than a few lines of code are needed for the example, the R code will be included in appendix C rather than the text.

To obtain more information about any function, using the help function (?) provides a definition of the function, the various options possible when calling it, and examples of how to use the function. Studying the help pages will usually be enough to understand how to use the function. Perhaps the biggest problem is remembering the amazing number of functions that are available. Various web pages devoted to just listing the most used functions are common (see, e.g., *R/Rpad reference card* at `http://www.rpad.org/Rpad/Rpad-refcard.pdf`).

2.4 Tournaments: Ordering people ($p_i > p_j$)

The most basic datum is probably comparing one person to another in terms of a direct order relationship. This may be some sort of competition. Say we are interested in chess playing skill and we have 16 people play everyone else in a series of round robin matches. This leads to matrix of wins and losses. In Table 2.2, let a 1 in a cell mean that the row person beat the column person. NAs are put on the diagonal since people do not play themselves. The data were created by first creating a P vector of abilities (ranging from -1.5 to 1.5) and then predicting the probability of a P_i beating P_j as a logistic function of the difference between the two Ps. The logistic function (eq 2.6) is used as one of many possible ways to create a matrix of wins and losses. A function to find a matrix of inner or outer sums of two matrices, %+%, from the **psych** package is used for convenience[4]. Given the matrix of probabilities, (prob), random numbers are drawn from a rectangular distribution and if the probability exceeds that number, the person was said to win the match[5]. Because A beating B implies B loses to A, elements below the diagonal of the matrix are just 1 - those above the diagonal.

$$Prob(win|P_i,P_j) = Prob(P_i > P_j|P_i,P_j) = \frac{1}{1+e^{(P_j-P_i)}} \tag{2.6}$$

move to an appendix

```
> set.seed(42)       # to reproduce the same sequence each time
> p <- seq(-1.5, 1.5, 0.2) #make a sequence from low to high with step size
>  n <- length(p)          # how long is the vector
> pdif <- p %+% -t(p)     #find the matrix of differences between two vectors
> prob <- 1/(1 + exp(pdif))  #convert to probability values using a logistic function
> tournament <- (matrix(runif(n * n), nrow = n) > prob) + 0 #win with this probability
> win <- lower.tri(tournament)* tournament    #upper off diagonal matrix
> tournament <- lower.tri(tournament)* tournament + upper.tri(tournament)*(1-t(win))
> diag(tournament) <- NA      #the diagonal of the matrix is missing
> colnames(tournament) <- rownames(tournament) <- c(paste("P", 1:n, sep = ""))
> tournament             #show the data
```

[4] Packages are explained in more detail in Appendix-A

[5] Computerized Random Number Generators start with a "seed" and produce a sequence of pseudo random numbers based upon a particular generator function. In order to allow for the reader to be able to reproduce the examples given, in many of the examples the initial seed is set to a particular memorable value Adams (1980). The default setting for the seed is a function of the computer clock.

(See Table - 2.2).

Table 2.2 Simulated wins and losses for 16 chess players. Entries reflect row beating column.

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

2.4.1 Scaling of People

There are multiple ways to score these results. One is the simple average score (found by taking the `rowMeans`), a second is convert the winning percentage to a normal score equivalent, the third is to convert the winning percentage to a logistic equivalent. It is easy to do all three scoring procedures and to graphically compare them. This is done using the standard `data.frame` function and the `pairs.panels` function from the **psych** package.

```
> score <-  rowMeans(tournament, na.rm = TRUE)
> qscore <- qnorm(score)
> logit <-  log(score/(1 - score))
> chess.df  <- data.frame(latent = p, observed = score, normed = qscore, logit)
```

Just assigning numbers is not enough, for it is important to evaluate how well the assigned numbers capture the data. This requires a model of how to combine the rankings to predict the outcome. The average percent wins would seem reasonable until we consider how to combine them. A simple difference would not work, for that could lead to values outside of the range. In analogy to the axioms of choice (Bradley and Terry (1952), Luce (1977),) we could predict that the probability of A beating B (p(A >B) is the ratio of the frequency of A winning divided by the sum of the frequencies that A or B wins:

$$p(A > B|A,B) = \frac{p(A)}{p(A) + p(B)}. \tag{2.7}$$

For the normal deviate scores, a natural model would be to find the probability that A > B by finding the cumulative normal value of the normal-score difference:

Table 2.3 Three alternative solutions to the chess problem of Table 2.2

```
    latent observed normed logit
P1    -1.5     0.13  -1.11 -1.87
P2    -1.3     0.20  -0.84 -1.39
P3    -1.1     0.27  -0.62 -1.01
P4    -0.9     0.40  -0.25 -0.41
P5    -0.7     0.60   0.25  0.41
P6    -0.5     0.47  -0.08 -0.13
P7    -0.3     0.33  -0.43 -0.69
P8    -0.1     0.27  -0.62 -1.01
P9     0.1     0.47  -0.08 -0.13
P10    0.3     0.60   0.25  0.41
P11    0.5     0.67   0.43  0.69
P12    0.7     0.80   0.84  1.39
P13    0.9     0.60   0.25  0.41
P14    1.1     0.53   0.08  0.13
P15    1.3     0.87   1.11  1.87
P16    1.5     0.80   0.84  1.39
```

$$p(A > B|A,B) = pnorm(normal_A - normal_B) \tag{2.8}$$

For the logistic scores, the natural model would be to find the probability based upon the difference of the logistic scores:

$$p(A > B|A,B) = \frac{1}{1 + e^{(logit_B - logit_A)}} \tag{2.9}$$

For the observed scores, any of these would probably make equally good sense. The following function `model.fit` can be used to find the goodness of fit for the choice (eq 2.7), normal (eq 2.8), or logistic (eq 2.9) model. The function introduces the use of a *list* to hold multiple returned values. Note that the choice model will not work for negative scale values, and thus when applying it to normal or logistic modeled data, it is necessary to add in the minimum value.

move to appendix or include in the psych package

```
> model.fit <- function(model, data, test="logit", digits = 2,rowwise=FALSE) {
+     model <- as.matrix(model)
+     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
+      diag(error) <- NA
+      sum.error2 <- sum(error^2, na.rm = TRUE)
```

```
> pairs.panels(chess.df)
```

