PLSC 504

Scaling and Item Response Theory

November 11, 2020

The Plan

- Scaling Overview
- Unidimensional Scaling
- Scale Reliability
- Multidimensional Scaling
- Item Response Theory

Some Terms

- Always $i \in \{1, 2, ...N\}$ observations on $k \in \{1, 2, ...K\}$ indicators.
- Z or $Z_1, Z_2, ...$ will indicate the underlying / latent trait(s) / phenomena
- $D_1, D_2, ...D_K$ are *dichotomous* indicators
- $Y_1, Y_2, ...Y_K$ are *continuous* indicators

Scaling, In General

Characteristics:

- *Comparative* vs. *Non-Comparative* Scaling
- *Subject-centered* vs. *Stimulus-centered*
- *Metric* vs. *Non-Metric*
- *Unidimensional* vs. *Multidimensional*

(Unidimensional) Scaling: History

Thurstone (1927, 1929)

- Non-comparative, subject-centered
- "Law of Comparative Judgment": The degree to which any two stimuli can be discriminated is a direct function of the difference in their status as regards the attribute in question.
- Methods: *paired comparisons*, *successive intervals*, and *equal-appearing intervals*.

Likert (1932)

- Non-comparative, subject-centered
- *Summative* + *unidimensional* $\rightarrow$ item construction & selection are key

Guttman (1944, 1950) ("scalogram analysis")

- Comparative; both subject- and stimulus-centered
- The response to each item is a simple function of the sum score

See McIver and Carmines (1981) for more...

Scaling: Dissimilarities and Distances

Dissimilarities matrix:

$$\underset{K \times K}{\Delta} = \begin{bmatrix} 0 & \delta_{21} & \delta_{31} & \dots & \delta_{K1} \\ \delta_{12} & 0 & \delta_{32} & \dots & \delta_{K2} \\ \delta_{13} & \delta_{23} & 0 & \dots & \delta_{K3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \delta_{1K} & \delta_{2K} & \delta_{3K} & \dots & 0 \end{bmatrix}$$

where

$$\delta_{ij} = f(\mathbf{Y}_i - \mathbf{Y}_j)$$

What's a "Distance"?

The function δ_{AB} of A and B is a distance function if it meets three criteria:

- $\delta_{AB} \geq 0$ and $\delta_{AB} = 0$ iff $A = B$
- $\delta_{AB} = \delta_{BA}$ (*symmetry*)
- $\delta_{AC} \leq \delta_{AB} + \delta_{BC}$ (*triangle inequality*)

Euclidean ("L2") Distance:

$$\delta_{ij} = \sqrt{(Y_{i1} - Y_{j1})^2 + (Y_{i2} - Y_{j2})^2 + ...(Y_{iN} - Y_{jN})^2}$$

Manhattan ("L1") Distance:

$$\delta_{ij} = |Y_{i1} - Y_{j1}| + |Y_{i2} - Y_{j2}| + ...|Y_{iN} - Y_{jN}|$$

Minkowski Order-p ("L^p") Distance:

$$\delta_{ij} = \left(\sum_{\ell=1}^{N} |Y_{i\ell} - Y_{j\ell}|^p \right)^{1/p}$$

Goal: Locate N points in a low (p)-dimensional space $p << N$ such that the (say) Euclidean distances between them approximate Δ.

That is, find a set of $N \times P$ points $\mathbf{D}$ such that $d_{ij}(\mathbf{D}) \approx \delta_{ij} \; \forall \, i, j$, where

$$d_{ij}(\mathbf{D}) = \sqrt{\sum_{p=1}^{P} (Y_{ip} - Y_{jp})^2}$$

and $p = \{1, 2, ...P\}$ denotes the dimensionality of the space.

For a $K \times K$ dissimilarity matrix Δ, the *stress* associated with a given set of coordinates $\mathbf{D}$ is:

$$\sigma(\mathbf{D}) = \sum_{i<j} w_{ij}[\delta_{ij} - d_{ij}(\mathbf{D})]^2$$

with weights

$$\mathbf{W} = \begin{bmatrix} 0 & w_{21} & w_{31} & \dots & w_{K1} \\ w_{12} & 0 & w_{32} & \dots & w_{K2} \\ w_{13} & w_{23} & 0 & \dots & w_{K3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ w_{1K} & w_{2K} & w_{3K} & \dots & 0 \end{bmatrix}$$

with the constraint that

$$\sum_{i<j} w_{ij}\delta_{ij}^2 = \frac{N(N-1)}{2}.$$

Scaling Types

Key: Transformations of the dissimilarities.

Ratio:

- $\hat{d}_{ij} = \delta_{ij}$ (no transformation)
- A special case of *metric* scaling

Interval:

- $\hat{d}_{ij} = a + b(\delta_{ij})$ (linear transformation)
- Also *metric*; "the ratio of differences of distances should be equal to the corresponding ratio of differences in the data"

Nonlinear:

- $\hat{d}_{ij} = g(\delta_{ij})$ (e.g., splines)
- Also *metric*

Ordinal:

- $\hat{d}_{ij} = f(\delta_{ij})$ such that $\delta_{ij} < \delta_{i'j'} \Rightarrow f(\delta_{ij}) < f(\delta_{i'j'})$
- *Monotone* / rank-preserving transformations
- *Nonmetric*

Unidimensional Scaling: The Sum Score

Simplest approach: the *sum score*:

$$\hat{Z}_i = \sum_{k=1}^{K} Y_{ik}$$

or

$$\hat{Z}_i = \frac{\sum_{k=1}^{K_i} Y_{ik}}{K_i}.$$

Requires:

- $Var(Y_j) = Var(Y_k) \forall j \neq k$
- $Cov(Y_j, Z) = Cov(Y_k, Z) \forall j \neq k$

For $p = 1$, the stress

$$\sigma(\mathbf{D}) = \sum_{i<j} w_{ij}[\delta_{ij} - d_{ij}(\mathbf{D})]^2$$

can be shown to be a function of the orders of the rank orders of the items (see Mair and deLeeuw 2014 for details). This means:

- No need to chose among distances / similarities.
- Solution is via combinatorics.

Example: Cities

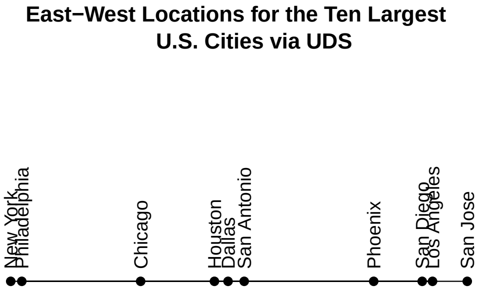
```
> Cities[,c(1,4)]
                    city  longitude
New York        New York  -74.00594
Los Angeles  Los Angeles -118.24368
Chicago          Chicago  -87.62980
Houston          Houston  -95.36980
Philadelphia Philadelphia  -75.16522
Phoenix          Phoenix -112.07404
San Antonio  San Antonio  -98.49363
San Diego      San Diego -117.16108
Dallas            Dallas  -96.79699
San Jose        San Jose -121.88633
```

Cities: Euclidean Distance Matrix (Δ)

```
> CityLong <- data.frame(t(Cities$longitude)) # longitudes in a row
> colnames(CityLong) <- t(Cities$city) # names
> D1long <- dist(t(CityLong)) # distance object
> D1long
             New York Los Angeles Chicago Houston Philadelphia Phoenix San Antonio San Diego Dallas
Los Angeles      44.2
Chicago          13.6        30.6
Houston          21.4        22.9     7.7
Philadelphia      1.2        43.1    12.5    20.2
Phoenix          38.1         6.2    24.4    16.7         36.9
San Antonio      24.5        19.8    10.9     3.1         23.3    13.6
San Diego        43.2         1.1    29.5    21.8         42.0     5.1        18.7
Dallas           22.8        21.4     9.2     1.4         21.6    15.3         1.7      20.4
San Jose         47.9         3.6    34.3    26.5         46.7     9.8        23.4       4.7   25.1
```

Unidimensional Scaling

```
> library(smacof)
> UDS <- uniscale(D1long)
> UDS

Call: uniscale(delta = D1long)

Final stress value: 4.1e-16
Number of accepted permutations: 180836
Number of possible permutations: 3628800
Number of objects: 10

> UDS$conf
   New York  Los Angeles      Chicago      Houston Philadelphia
   -1.04236      0.75350     -0.48929     -0.17508     -0.99530
    Phoenix  San Antonio    San Diego       Dallas     San Jose
    0.50304     -0.04827      0.70955     -0.11715      0.90137
```


East−West Locations for the Ten Largest U.S. Cities via UDS
New York
Philadelphia
Chicago
Houston
Dallas
San Antonio
Phoenix
San Diego
Los Angeles
San Jose

Another Example: SCOTUS Votes

```
> head(SCOTUS)
  id Rehnquist Stevens OConnor Scalia Kennedy Souter Thomas Ginsburg Breyer
  1          0       1       0      0       0      0      0        0      0
  2          0       0       0      0       0      0      0        0      0
  3          0       0       0      0       0      0      0        0      0
  4          1       1       1      1       1      1      1        1      1
  5          0       1       0      0       1      1      0        1      1
  7          1       1       1      0       1      1      0        1      1
```

SCOTUS: Sum Scores and Distances

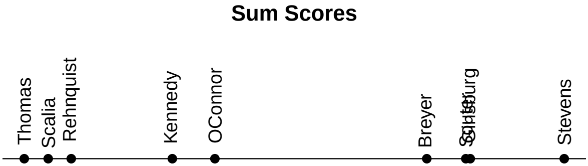
```
> SumScores <- colSums(SCOTUS[,2:10],na.rm=TRUE) / nrow(SCOTUS)
> SumScores
Rehnquist   Stevens   OConnor    Scalia   Kennedy    Souter    Thomas  Ginsburg  Breyer
   0.2838    0.6906    0.4023    0.2648    0.3672    0.6094    0.2451    0.6130  0.5772
```

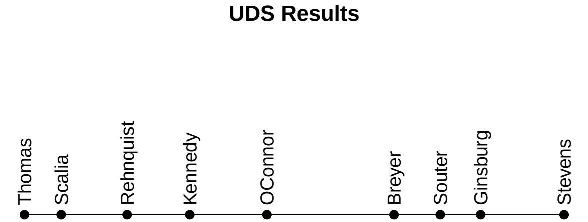
```
> D1SCOTUS <- dist(t(SCOTUS[,2:10]))
> D1SCOTUS
         Rehnquist Stevens OConnor Scalia Kennedy Souter Thomas Ginsburg
Stevens      25.61
OConnor      15.94   22.32
Scalia       13.49   26.80   18.38
Kennedy      13.27   23.62   16.31  15.62
Souter       22.52   15.39   18.73  23.56   20.57
Thomas       13.45   26.81   18.36  10.15   16.03  23.96
Ginsburg     22.72   15.10   19.65  24.00   20.88  11.27  24.64
Breyer       22.29   16.09   18.25  24.02   20.66  13.42  24.78    12.69
```

SCOTUS: UDS

```
> SCOTUS.UDS <- uniscale(D1SCOTUS)

> SCOTUS.UDS

Call: uniscale(delta = D1SCOTUS)

Final stress value: 0.317
Number of accepted permutations: 347136
Number of possible permutations: 362880
Number of objects: 9
```

Sum Scores and UDS





For a scale S that is a sum of K separate items $Y_1, Y_2, ... Y_K$,

$$\alpha = \frac{K}{K-1}\left(1 - \frac{\sum_{k=1}^{K} \sigma^2_{Y_k}}{\sigma^2_S}\right)$$

where

- $\sigma^2_{Y_k}$ is the variance of item k and
- σ^2_S is the variance of the scale S.

α: **"The expected correlation of two tests that measure the same construct."**

Cronbach's α (continued)

α:

- A *lower bound* to reliability.
- If $S = D_1 + D_2 + ... + D_K$, then $\sigma^2_{Y_k} = P_k(1 - P_k)$.
- $\alpha \in [0, 1]$ (theoretically)
- Rule of thumb: $\alpha \geq 0.7$ is "adequate"

Limitations (from Sijtsma 2009):

- Requires equal item variances, equal item covariances, and unidimensionality
- Among the "lower bounds," it's among the smallest
- A better one is the "greatest lower bound" (*glb*), but even it has problems...

Reliability: SCOTUS Data

```
> SCOTUSAlpha <- alpha(SCOTUS[,2:10],check.keys=TRUE)
> SCOTUSAlpha

Reliability analysis
Call: alpha(x = SCOTUS[, 2:10], check.keys = TRUE)

  raw_alpha std.alpha G6(smc) average_r S/N    ase mean   sd
       0.9       0.9    0.93      0.49 8.6 0.0044 0.45 0.35

 lower alpha upper     95% confidence boundaries
0.89 0.9 0.9

 Reliability if an item is dropped:
          raw_alpha std.alpha G6(smc) average_r S/N alpha se
Rehnquist      0.88      0.88    0.91      0.48 7.4   0.0049
Stevens        0.89      0.89    0.92      0.51 8.4   0.0045
OConnor        0.88      0.88    0.92      0.48 7.3   0.0051
Scalia         0.89      0.89    0.91      0.50 7.9   0.0046
Kennedy        0.88      0.88    0.92      0.48 7.3   0.0051
Souter         0.88      0.88    0.91      0.47 7.2   0.0051
Thomas         0.89      0.89    0.91      0.50 8.0   0.0046
Ginsburg       0.88      0.88    0.91      0.48 7.4   0.0051
Breyer         0.88      0.88    0.91      0.49 7.5   0.0050
```

Types...

- **Metric** - Inputs are interval/ratio-level measures; transformations from δ_{ij} to d_{ij} are cardinal-valued.
- **Non-Metric** - Inputs are binary or ordinal-level; transformations are rank-preserving *only*.