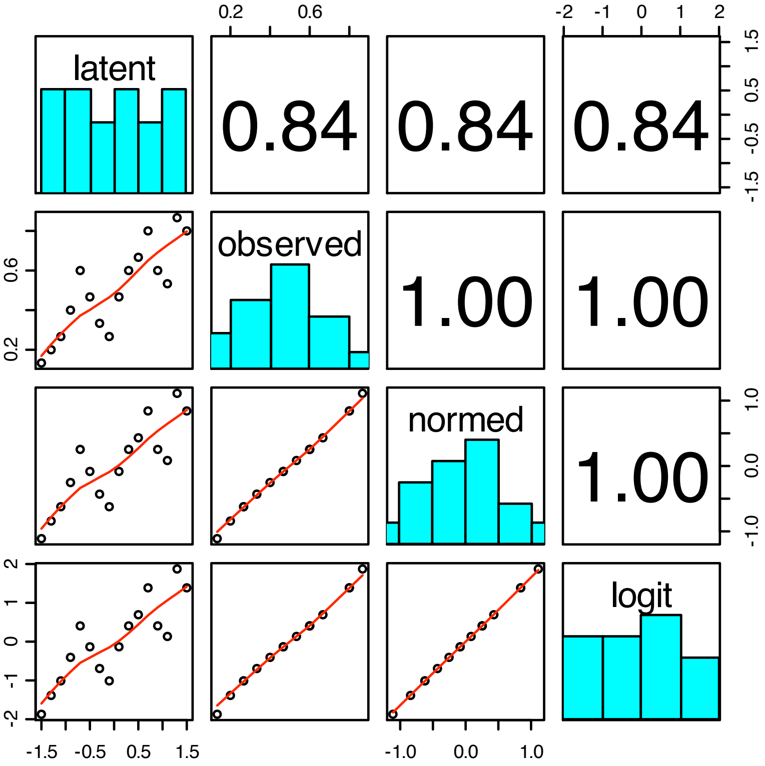

Fig. 2.2 Original model and three alternative scoring systems. This is also an example of a SPLOM (scatter plot matrix) using the pairs.panels function from the psych package.

```
+       sum.data2 <- sum(data^2, na.rm = TRUE)
+       gof <- 1 - sum.error2/sum.data2
+      fit <- list(fit = gof, original = sum.data2, resid = sum.error2,
+      residual = round(error,
+           digits))
+      return(fit)  }
```

Applying the model.fit function to the three scaling solutions found in chess.df yields 9 different estimates of goodness of fit (choice, logistic and normal models for each of the three scoring systems of the basic data set). The output of any R function is an object that can be used as input to any other function. We take advantage of this to make repeated calls to the model.fit function and collect the output in a matrix (fits).

probably remove this one as well

```
> tests <- c("choice","logit","normal")
>   fits <- matrix(rep(0, 9), ncol = 3)
>  for (i in 1:3) {
```

```
+      for (j in 1:3) {
+          fits[i, j] <- model.fit(chess.df[i + 1], data = tournament,
+          test = tests[j])$fit   } }
>  rownames(fits) <- c("observed", "normed", "logistic")
>  colnames(fits) <- c("choice", "logistic", "normal")
>  round(fits, 2)
```

```
         choice logistic normal
observed   0.64     0.58   0.61
normed     0.64     0.66   0.67
logistic   0.64     0.67   0.66
```

The goodness of fit tests suggests that both the normal and logistic procedures provide a somewhat better fit of the data than does merely counting the percentage of wins.

2.4.2 Alternative approaches to scaling people

The previous example of a tournament forced all players to play each other. Although this can be done with small groups, it is not feasible for larger groups. Tournament play for large sets of players (or teams) needs to use alternative measurement models. The NCAA basketball tournament of 65 teams is a well known alternative in which teams are eliminated after loses. This allows a choice of an overall winner, but does not allow for precise rankings below that. In addition, it is important to note in the simulated example (Table 2.2), that a better player (P16) was defeated by a weaker player (P5), even though P16 had a much higher winning percentage (80%) than did P5 (60%). In addition, in this 16 player match, the observed rankings were correlated only .84 with the underlying ability (the latent score) used to generate the data. Thus, neither a sudden death tournament nor a round robin tournament necessarily leads to identifying the strongest over all player or team.

Using a system developed by Arpad Elo (and hence called the Elo scale) chess players are ranked on a logistic scale where two players with equal scores have a 50% probability of winning and a player 200 points higher than another would win 75% of the time. The Elo system does not require all players to play each other, but rather adds and subtracts points for each player following a match. Beating a better player adds more points than does beating a weaker player, and losing to a player with a lower ranking subtracts more points than losing to a player with a higher ranking. Revisions have been suggested by Mark Glickman and others.

Glickman and his colleagues haves applied logistic rating to the problem of rankings of colleges **?** as well as sports teams. Using the pattern of student choice as analogous to direct competition (School A beats B if student X chooses to attend A rather than B), and then scaling the schools using the logistic model provides a better metric for college rankings than other procedures. (As Glickman and his colleagues point out, selectivity ratings can be "gamed" by encouraging many weaker students to apply and then rejecting them, and yield ratings can be inflated by rejecting students who are more likely to go somewhere else.)

2.4.3 Assigning numbers to people – the problem of rewarding merit

The previous examples considered that people are in direct competition with each other and are evaluated on the same metric. This is not just an abstract problem for evaluating chess players or basketball teams, but is a real life problem when it comes to assigning salaries and raises. But it is not just a problem of evaluating merit, it is also a problem in how to treat equal merit. Consider a department chairman with $100,000 of merit raises to distribute to 20 faculty members who have an average salary of $100,000. That is, the salary increment available is 5% of the total salary currently paid. Assume that the current range of salaries reflects differences in career stage and that the range of current salaries is uniformly distributed from $50,000 to $150,000. Furthermore, make the "Lake Wobegon" assumption that all faculty members are above average and all deserve an "equal" raise. What is the correct average? Is it $5,000 per faculty member ($100,000/20) or is 5% for each faculty member (with the lecturers getting $2,500 and the full professors getting three times as much, or $7,500)? This problem is considered in somewhat more detail when comparing types of scales in section 3.15

elaborate the distinction between linear, market, and Weber law

2.5 Social Networks: Proximities of People ($|p_i - p_j| < \delta$)

An alternative to ordering people is to ask how close two people are to each other. This can be done either for all possible pairings of people, or for a limited set of targets. In both cases, the questions are the same: Are two people closer than some threshold, X: if $\delta = |P_i - P_j|$, then is $\delta < X$? This very abstract representation allows us to consider how well known or liked or desirable someone is, depending upon the way we phrase the question to person i:

1. Do you know person j?
2. Do you like person j? or as an alternative:
3. Please list all your friends in this class (and is j included on the list)
4. Would you be interested in having a date with person j?
5. Would you like to have sex with person j?
6. Would you marry person j?

2.5.1 Rectangular data arrays of similarity

If we ask a large group of people about a smaller set (perhaps of size one) of people, we will form a rectangular array of proximities. For example, evolutionary psychologists have used responses to items 4-6 asked by an attractive stranger (person j) to show strong sex differences in interest in casual sex Buss and Schmitt (1993). Typical data might look like Table 2.4. Sociologists might use questions 1 - 3 to exam social networks in classrooms.

Note that the data will not necessarily, and in fact probably will not, be symmetric. For more people know (or know of) Barack Obama than he could possibly know.

Table 2.4 A hypothetical response matrix for questions 4-6 about social interaction with an attractive stranger.