General steps:

- Generate a dissimilarity matrix Δ (chosing distance metric)
- Choose the dimensionality p
- Choose the type of scaling
- Fit model + interpret the results
- Assess model fit + conduct diagnostics

MDS: Model Fit / Diagnostics

Recall: Stress is:

$$\sigma(\mathbf{D}) = \sum_{i<j} w_{ij}[\delta_{ij} - d_{ij}(\mathbf{D})]^2$$

Kruskal's rule of thumb for stress:

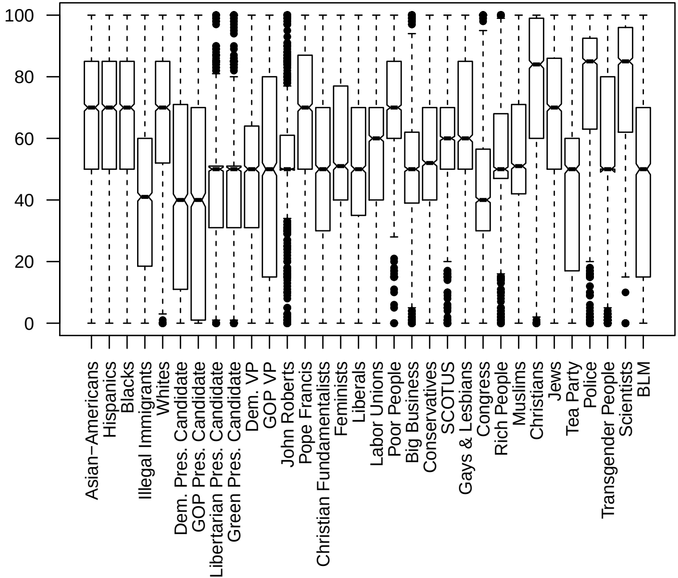
- $\sigma(\mathbf{D}) = 0.20 \rightarrow$ "poor"
- $\sigma(\mathbf{D}) = 0.10 \rightarrow$ "fair"
- $\sigma(\mathbf{D}) = 0.05 \rightarrow$ "good"
- $\sigma(\mathbf{D}) = 0.025 \rightarrow$ "excellent"
- $\sigma(\mathbf{D}) = 0 \rightarrow$ "perfect"

Key diagnostic: **Shepard plot**: a scatterplot of δ_{ij} vs. $\hat{d}_{ij}$...

- Illustrates model "fit"
- Also illustrates the transformation of the δ_{ij}s

Also **permutation tests** for model fit...

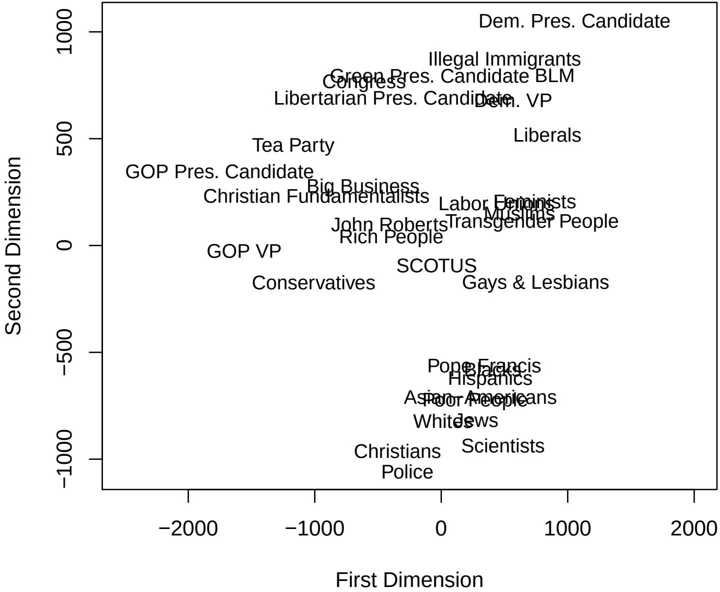
Example: 2016 ANES Thermometer Scores



MDS using `cmdscale`

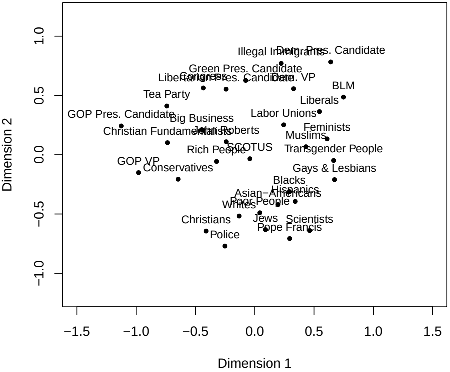
```
> MDS2.alt <- cmdscale(ThermDist,k=2)

> head(MDS2.alt)
                          [,1]   [,2]
Asian-Americans         309.90 -709.9
Hispanics               387.51 -627.6
Blacks                  407.03 -580.7
Illegal Immigrants      501.74  866.3
Whites                   14.72 -821.5
Dem. Pres. Candidate   1054.53 1053.2
```

MDS Plot, using `cmdscale`

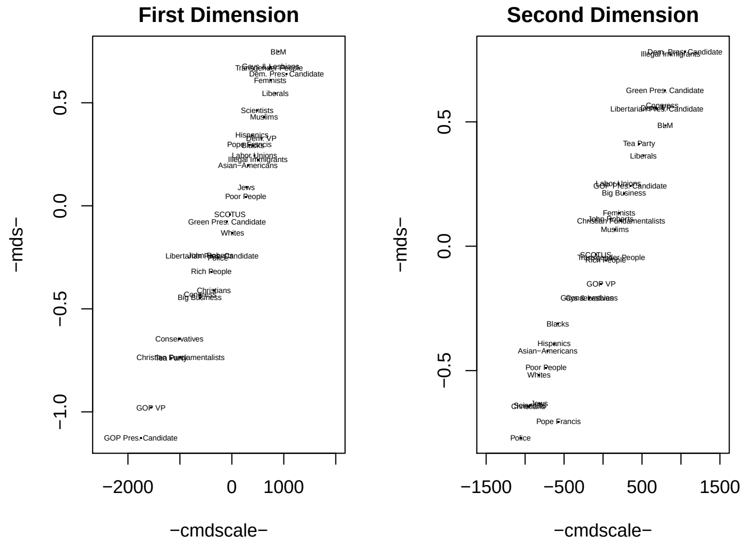


MDS using mds (ratio scaling)

```
> MDS2 <- mds(ThermDist, ndim=2)
> MDS2

Call:
mds(delta = ThermDist, ndim = 2)

Model: Symmetric SMACOF
Number of objects: 32
Stress-1 value: 0.221
Number of iterations: 87
```



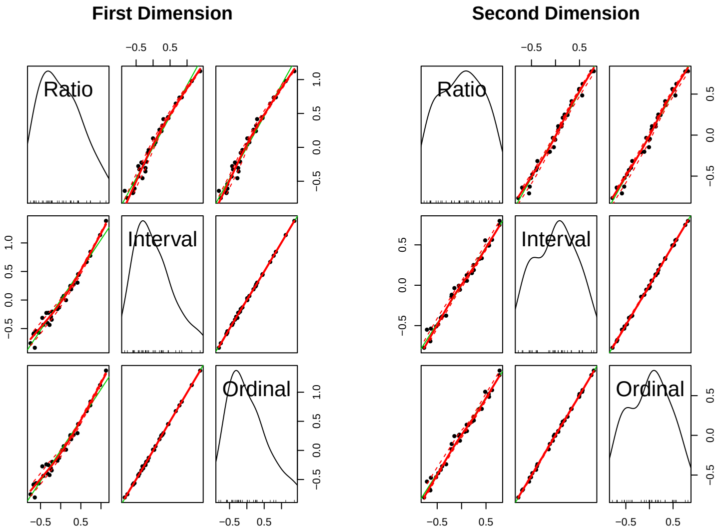
cmdscale and mds Comparison



Alternatives: Interval and Ratio Scaling

```
> MDS2.int <- mds(ThermDist, ndim=2, type="interval")
> MDS2.int

Call:
mds(delta = ThermDist, ndim = 2, type = "interval")

Model: Symmetric SMACOF
Number of objects: 32
Stress-1 value: 0.087
Number of iterations: 17

> # Ordinal:
>
> MDS2.ord <- mds(ThermDist, ndim=2, type="ordinal")
> MDS2.ord

Call:
mds(delta = ThermDist, ndim = 2, type = "ordinal")

Model: Symmetric SMACOF
Number of objects: 32
Stress-1 value: 0.071
Number of iterations: 22
```

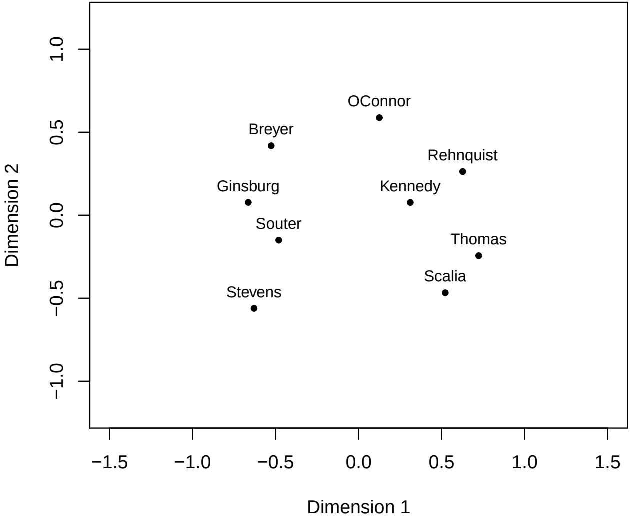
Ratio-Interval-Ordinal Comparison



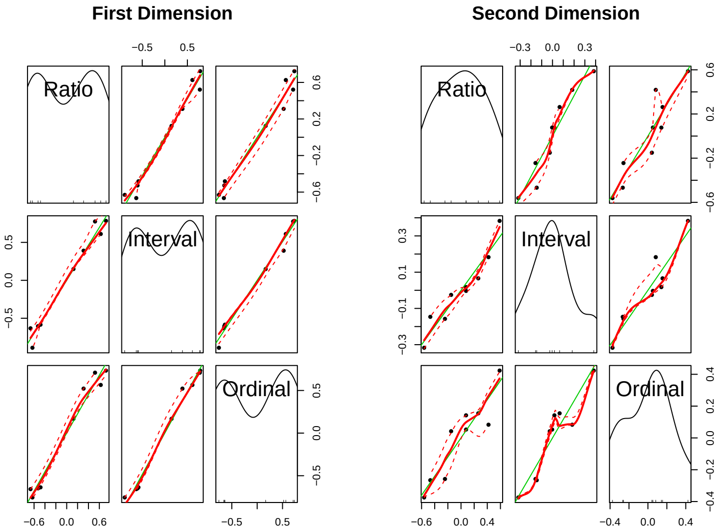
SCOTUS Redux ($p = 2$)

```
> SCR <- mds(D1SCOTUS, ndim=2, type="ratio")
> SCR

Call:
mds(delta = D1SCOTUS, ndim = 2, type = "ratio")

Model: Symmetric SMACOF
Number of objects: 9
Stress-1 value: 0.184
Number of iterations: 73
```

SCOTUS MDS Plot



Ratio-Interval-Ordinal Comparison



Useful References

- Kruskal, J.B., and M. Wish 1978. *Multidimensional Scaling.* Sage.
- McIver, John, and Edward C. and Carmines. 1981. *Unidimensional Scaling*. Sage Publications.
- Davison, Mark L. *Multidimensional Scaling*. 1983. New York: Wiley.
- Cox, Trevor F. and Michael A. A. Cox. 2000. *Multidimensional Scaling*, 2nd Ed. New York: Chapman & Hall.
- Borg, Ingwer, and Groenen, Patrick. 2005. *Modern Multidimensional Scaling: Theory and Applications*, 2nd Ed. Berlin: Springer-Verlag.
- Borg, Ingwer, Patrick Groenen, and Patrick Mair. 2013. *Applied Multidimensional Scaling*. Berlin: Springer-Verlag.

Useful R Packages and Routines

Distances, Proximities, etc.

- `dist` function (base R)
- `distances` package
- `proxy` package

Scaling

- `stats::cmdscale` (classical MDS, in base R)
- `smacof` (state-of-the-art MDS package)
- `vegan` (ecology package; has some good MDS routines)
- Others...

Useful Links

- `smacof` documentation.
- Seven ways to do MDS in R.
- Jan De Leeuw's website.

Item Response Theory (IRT)

Item Response Theory ("IRT")

- Origins in psychometrics / testing
- *Measurement* model – (typically) *no* **X**
- *Unidimensional*
- *Discrete* responses **Y**
- Equally descriptive and inferential

$$Y^* = \text{ latent trait (“ability”)}$$

$$Y = \text{observed measures}$$

- $i \in \{1, 2 \ldots N\}$ indexes *subjects* / *units*, and
- $j \in \{1, 2, \ldots J\}$ indexes *items* / *measures*.

$$Y_{ij} = \begin{cases} 0 & \text{if subject } i \text{ gets item } j \text{ “incorrect,”} \\ 1 & \text{if subject } i \text{ gets item } j \text{ “correct.”} \end{cases}$$

Data

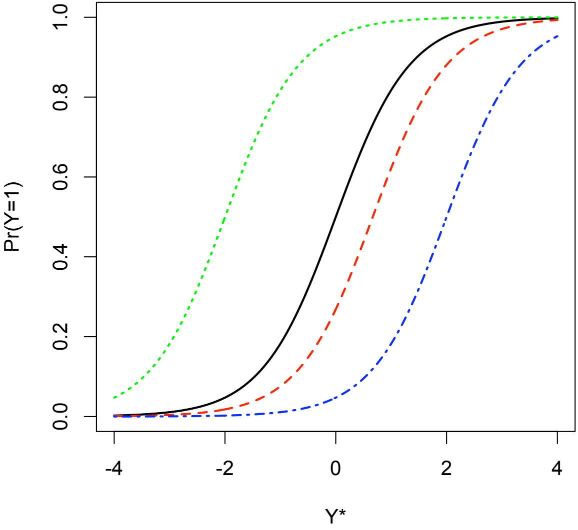
```
> head(SCOTUS,10)

   id Rehnquist Stevens OConnor Scalia Kennedy Souter Thomas Ginsburg Breyer
1   1         0       1       0      0       0      0      0        0      0
2   2         0       0       0      0       0      0      0        0      0
3   3         0       0       0      0       0      0      0        0      0
4   4         1       1       1      1       1      1      1        1      1
5   5         0       1       0      0       1      1      0        1      1
6   6         0       0       0      0       0      0     NA        0      0
7   7         1       1       1      0       1      1      0        1      1
8   8         0       1       0      0       0      0      0       NA      0
9   9         0       0       0      0       0      0      0        0      0
10 10         1       1       1      1       1      1      1        1      1
```