Person	Gender	Item 4	Item 5	Item 6
1	F	0	0	0
2	F	1	0	0
3	F	1	1	0
...				
98	M	1	1	0
99	M	1	1	1
100	M	1	1	1

2.5.2 Square arrays of similarity

Another example of proximity data for pairs of people would be the results of "speed dating" studies Finkel et al. (2007). Given N people, each person spends a few minutes talking to each other person in the room. After each brief conversation each person is asked whether they want to see the other person again. (Abstractly, the assumption is that the smaller the distance, δ, between two people, the more a person would want to see the other person). If $\delta < X$ the person responds yes.) Here, although the matrix is square (everyone is compared with everyone else, the proximities are not symmetric, for some people are liked (close to) more people than others.

We simulate the data using an "experimental design" where each of 10 males interact for 3 minutes with each of 10 females, and vice versa. After each interaction both members of the pair were asked whether they wanted to see the other person again. To simulate an example of such data we create a 20 x 20 array of person "interest" by randomly sampling with replacement from the numbers 0 and 1. Let the rows represent our participants and the columns the expression of interest they have in the other participants. The first 10 participants are females, the second 10 males (Table 2.5).

```
> set.seed(42)
> prox <- matrix(rep(NA, 400), ncol = 20)
> prox[11:20, 1:10] <- matrix(sample(2, 100, replace = TRUE) - 1)
> prox[1:10, 11:20] <- matrix(sample(2, 100, replace = TRUE) - 1)
> colnames(prox) <- rownames(prox) <- c(paste("F", 1:10, sep = ""),
+ paste("M", 1:10, sep = ""))
> prox
```

Because of the experimental design, the data matrix has missing values for same sex pairs. We find the row and column means but specify that we want to not include the missing values.

```
> colMeans(prox, na.rm = TRUE)
```

```
 F1  F2  F3  F4  F5  F6  F7  F8  F9 F10  M1  M2  M3  M4  M5  M6  M7  M8  M9 M10
0.8 0.5 0.6 0.6 0.6 0.4 0.7 0.2 0.4 0.7 0.6 0.7 0.6 0.9 0.4 0.5 0.4 0.5 0.6 0.4
```

```
> rowMeans(prox, na.rm = TRUE)
```

Table 2.5 Hypothetical results from a speed dating study.

	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10
F1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	1	1	0	1	0	1	1	1	1	0
F2	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	0	1	0	1	0	0	1	0	1	1
F3	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	0	1	1	1	0	1	1	0	0	0
F4	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	0	0	1	1	1	0	0	0	0	0
F5	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	1	1	1	1	1	1	1	0	1	0
F6	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	1	1	0	1	1	1	0	0	1	0
F7	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	1	0	1	1	0	0	0	1	1	1
F8	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	1	0	1	0	1	0	0	1	1	0
F9	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	1	1	0	1	0	0	0	1	0	1
F10	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	0	1	1	1	0	1	0	1	0	1
M1	1	0	1	1	0	0	1	0	1	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
M2	1	1	0	1	0	0	1	0	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
M3	0	1	1	0	0	0	1	0	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
M4	1	0	1	1	1	1	1	0	1	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
M5	1	0	0	0	0	0	1	0	1	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
M6	1	1	1	1	1	1	0	1	1	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
M7	1	1	0	0	1	1	0	0	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
M8	0	0	1	0	1	0	1	0	0	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
M9	1	0	0	1	1	0	1	1	0	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
M10	1	1	1	1	1	1	0	0	0	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

```
 F1  F2  F3  F4  F5  F6  F7  F8  F9 F10  M1  M2  M3  M4  M5  M6  M7  M8  M9 M10
0.7 0.5 0.5 0.3 0.8 0.6 0.6 0.5 0.5 0.6 0.6 0.4 0.3 0.8 0.4 0.9 0.4 0.4 0.6 0.7
```

Given the data matrix, prox, the column means represent how much each person was liked (participants F8 and M3 seem to be the least popular, and participant F1 and M4 the most popular). The row means represent differences in how much people liked others with participants F5 and M6 liking the most people, and participants F4 and M3 liking the least number. What is not known from the simple assignment of average numbers is whether this is the appropriate metric. That is, is the difference between having 50% of the people like you versus 60% the same as the difference between 80% and 90%?

The Social Network Analysis, **sna**, package in R allows for a variety of statistical and graphical analyses of similarity matrices. One such analysis is the comparison of different networks. Another is the ability to graph social networks (Figure 2.3).

2.6 The Scaling of Objects ($o_i < o_j$)

When judging the value of a particular object and a specific dimension, we can form scales based upon the effect the object has on similar objects. This technique has long been used in the physical sciences. For example, Friedrich Mohs' (1773-1839) scale of hardness is an ordinal scale based upon direct comparisons of minerals . If one mineral will scratch another, we say the first is harder than the second. Hardness is transitive, in that if X scratches Y and Y scratches Z, then X will scratch Z (Table 2.6). The Moh's hardness scale is an ordinal scale and does not reflect that the differences between the 10 minerals can be measured more precisely in terms of the amount of force required to make the scratch using a diamond tip

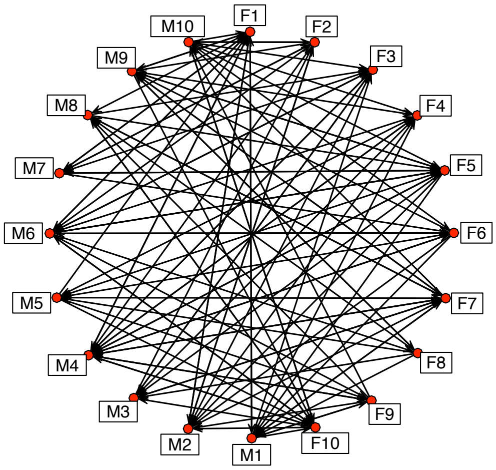

Fig. 2.3 Social networks analysis of the data from Table 2.5

and a sclerometer. Even more precise measures may be made by measuring the size of an indentation made by diamond tip under various levels of pressure **?**. When this is done the Mohs scale can be converted from an ordinal to a relative hardness scale.

check absolute hardness values

Table 2.6 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 the indentation test.

Mohs Hardness	Mineral	Indentation hardness
1	Talc	1
2	Gypsum	3
3	Calcite	9
4	Fluorite	21
5	Apaptite	48
6	Orthoclase Feldspar	72
7	Quartz	100
8	Topaz	200
9	Corundum	400
10	Diamond	1500

Another way of ordering objects is in terms of their effects upon the external environment. For sailors, it is important to be able to judge wind conditions by observation. Although the effect of wind varies as the square of the velocity, a roughly linear metric of wind speed based upon observed sea state was developed by Sir Francis Beaufort (1774-1857) and is still in common use among windsurfers and sailors. Beaufort's original scale was in terms of how a British frigate would handle and what sails she could carry but was later revised in terms of observations of sea state. Beaufort did not classify wind speed in terms of velocity and these estimated equivalents as well as the current descriptions have been added by meteorologists (Table 2.7). What is most important to notice is that because of the non-linear (squared) effect of wind velocity on sailors, equal changes in the Beaufort scale (e.g., from 1 to 2 or from 4 to 5) do not lead to equal changes in such important outcomes as the probability of capsizing!

Table 2.7 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

Wave length, periodicity, and energy

2.6.1 Weber-Fechner scales of subjective experience

get a better Fechner cite

Early studies of psychophysics by Weber (1834, 1948) and subsequently Fechner et al. (1997) demonstrated that the human perceptual system does not perceive stimulus intensity as a linear function of the physical input. 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. What was the Δ that was just detectable? The finding was that the perceived intensity follows a logarithmic function. Examining the magnitude of the "*just noticeable differece*" or *JND*, Weber (1834) found that

$$JND = \frac{\Delta Intensity}{Intensity} = constant \tag{2.10}$$

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

$$L_p = 20Log_{10}\frac{P_o}{P_{ref}} \tag{2.11}$$

where the reference pressure,P_{ref}, in the air is $20\mu Pa$. 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.

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. 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.

> ... *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).

2.6.2 Thurstonian Scalilng

Louis L. Thurstone was a pioneer in psychometric theory and measurement of attitudes, interests, and abilities. Among his many contributions was a systematic analysis of the process of comparative judgment (Thurstone, 1927). He considered the case of asking subjects to successively compare pairs of objects. If the same subject does this repeatedly, or if subjects act as random replicates of each other, their judgments can be thought of as sampled from a normal distribution of underlying (latent) scale scores for each object, Thurstone proposed that the comparison between the value of two objects could be represented as representing the differences of the average value for each object compared to the standard deviation of the differences between objects. The basic model is that each item has a normal distribution of response strength and that choice represents the stronger of the two response strengths (Figure-2.4). A justification for the normality assumption is that each decision represents the sum of many independent inputs and thus, through the central limit theorem, is normally distributed.

Thurstone considered five different sets of assumptions about the equality and independence of the variances for each item (Thurstone, 1927). Torgerson expanded this analysis slightly by considering three classes of data collection (with individuals, between individuals and mixes of within and between) crossed with three sets of assumptions (equal covariance of decision process, equal correlations and small differences in variance, equal variances) (Torgerson, 1958).

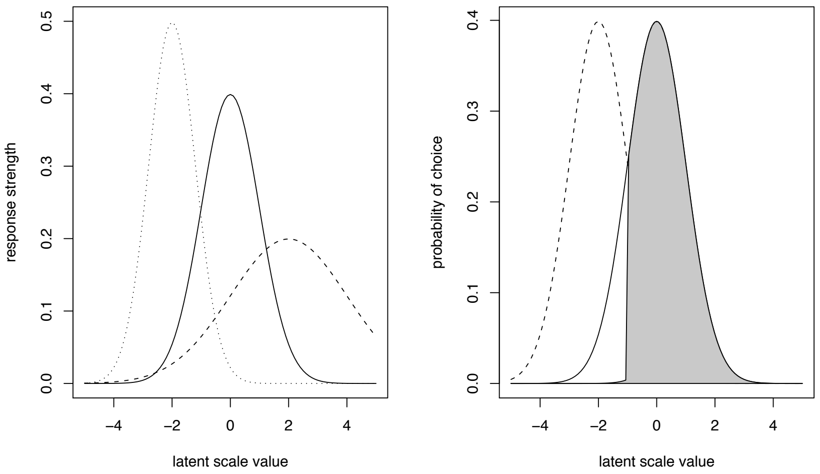

Fig. 2.4 Thurstone's model of paired discrimination. Left panel: three items differ in their mean level as well as their variance. Right panel: choice between two items with equal variance reflects the relative strength of the two items. The shaded section represents choosing item 2 over item 1.

Thurstone scaling has been used in a variety of contexts, from scaling the severity of crimes (Coombs, 1964) to the severity of cancer symptoms to help nurses understand their patients (Degner et al., 1998) or in market research to scale alternative products (Green et al., 1989). Consider an example of scaling of vegetables, discussed in great detail in Guilford (1954). Participants were asked whether they preferred vegetable A to vegetable B for a set of nine vegetables (Turnips, Cabbage, Beets, Asparagus, Carrots, Spinach, String Beans, Peas, and Corn). This produced the following data matrix where the numbers represent the percentage of time that the column vegetable was chosen over the row vegetable Guilford (1954).

Table 2.8 Paired comparisons of nine vegetables (from Guilford, 1954). The numbers represent the probability with which the column vegetable is chosen over the row vegetable.

	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

Just as when scaling individuals (section-2.4), there are several natural ways to convert these data into scale values. The easiest is simply to find the average preference for each item. To do this, copy the data table into the clipboard and then using the read.clipboard function (part of the **psych** package, create the new variable, veg. The read.clipboard function will create a data.frame, veg. Although a data.frame looks like a matrix, each separate column is able to be treated individually and may be of different types of data (e.g., characters, logical, or numeric). The mean function will report separate means for each variable in a data frame whereas it will report just the overall mean for a matrix. Were veg a matrix, mean would report just one value but colMeans would report the mean for each column.

```
> veg <- read.clipboard()
> round(mean(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
```

Given that the zero point is arbitrary, subtracting the Turnip value from other vegetables does not change anything:

```
> veg.t <- mean(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
```

If these values were useful, then it should be possible to recreate the rows of the original matrix (Table 2.8) by taking the differences between any two scale values + .5 (since a vegetable is preferred over itself 50% of the time, adding .5 to predicted choice). But this will produce values greater than 1 and less than 0! The predicted probability of choosing Corn over Turnips would be .5 + .54 = 1.04!. Clearly, this is not a good solution.

Thurstone's proposed solution was to assign scale values based upon the average normal deviate transformation of the raw probabilities. This is found by first converting all the observed probabilities into their corresponding standard normal values, *z-scores*, and then finding the average z-score. This is done using the qnorm, as.matrix, and colMeans functions. (Table 2.9).

Table 2.9 Convert the paired comparison data of Table 2.8 into z-scores.

	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

Finding the average z-score for each column of Table 2.9 is the equivalent to finding the least squares solution to the series of equations finding the pairwise distances (Torgerson, 1958). Adding a constant does not change the distances, so by subtracting the smallest to all the numbers we find the following scale values:

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

But how well does this set of scale values fit? One way to evaluate the fit is to find the *predicted* paired comparisons given the model and then subtract them from the observed. The model matrix is found by taking the differences of the row and column values for the items. For example, the modeled value for String Beans vs. Corn is 1.40 - 1.63 or -.23.