One-Parameter Logistic Model ("1PLM")

$$\Pr(Y_{ij} = 1) = \frac{\exp(\theta_i - \beta_j)}{1 + \exp(\theta_i - \beta_j)}$$

Here,

- θ_i = respondent i's *ability*,
- β_j = item j's *difficulty*.
- $\beta_j \equiv$ value of Y^* where $\Pr(Y_{ij} = 1) = 0.50$



1.0
0.8
0.6
0.4
0.2
0.0
Pr(Y=1)
-4
-2
0
2
4
Y*

1PLM

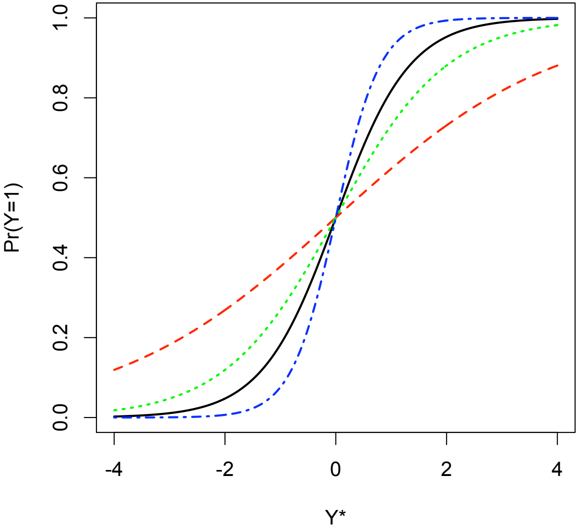
- a.k.a. "Rasch" model (Rasch 1960)
- Implicit "slope" $= 1.0$
- Implies items are equally "discriminating"
- If not...

Two-Parameter Logistic Model ("2PLM")

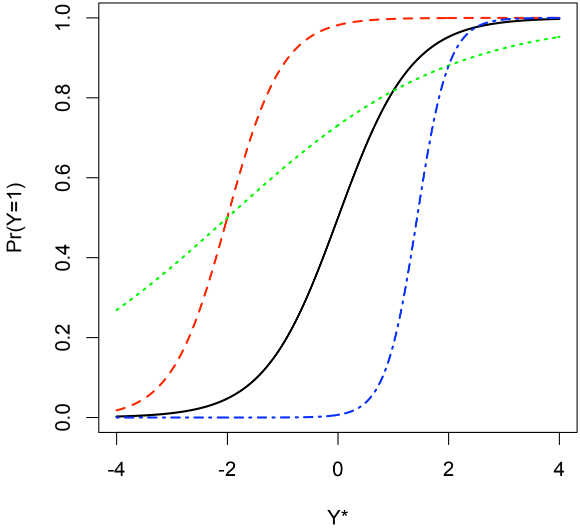
$$\Pr(Y_{ij} = 1) = \frac{\exp[\alpha_j(\theta_i - \beta_j)]}{1 + \exp[\alpha_j(\theta_i - \beta_j)]}$$

- θ_i = respondent i's *ability*,
- β_j = item j's *difficulty*,
- α_j = item j's *discrimination*.

Identical Difficulty, Different Discrimination



Different Difficulty & Discrimination



2PLM

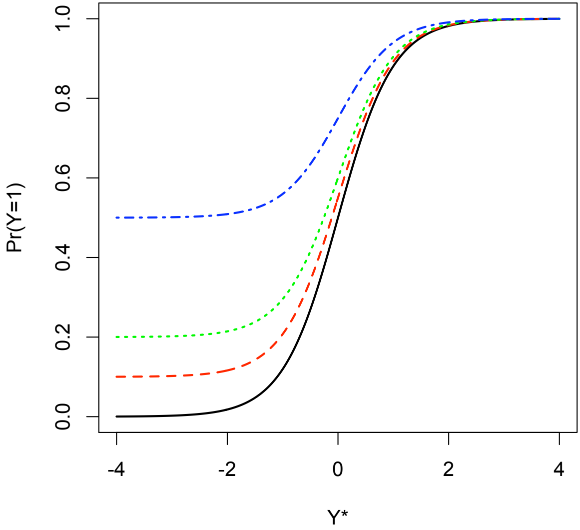
- Due to Birnbaum (1968)
- Similar to "typical" logit...
- Nests the 1PLM as a special case ($\alpha_j = 1 \forall j$)

Three-Parameter Logistic Model ("3PLM")

$$\Pr(Y_{ij} = 1) = \delta_j + (1 - \delta_j) \left\{ \frac{\exp[\alpha_j(\theta_i - \beta_j)]}{1 + \exp[\alpha_j(\theta_i - \beta_j)]} \right\}$$

- θ_i = respondent i's *ability*,
- β_j = item j's *difficulty*,
- α_j = item j's *discrimination*.
- δ_j = *lower asymptote* of $\Pr(Y_{ij} = 1)$ (incorrectly: "guessing" parameter).

3PLM, Constant α & β, Varying δ



- *Unidimensionality*
- *Local Item Independence* ("No LID"):

$$\text{Cov}(Y_{ij}, Y_{ik}|\theta_i) = 0 \; \forall \; j \neq k$$

Estimation: Notation

$$P_{ij} = \Pr(Y_{ij} = 1),$$

$$\begin{aligned} Q_{ij} &= \Pr(Y_{ij} = 0) \\ &= 1 - \Pr(Y_{ij} = 1), \end{aligned}$$

$$\Psi = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_J \\ \alpha_1 \\ \vdots \\ \alpha_J \\ \delta_1 \\ \vdots \\ \delta_J \end{pmatrix}.$$

Known $\Psi = \alpha,\ \beta,\ \delta$:

$$L(\mathbf{Y}|\Psi) = \prod_{j=1}^{J} P_{ij}^{Y_{ij}} Q_{ij}^{1-Y_{ij}}.$$

Known θ:

$$L(\mathbf{Y}|\theta) = \prod_{i=1}^{N} P_{ij}^{Y_{ij}} Q_{ij}^{1-Y_{ij}}.$$

Estimation: Likelihoods

$$L(\mathbf{Y}|\Psi,\theta) = \prod_{i=1}^{N}\prod_{j=1}^{J} P_{ij}^{Y_{ij}} Q_{ij}^{1-Y_{ij}}$$

$$\ln L(\mathbf{Y}|\Psi,\theta) = \sum_{i=1}^{N}\sum_{j=1}^{J} Y_{ij}\ln P_{ij} + (1-Y_{ij})Q_{ij}.$$

Parameterization

- $N + J$ parameters in the 1PLM,
- $N + 2J$ parameters in the 2PLM,
- $N + 3J$ parameters in the 3PLM.

But...