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

Table 2.10 Modeled z score differences based upon the modeled scale values

```
> 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
```

Then, convert these modeled values to probabilities (Table 2.11) and then find the residuals or errors (Table 2.12) by comparing to the original data (Table 2.8).

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


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

These residuals seem small. But how small is small? The mean residual is (as it should be) 0. Describing the residuals (using the `describe` function) suggests that they are indeed small:

Alternatively, using a goodness of fit test (e.g., Equation 2.5) finds that the sum squared residual of .31 is much less than the sum squared data, 24.86 for a Goodness of Fit of .994.

Table 2.11 Modeled probability of choice based upon the modeled scale values

	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

Table 2.12 Residuals = data - model

	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

Table 2.13 Basic summary statistics of the residuals suggest that they are very small

```
> describe(resid)
        var n  mean   sd median  mad   min  max range   se
Turn      1 9 -0.01 0.05   0.00 0.03 -0.12 0.03  0.15 0.02
Cab       2 9  0.00 0.05   0.00 0.02 -0.05 0.12  0.17 0.02
Beet      3 9  0.00 0.05   0.01 0.04 -0.07 0.07  0.14 0.02
Asp       4 9  0.00 0.04  -0.01 0.03 -0.07 0.08  0.14 0.01
Car       5 9  0.00 0.05   0.01 0.01 -0.08 0.06  0.14 0.02
Spin      6 9  0.00 0.03   0.00 0.03 -0.06 0.06  0.12 0.01
S.Beans   7 9  0.00 0.04   0.00 0.03 -0.05 0.07  0.12 0.01
Peas      8 9  0.00 0.05   0.01 0.06 -0.08 0.08  0.16 0.02
Corn      9 9  0.01 0.04  -0.01 0.02 -0.06 0.07  0.13 0.01
```

2.6.3 Alternative solutions to the ranking of objects

Just as in the scaling of people in tournaments there were alternative ways to assign scale values and to evaluate the scale values, so we can consider alternatives to the Thurstone scale. Although we can not take simple differences between scale values to predict choice, using Equation 2.7 or Equation 2.9 does allow for alternative solutions. Consider assigning a constant to all values, rank orders (1-9) or squared rank orders, or even reversed rank orders (just to be extreme)! Just as we can compare the three ways of scaling people from tournament outcomes, so can we compare using the Choice model, Thurstone Case V, or logistic models for fitting alternative scalings of the vegetable data.

First form a data frame made up of the six alternative scaling models (a constant, simple rank orders of the vegetable choices, squared rank orders, reversed rank orders, frequency of choice and Thurstonian scale values) and then apply the model.fit function from above to all six scales. What is most interesting is that although the scale values differ greatly, the Choice model fits almost as well for four of the six scaling solutions. . The Thurstone and logistic fitting techniques differ a great deal between the six methods of forming the scales.

```
> veg.scales.df <- data.frame(constant= rep(.5,9),equal= seq(1,9),squared = seq(1,9)^2,
+ reversed = seq(9,1),raw = veg.t,thurstone=scaled)
> round(veg.scales.df,2)
>  fits <- matrix(rep(0, 3*dim(veg.scales.df)[2]), ncol = 3)
>  for (i in 1:dim(veg.scales.df)[2]) {
+      for (j in 1:3) {
+          fits[i, j] <- model.fit(veg.scales.df[i],rowwise=TRUE, data = as.matrix(veg),
+          test = tests[j])$fit   } }
>  rownames(fits) <- c("Constant", "Equal","Squared", "Reversed","Choice", "Thurstone")
>  colnames(fits) <- c("choice", "logistic", "normal")
>  round(fits, 2)
```

```
        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
Choice      0.97     0.89   0.93
Thurstone   0.97     0.97   0.99
```

What should we conclude from this comparison? Not that the Thurstone techniques are useless, but rather that the scaling solutions need to be considered in comparison with alternative hypothetical solutions. That is, just because one procedure yields a very good fit and makes psychometric sense does not imply that it is necessarily better than a simpler procedure.

2.6.4 Why emphasize Thurstonian scaling?

Perhaps it seems as if too much emphasis has been given to the ordering of vegetables. However, the process of model fitting and model testing outlined in the previous section is similar to the process that needs to be followed in all analyses.

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

2.7 Multiple Dimensional Scaling: Distances between Objects ($|o_i - o_j| < |o_k - o_l|$)

The tournament rankings of chess players and the frequency of choice of vegetables (or the severity of crimes) are found by comparing two stimuli, either subjects (2.4) or objects (2.6.2), to each other. The social networks of friendship groups can be thought of as a representing distances less than a certain value ($|o_i - o_j| < \delta$) (2.5). But to compare these distances to other distances leads to ordering relationships of distances (($|o_i - o_j| < |o_k - o_l|$). The typical application is to order pairs of distances in a multidimensional space. One classic example is multidimensional scaling of distances based upon the Euclidian distance between two points, x and y, which, in an n-dimensional space, is

$$Distance_{xy} = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}. \tag{2.12}$$

Alternative scaling models attempt to fit a monotone function of distance rather than the Euclidian distance. There are a variety of metric and non-metric algorithms, the basic procedure of the non-metric procedures is fit a monotonically increasing function of distance (e.g., their ranks) rather than the distances themselves.

Table 2.14 Distances between 9 American cities.

	BOS	CHI	DC	DEN	LA	MIA	NY	SEA	SF
BOS	0	963	429	1949	2979	1504	206	2976	3095
CHI	963	0	671	996	2054	1329	802	2013	2142
DC	429	671	0	1616	2631	1075	233	2684	2799
DEN	1949	996	1616	0	1059	2037	1771	1307	1235
LA	2979	2054	2631	1059	0	2687	2786	1131	379
MIA	1504	1329	1075	2037	2687	0	1308	3273	3053
NY	206	802	233	1771	2786	1308	0	2815	2934
SEA	2976	2013	2684	1307	1131	3273	2815	0	808
SF	3095	2142	2799	1235	379	3053	2934	808	0

Consider the airline distances between nine american cities in Table 2.14. Even considering issues of the spherecity of the globe, it is not surprising that these can be arranged in a two dimensional space. Using the `cmdscale` function, and specifying a two dimensional solution finds the best fitting solution (values for all the cities on two dimensions):

Table 2.15 Two dimensional representation for 9 American cities.

```
> city.location <- cmdscale(cities, k=2)    #ask for a 2 dimensional solution
> round(city.location,0)
     [,1] [,2]
BOS -1349 -462
CHI  -428 -175
DC  -1077 -136
DEN   522   13
LA   1464  561
MIA -1227 1014
NY  -1199 -307
SEA  1596 -639
SF   1697  132

>plot(city.location,type="n", xlab="Dimension 1", ylab="Dimension 2",main ="cmdscale(cities)")
>text(city.location,labels=names(cities))     #put the cities into the map
```

Representing these cities graphically produces a rather strange figure fig 2.5) . Reversing both axes produces a figure that is more recognizable (fig 2.6.