- NJ observations,
- Asymptotics as $N \to \infty$, $J \to \infty$...

Estimation: Conditional Likelihood

Total score is:

$$T_i = \sum_{j=1}^{J} Y_{ij} \; \in \; \{0, 1, ...J\}$$

$$L = \prod_{i=1}^{N} \frac{\exp[\alpha_j(\theta_t - \beta_j)]}{1 + \exp[\alpha_j(\theta_t - \beta_j)]}$$

θ_t are "score-group" parameters corresponding to the $J + 1$ possible values of T.

- Equivalent to fitting a conditional logit model:

$$\Pr(Y_{ij} = 1) = \frac{\exp(\mathbf{Z}_{ij}\gamma)}{\sum_{j=1}^{J} \exp(\mathbf{Z}_{ij}\gamma)}$$

  with $\mathbf{Z}_{ij} =$ "item dummies."

- Useful only for 1PLM (since T_i is a sufficient statistic for θ_i).

$$L(\mathbf{Y}|\Psi, \theta) = \prod_{i=1}^{N} \left[\int_{-\infty}^{\infty} \prod_{j=1}^{J} P_{ij}^{Y_{ij}} Q_{ij}^{1-Y_{ij}} d\theta \right]$$

- Analogous to "random effects"...
- Eliminates inconsistency as $N \rightarrow \infty$, *but*
- Requires *strong* exogeneity of θ and Ψ.

Estimation: Bayesian Approaches

- Place priors on θ, Ψ;
- Estimate via sampling from posteriors, via MCMC.
- Eliminates problems with $\hat{\alpha}$, $\hat{\beta}$, $\hat{\theta} = \infty$ (see below).
- Easily extensible to other circumstances (hierarchical/multilevel, etc.)

Identification

Two Issues:

- *Scale* invariance: $L(\hat{\Psi}) = L(\hat{\Psi} + c)$
- *Rotational* invariance: $L(\hat{\Psi}) = L(-\hat{\Psi})$

Fixes:

- Set one (arbitrary) $\beta_j = 0$, and another (arbitrary) $\beta_k > 0$, or
- Fix two θ_is at specific values.

Further (Potential) Concerns

- $Y_{ij} = 0/1 \; \forall \; i \;\rightarrow\; \beta_j = \pm\infty$.
- $Y_{ij} = 0/1 \; \forall \; j \;\rightarrow\; \theta_i = \pm\infty$.
- Separation / "empty cells" $\rightarrow\; \alpha_j = \pm\infty$.
- Problematic for joint and conditional approaches; more easily dealt with in the Bayesian framework.

- Estimates of $\hat{\alpha}$s, $\hat{\beta}$s, and/or $\hat{\delta}$s, plus $\hat{\theta}$s
- Associated s.e.s / c.i.s
- "Scale-free" quantities of interest...

IRT Models in R

- Library `ltm` (marginal estimation)
  - `rasch` (1PLM)
  - `ltm` (2PLM)
  - `tpm` (3PLM)
- Library `MCMCpack` (Bayesian estimation)
  - 1 and 2PLM
  - Standard, hierarchical, dyamic, multidimensional
- `ideal` (in library `pscl`) (Bayesian estimation)
  - 1 and 2PLM
  - *k*-dimensional
  - takes a `rollcall` object
- Other packages: `eRm`, `irtoys`, `irtProb`, `MiscPsycho`, etc.

Example: SCOTUS Voting, 1994-2004

```
> summary(SCOTUS)

       id          Rehnquist        Stevens         OConnor        Scalia
 Min.   :   1   Min.   :0.00   Min.   :0.00   Min.   :0.0   Min.   :0.00
 1st Qu.: 377   1st Qu.:0.00   1st Qu.:0.00   1st Qu.:0.0   1st Qu.:0.00
 Median : 753   Median :0.00   Median :1.00   Median :0.0   Median :0.00
 Mean   : 753   Mean   :0.28   Mean   :0.69   Mean   :0.4   Mean   :0.27
 3rd Qu.:1129   3rd Qu.:1.00   3rd Qu.:1.00   3rd Qu.:1.0   3rd Qu.:1.00
 Max.   :1505   Max.   :1.00   Max.   :1.00   Max.   :1.0   Max.   :1.00
                NA's   :49     NA's   :51     NA's   :55    NA's   :41
    Kennedy         Souter       Thomas         Ginsburg        Breyer
 Min.   :0.00   Min.   :0.0   Min.   :0.00   Min.   :0.00   Min.   :0.00
 1st Qu.:0.00   1st Qu.:0.0   1st Qu.:0.00   1st Qu.:0.00   1st Qu.:0.00
 Median :0.00   Median :1.0   Median :0.00   Median :1.00   Median :1.00
 Mean   :0.37   Mean   :0.6   Mean   :0.25   Mean   :0.61   Mean   :0.57
 3rd Qu.:1.00   3rd Qu.:1.0   3rd Qu.:0.00   3rd Qu.:1.00   3rd Qu.:1.00
 Max.   :1.00   Max.   :1.0   Max.   :1.00   Max.   :1.00   Max.   :1.00
 NA's   :32     NA's   :37    NA's   :44     NA's   :39     NA's   :61
```

1PLM Using `rasch`

```
> # 1PLM / Rasch Model:
> require(ltm)
> OnePLM<-rasch(SCOTUS[c(2:10)])
> summary(OnePLM)

Model Summary:
 log.Lik   AIC   BIC
   -5529 11079 11132

Coefficients:
                 value std.err z.vals
Dffclt.Rehnquist  0.46   0.040   11.5
Dffclt.Stevens   -0.59   0.030  -19.8
Dffclt.OConnor    0.14   0.030    4.6
Dffclt.Scalia     0.52   0.041   12.5
Dffclt.Kennedy    0.21   0.032    6.5
Dffclt.Souter    -0.36   0.027  -13.1
Dffclt.Thomas     0.60   0.043   13.8
Dffclt.Ginsburg  -0.37   0.027  -13.4
Dffclt.Breyer    -0.26   0.027   -9.9
Dscrmn            3.74   0.130   28.9

Integration:
method: Gauss-Hermite
quadrature points: 21

Optimization:
Convergence: 0
max(|grad|): 0.0027
quasi-Newton: BFGS
```

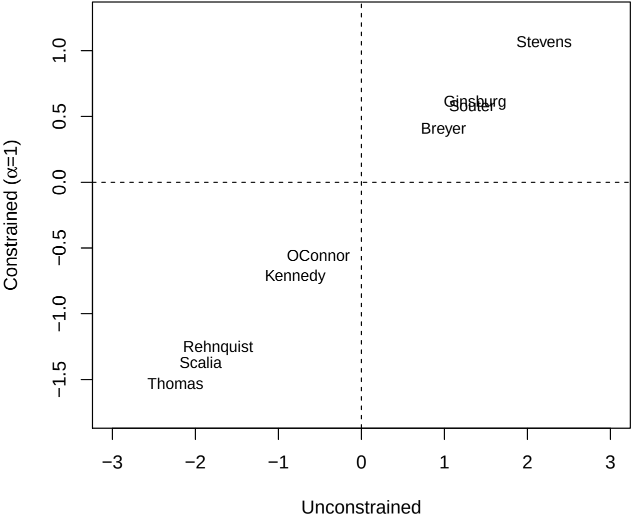
Converted to $\widehat{\Pr(Y_i = 1|\theta_i} = 0)$

```
> # Convert to probabilities given theta=0
>
> coef(OnePLM, prob=TRUE, order=TRUE)
          Dffclt Dscrmn P(x=1|z=0)
Stevens    -0.59    3.7      0.900
Ginsburg   -0.37    3.7      0.797
Souter     -0.36    3.7      0.791
Breyer     -0.26    3.7      0.729
OConnor     0.14    3.7      0.373
Kennedy     0.21    3.7      0.311
Rehnquist   0.46    3.7      0.151
Scalia      0.52    3.7      0.126
Thomas      0.60    3.7      0.096
```