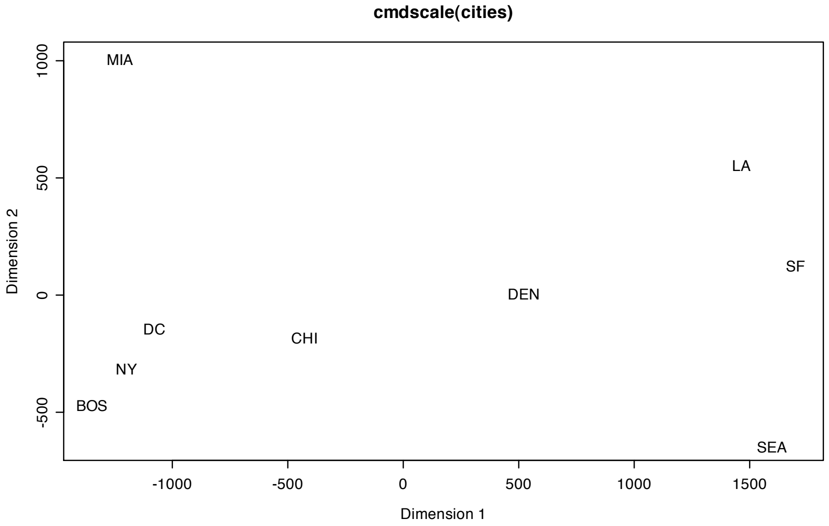

Fig. 2.5 Original solution for 9 US cities. What is wrong with this figure? Axes of solutions are not necessarily directly interpretable. Compare to Figure 2.6

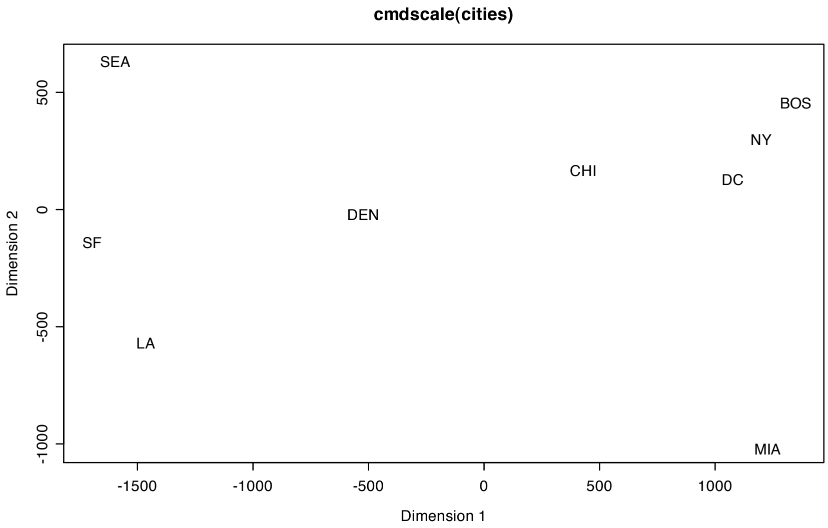

Fig. 2.6 Revised solution for 9 US cities after making city.location <- -city.location

Extensions of the metric multidimensional scaling procedures fit data where distances are ordinal rather than interval (e.g, Borg and Groenen (2005); Carroll and Arabie (1980); Green et al. (1989); Kruskal and Wish (1978)) are known as non-metric multidimensional scaling and are available in the **MASS** package as `isoMDS` and `sammon`. In addition, some programs are able to find the best fit for arbitrary values of r for distance in a Minkowski R space. An r value of 1 produces a "city block" or "Manhattan" metric, r of 2 is the standard Euclidean, and r values greater than 2 emphasize the larger distance much more than smaller distances.

$$Distance_{xy-r} = \sqrt[r]{\sum_{i=1}^{n} (x_i - y_i)^r}. \tag{2.13}$$

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

"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). 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. Although the comparisons are ordinal, there is a surprising amount of metric information in the analysis.

2.8.1 Individual Preferences – the I scale

When an individual is asked whether they prefer one object to another, the assumption is that the preferred object is closer (in an abstract, psychological space) to the person than is the non-preferred item. The person's location is known at his or her "ideal point" and the closer an object is to the ideal point, the more it is preferred. Consider the case of how many children someone would like to have. For the purpose of analysis we limit this to 0, 1, 2, 3, 4, or 5 children.

Suppose we ask each individual in a sample of people how many children they would like to have. It is likely that the first choices would range from 0 - 5. Then, except for those whose first choice was either 0 or 5, they are then asked "if you could not have that number, would you rather have one more or one less?" . For people whose second choice was neither 0 nor 5, this procedure is then continued with question #3: "If you could not have X (the first choice) or Y (the second choice), would you rather have (one less than the minimum of X and Y) or (one more than the maximum of X and Y)" where the questioner replaces X and Y with the values from the subject.

There are many possible Individual preference orderings (I-scales) that will be single peaked: For the person who prefers no children, the preference ordering is 012345, while for the person who would like a very large family the I-scale would be 54321. For someone whose ideal point is between 2 and 3, the orderings 231450, 234150, 231450, or 231045 are all possible.

Assign the scale value for no children to be 0 and for 5 children to be 100. Where on this scale are the values for 1, 2, 3 or 4 children? The naive answer is to assume equal spacing and give values of 20, 40, 60, and 80. But psychologically, is the difference between 0 and 1 the same as between 4 and 5? This can be determined for individual subjects by careful examination of their preferential orders.

Table 2.16 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>1 2 3 4</td><td>5</td><td>0</td><td>1 2 3 4</td><td>5</td></tr>
<tr><td>0</td><td>0|5</td><td>5</td><td>0</td><td>0|5</td><td>5</td></tr>
<tr><td>0</td><td>2|3</td><td>5</td><td>0</td><td>2|3</td><td>5</td></tr>
<tr><td>0</td><td>1|2</td><td>5</td><td>0</td><td>1|2</td><td>5</td></tr>
<tr><td>0</td><td>0|1 4|5</td><td>5</td><td>0</td><td>0|1 4|5</td><td>5</td></tr>
<tr><td>0</td><td>3|4</td><td>5</td><td>0</td><td>3|4</td><td>5</td></tr>
<tr><td>0</td><td>2|5</td><td>5</td><td>0</td><td>2|5</td><td>5</td></tr>
</table>

Consider the ordering 231450 versus the 321045 ordering. For the first person, because 5 is preferred to 0 we can say that the (0|5) midpoint has been crossed (the person is to the right of that midpoint). But the person prefers 2 to 3, and thus the (2|3) midpoint has not been crossed. This implies that the distance from 0 to 2 is greater than the distance from 3 to 5. The data from the second person, 321045, because the (2|3) midpoint has been crossed, but the (0|5) has not been imply that the distance 0 to 2 is less than the distance from 3-5. (See Table 2.16).