Alternative Model Constraining $\alpha = 1.0$

```
> AltOnePLM<-rasch(IRTData, constraint=cbind(length(IRTData)+1,1))
> summary(AltOnePLM)

Model Summary:
 log.Lik   AIC   BIC
   -6452 12923 12971

Coefficients:
                 value std.err z.vals
Dffclt.Rehnquist  1.26   0.073   17.3
Dffclt.Stevens   -1.07   0.071  -15.1
Dffclt.OConnor    0.56   0.069    8.1
Dffclt.Scalia     1.37   0.074   18.6
Dffclt.Kennedy    0.72   0.069   10.4
Dffclt.Souter    -0.58   0.068   -8.6
Dffclt.Thomas     1.53   0.075   20.3
Dffclt.Ginsburg  -0.61   0.068   -8.9
Dffclt.Breyer    -0.40   0.068   -5.9
Dscrmn            1.00      NA     NA
```

Constrained and Unconstrained 1PLM $\hat{\beta}$s



2PLM

```
> TwoPLM<-ltm(IRTData ~ z1)
> summary(TwoPLM)

Coefficients:
                   value std.err z.vals
Dffclt.Rehnquist   0.44   0.035   12.3
Dffclt.Stevens    -0.63   0.038  -16.7
Dffclt.OConnor     0.14   0.026    5.6
Dffclt.Scalia      0.59   0.042   14.1
Dffclt.Kennedy     0.20   0.028    7.2
Dffclt.Souter     -0.27   0.025  -10.7
Dffclt.Thomas      0.68   0.044   15.2
Dffclt.Ginsburg   -0.29   0.025  -11.8
Dffclt.Breyer     -0.24   0.025   -9.6
Dscrmn.Rehnquist   4.77   0.377   12.7
Dscrmn.Stevens     2.46   0.165   14.9
Dscrmn.OConnor     4.14   0.341   12.1
Dscrmn.Scalia      2.82   0.188   15.0
Dscrmn.Kennedy     4.74   0.448   10.6
Dscrmn.Souter      6.69   0.535   12.5
Dscrmn.Thomas      2.84   0.190   14.9
Dscrmn.Ginsburg    5.83   0.439   13.3
Dscrmn.Breyer      3.76   0.253   14.9
```

2PLM: Probabilities and Testing

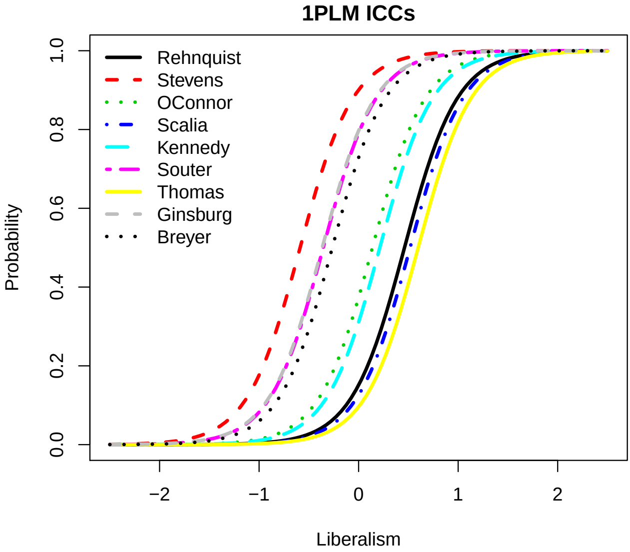
```
> coef(TwoPLM, prob=TRUE, order=TRUE)

          Dffclt Dscrmn P(x=1|z=0)
Stevens    -0.63    2.5       0.82
Ginsburg   -0.29    5.8       0.85
Souter     -0.27    6.7       0.86
Breyer     -0.24    3.8       0.71
OConnor     0.14    4.1       0.35
Kennedy     0.20    4.7       0.28
Rehnquist   0.44    4.8       0.11
Scalia      0.59    2.8       0.16
Thomas      0.68    2.8       0.13
```

```
> anova(OnePLM, TwoPLM)

 Likelihood Ratio Table
         AIC   BIC log.Lik   LRT df p.value
OnePLM 11079 11132   -5529
TwoPLM 10882 10978   -5423 212.7  8  <0.001
```

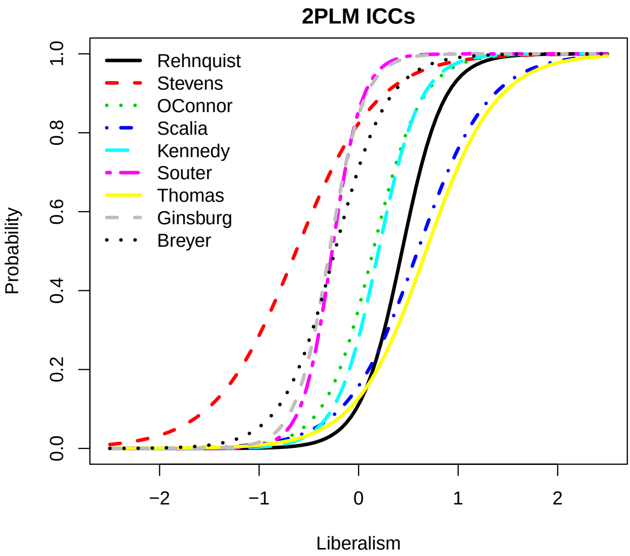
3PLM

```
> ThreePLM<-tpm(IRTData)
> summary(ThreePLM)

Coefficients:
                   value  std.err  z.vals
Gussng.Rehnquist  0.049    0.008   6.260
Gussng.Stevens    0.000    0.001   0.018
Gussng.OConnor    0.043    0.013   3.415
Gussng.Scalia     0.097    0.011   9.119
Gussng.Kennedy    0.071    0.014   5.162
Gussng.Souter     0.011    0.029   0.386
Gussng.Thomas     0.087    0.010   8.900
Gussng.Ginsburg   0.000    0.000   0.009
Gussng.Breyer     0.000    0.000   0.004
Dffclt.Rehnquist  0.716    0.030  23.511
Dffclt.Stevens   -0.630    0.038 -16.434
Dffclt.OConnor    0.340    0.040   8.537
Dffclt.Scalia     0.759    1.766   0.430
Dffclt.Kennedy    0.500    0.041  12.170
Dffclt.Souter    -0.294    0.063  -4.642
Dffclt.Thomas     0.808   10.610   0.076
Dffclt.Ginsburg  -0.329    0.030 -10.970
Dffclt.Breyer    -0.232    0.031  -7.439
Dscrmn.Rehnquist  8.735    4.259   2.051
Dscrmn.Stevens    2.577    0.181  14.214
Dscrmn.OConnor    3.979    0.439   9.068
Dscrmn.Scalia    26.537  578.889   0.046
Dscrmn.Kennedy    4.408    0.588   7.498
Dscrmn.Souter     6.698    1.416   4.731
Dscrmn.Thomas    34.074 2779.161   0.012
Dscrmn.Ginsburg   5.800    0.509  11.394
Dscrmn.Breyer     3.538    0.231  15.335
```

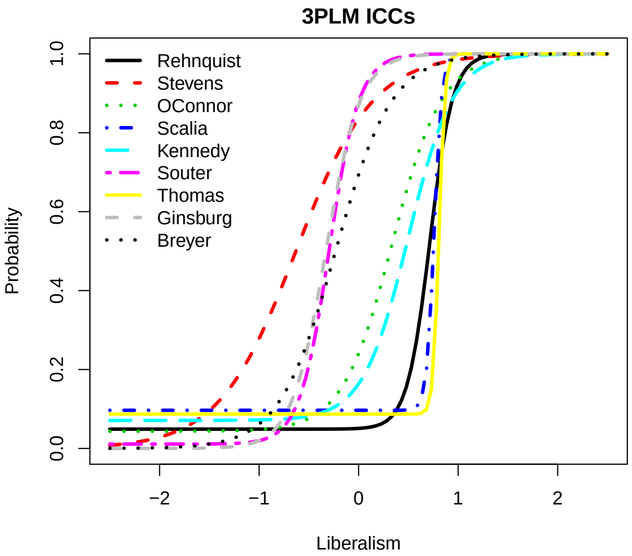
```
> anova(TwoPLM, ThreePLM)

 Likelihood Ratio Table
           AIC   BIC log.Lik    LRT df p.value
TwoPLM   10882 10978   -5423
ThreePLM 10737 10881   -5342 162.94  9  <0.001
```


1PLM ICCs
Rehnquist
Stevens
OConnor
Scalia
Kennedy
Souter
Thomas
Ginsburg
Breyer
Probability
0.0
0.2
0.4
0.6
0.8
1.0
−2
−1
0
1
2
Liberalism

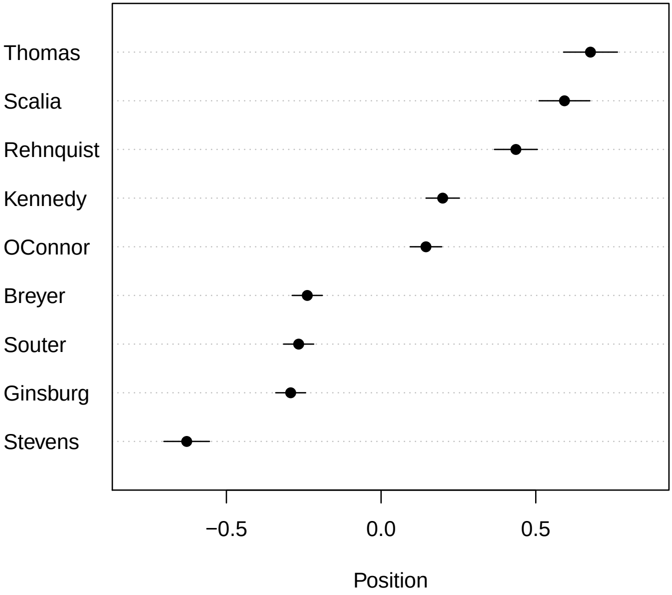
Cool Plots, II



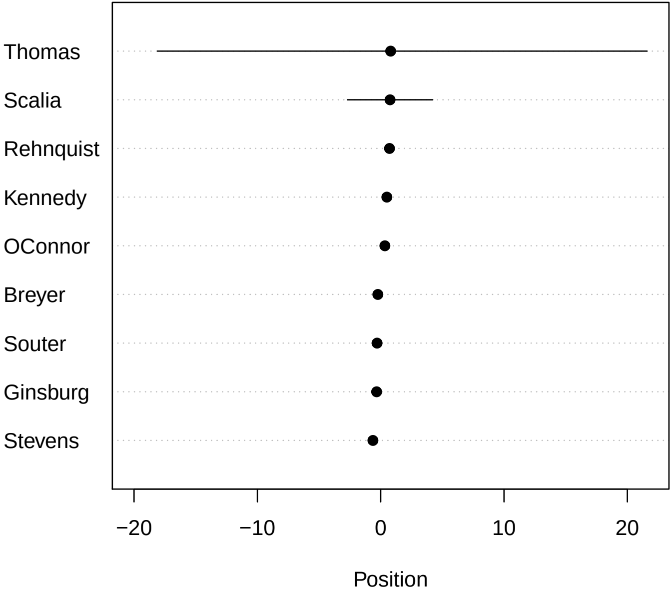
Cool Plots, III



Presenting Measures: Ladderplots (2PLM)



3PLM Ladderplot (#wtf)



Miscellaneous Things, I: Dimensionality

- Usually, *unidimensional*

- Sometimes, *two-dimensional*

- Tests:
    - Tetrachoric correlations among items
    - DIMTEST (Stout & Zhang, etc.)
    - Yen's Q_3
    - 1-D vs. 2-D comparisons (LR tests, etc.)

Miscellaneous Things, II: "DIF"

- *Differential item functioning*

- Formally,

$$\Pr(Y_{ij} = 1) = \Lambda[\alpha_j(\theta_i - \mathbf{X}_i\beta_j)].$$

- $\rightarrow$ violates *local item independence*

Extensions

- Nominal/Multinomial Y
- Ordinal Y:
  - *Graded response model* ("GRM") (Samejima 1969)
  - *Partial credit model* (Masters 1982)
  - *Generalized partial credit model* (Muraki 1992)
- Models for mixed response types (Thissen and Wainer 2001, 2003)
- Hierarchical IRT models (e.g. Bolt and Kim 2005)
- Models with covariates (e.g., DeBoeck and Wilson 2004)

Further Reading / Useful References

Hambleton, Ronald K., H. Swaminathan, and H. Jane Rogers. 1991. *Fundamentals of Item Response Theory*. Newbury Park CA: Sage Publications.

Fahrmeier, L., and G. Tutz. 2000. *Multivariate Statistical Modelling Based on Generalized Linear Models*. Berlin: Springer-Verlag.

De Boeck, Paul, and Mark Wilson, Eds. 2004. *Explanatory Item Response Models: A Generalized Linear and Nonlinear Approach.* New York: Springer.

de Ayala, R. J. 2008. *The Theory and Practice of Item Response Theory*. New York: The Guilford Press.

Baker, Frank B., and Seock-Ho Kim. 2017. *The Basics of Item Response Theory Using R*. New York: Springer.

van der Linden, Wim J., Ed. 2018. *Handbook of Item Response Theory* (3 vol.). New York: Chapman & Hall CRC.