2.8.2 Joint Preferences – the J scale

When multiple preference orderings are examined, they can be *partially ordered* with respect to their implied midpoints (Figure 2.7). The highlighted I-scales reflect the hypothesis that for all subjects, the distance between progressive numbers of children is a deaccelerating distance.

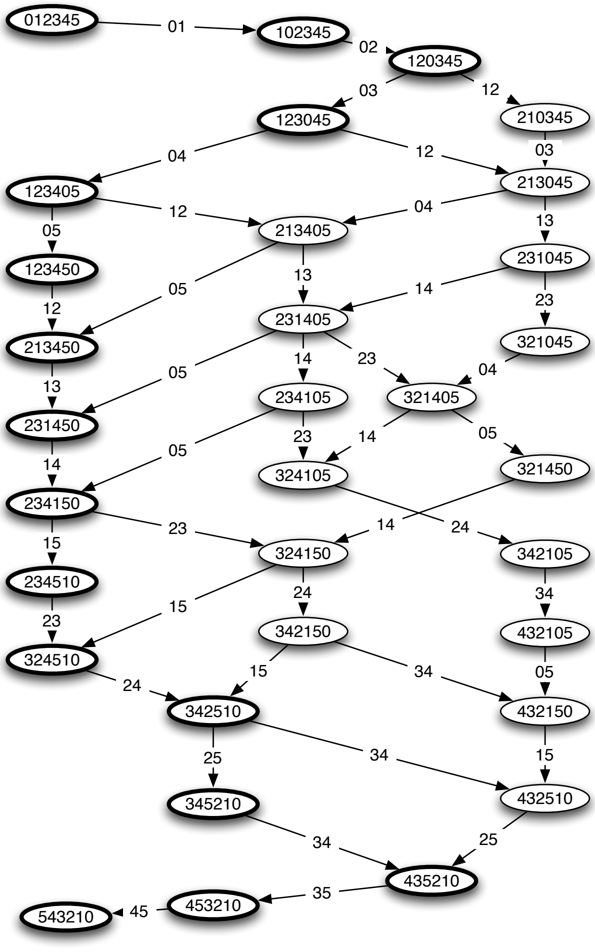

Fig. 2.7 Possible I-scales arranged to show ordering of mid-points. The highlighted I-scales reflect an deaccelerating Joint scale. The labels for each path show the midpoint “crossed” when going from the first I scale to the second I scale.

2.8.3 Partially ordered metrics

The ordering of the midpoints for the highlighted I-scales seen in Figure 2.7 allow distances in Joint scale to be partially ordered. The first and last two midpoints provide no information, for that order is fixed. But that (0|3), (0|4) and (0|5) come before (1|2), (1|3), and (1|4) gives a great deal of metric information (Table 2.17).

add a few more partial orders here

Table 2.17 Midpoint orders imply partial ordering of distances

Midpoint order	Distance information
$(0\|3) < (1\|2)$	$23 < 01$
$(0\|4) < (1\|2)$	$24 < 01$
$(0\|5) < (1\|2)$	$25 < 01$
$(0\|5) < (2\|3)$	$35 < 02$
$(2\|5) < (3\|4)$	$45 < 23$

2.8.4 Multidimensional Unfolding

Generalizations of Coombs' original unfolding to the multidimensional case, with both metric and non metric applications are discussed by de Leeuw (2005). The basic technique is to treat the problem as a multidimensional scaling problem of an off diagonal matrix (that is to say, objects by subjects).

2.9 Measurement of Attitudes and Abilities (comparing s_i, o_j)

The measurement of abilities and attitudes compares single items (objects) to single subjects. The comparison may be either one of order (for abilities) or one of proximity (for attitudes). The difference in these two models may be seen in Figures 2.9 and 2.10. Although most personality inventories are constructed using the abilities model, it has been pointed out that the ideal point (proximity) model is probably more appropriate Chernyshenko et al. (2007).

In the discussion of classic (Chapter 6) and modern test theory (Chapter 7), the implications of the ordering model will be discussed in detail. Here it is just discussed in terms of the Coombs' models.

2.9.1 Measurement of abilities ($s_i > o_j$)

The basic model is that for ability = θ and difficulty = δ that

$$prob(correct|\theta,\delta) = f(\theta - \delta) \tag{2.14}$$

2.9.1.1 Guttman scales

Guttman considered the case where there is no error in the assessment of the item difficulty or if the items are sufficiently far apart so that the pattern of item response is completely redundant with the total score Guttman (1950). That is,

$$prob(correct|\theta,\delta) = 1|\theta > \delta \\ 0|\theta < \delta.$$

One example of items which can be formed into a Guttman scale are those from the social distance inventory developed by Bogardus in 1925 to assess "social distance" which "refers to the degrees and grades of understanding and feeling that persons experience regarding each other. It explains the nature of a great deal of their interaction. It charts the character of social relations." Bogardus (1925)).

Table 2.18 The Bogardus Social Distance Scale is one example of items that can be made to a Guttman scale

The Bogardus social distance scale gave the following stem with a list of various ethnicities. "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

Such items (with the omission of item 1 and a rewording of item 2) will typically produce a data matrix similar to that in table 2.9.1.1. The scaling is redundant in that for perfect data the total number of items endorsed always matches the highest item endorsed. With the exception of a few examples such as social distance or sexual experience, it is difficult to find examples of sets of more than a few items that meet the scaling requirements for a Guttman scale.

2.9.1.2 Normal and logistic trace line models

The Guttman representation of equation 2.14 does not allow for error. A somewhat more relaxed model that does allow for error is the Mokken scale where each item has a different degree of difficulty (as in the Guttman scale) but some errors are allowed Mokken (1971). More generally, two models of item responding that do allow for error that do not require different difficulties and have been studied in great detail are the cumulative normal and the logistic model. Both of these models consider that the probability of being correct on an item is an increasing function of the difference between the person's ability (θ) and the item's difficulty (δ). These two equations are the cumulative normal of $\theta - \delta$

Table 2.19 Hypothetical response patterns for eight subjects to seven items forming a Guttman scale. For a perfect Guttman scale the total number of items endorsed (rowSums) reflects the highest item endoresed.

```
> 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
```

$$prob(correct|\theta,\delta) = \frac{1}{\sqrt{2\pi}} \int_{-\mathrm{inf}}^{\theta-\delta} e^{-\frac{u^2}{2}} du \qquad (2.15)$$

and the logistic function

$$prob(correct|\theta,\delta) = \frac{1}{1+e^{\delta-\theta}}. \qquad (2.16)$$

With addition of a multiplicative constant (1.7) to the difference between δ and θ in the logistic equation, the two functions are almost identical over the range from -3 to 3 (Figure 2.8).

$$prob(correct|\theta,\delta) = \frac{1}{1+e^{1.7(\delta-\theta)}}. \qquad (2.17)$$

Latent Trait Theory (and the associated Item Response Theory, IRT) tends to use the equations 2.16 or 2.17 in estimating ability parameters for subjects given a set of known (or estimated) item difficulties (Figure 2.9). People are assumed to differ in their ability in some domain and items are assumed to differ in difficulty or probability of endorsement in that domain. The basic model of measuring ability is equivalent to a high jump competition. Given a set of bars on a high jump, what is the highest bar that one can jump?

Classical test theory (Chapter 6) may be thought of as a high jump with random height bars and many attempts at jumping. The total number of bars passed is the person's score. Item Response Theory approaches (Chapter 7) recognize that bars differ in height and allow jumpers to skip lower bars if they are able to pass higher bars. For example, in the mathematical ability domain, item difficulties may be ordered from easy to hard, (knowing your arithmetic facts, knowing long division, basic algebra, differential calculus, integral calculus, matrix algebra, etc.).

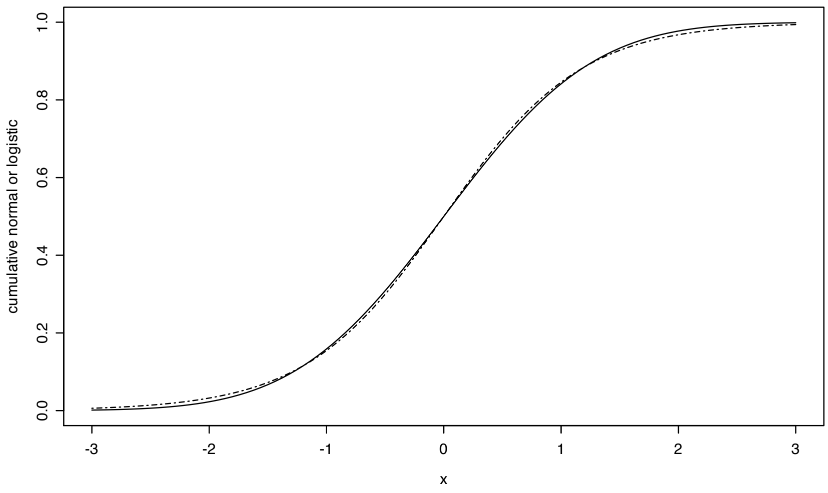

Fig. 2.8 The cumulative normal and the logistic (with a constant 1.7) are almost identical functions. The code combines a curve for the normal probability with a curve for the logistic function. curve(pnorm(x),-3,3,ylab=”cumulative normal or logistic”) curve(1/(1+exp(-1.7*x)),-3,3,add=TRUE,lty=4)

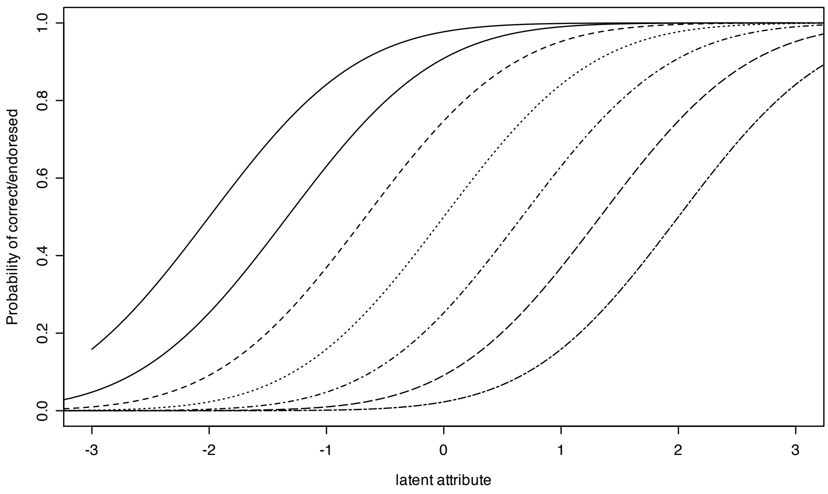

Fig. 2.9 The basic ability model. Probability of endorsing an item, or being correct on an item varies by item difficulty and latent ability.

2.9.2 Measurement of attitudes ($|s_i - o_j| < \delta$)

The alternative comparison model is one of proximity. This leads to a single peaked function (Figure 2.10). Some items are more likely to be endorsed the lower the the subject's attribute value, some are most likely to be endorsed at moderate levels, and others have an increasing probability of endorsement. Thus, if assessing neatness, the item "I am a messy person" will tend to peak at the lowest levels, while the item "My room neatness is about average" will peak at the mid ranges, and the item "I am a very neat person" will peak at the highest levels of the attribute. An item at the highest end of the dimension can be modeled using the ability model, and an item at the lowest level of difficulty can be modeled by reverse scoring it (treating rejections as endorsements, and treating endorsements as rejections). However, items in the middle range do not fit the ordering model and need to be modeled with a single peaked function Chernyshenko et al. (2007).

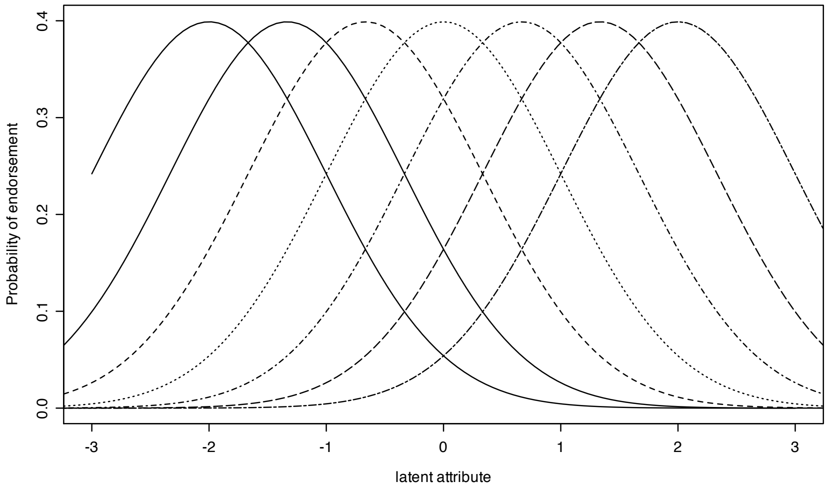

Fig. 2.10 Basic attitude model. Probability of endorsing an item is a single peaked function of the latent attribute and item location.

2.10 Theory of Data: some final comments

The kind of data we collect reflects the questions we are trying to address. Whether we compare objects to objects, subjects to subjects, or subjects to objects depends upon what is the primary question. Considering the differences between order and proximity information is also helpful, for some questions are more appropriately thought of as proximities (unfoldings) rather than ordering. Finally, that simple questions asked of the subject can yield stable metric and partially ordered metric information is an important demonstration of the power of modeling.

In the next chapter we consider what happens if we incorrectly assume metric properties of our data where in fact the mapping function from the latent to observed is not linear.