APSTA-GE 2094
APSY-GE 2524

Modern Approaches in Measurement: Lecture 4

Klint Kanopka

New York University

Table of Contents

Announcements

- PS1 is due tomorrow (2.13 @ 11.59p)
- PS0 and PS1 solutions and grades will come out at the same time
- Next we need to get some code running:
  - Download the code
  - Download the data
  - Run everything up through fitting `m1` and `m2` using `mirt`
  - Depending on your computer, this could take a while!

Check-In

- PollEv.com/klintkanopka

More Item Response Theory

Looking at IRFs

- These plots are also called Item Characteristic Curves (ICCs)
- Go to: https://www.desmos.com/calculator/5rcatk0pak
- With 3 ± 1 people, answer the following questions:

1. What happens to each of the plots as you modify b_1, a, and b_2?
2. Can you notice any weird behaviors?
3. What value of a makes the two IRFs the same shape?
4. Do the b parameters do the same thing in each model? Does the behavior of b change as a changes?

Rasch Measurement

With 3 ± 1 people, answer the following questions:

1. What is *specific objectivity*?
2. Why was Rasch obsessed with sufficient statistics?
3. What were your big thoughts about this reading?
4. Why won't a 2PL work with Rasch's methods?

Missing Data

- You may have heard me casually said, "missing data doesn’t matter in IRT models"
- What do I mean by this?
- Why doesn’t it matter?
- What are the consequences of missing data?

Missing Data

- Back to the likelihood!
- Here's the probability of one response and the likelihood across all our data

$$P(X_{ij} = 1 | \theta_i, a_j, b_j) = \frac{1}{1 + e^{-a_j(\theta_i - b_j)}}$$

- The likelihood only involves responses that we directly observe

$$\mathcal{L}(\theta_i, a_j, b_j | X) = \prod_i \prod_j \frac{1}{1 + e^{-a_j(\theta_i - b_j)}}$$

- If we don't see a response, it doesn't appear in the product!
- Missing data just reduces the *precision* on our estimates of our parameters: θ_i, a_j, b_j

Break

Supreme Court Voting Records

Supreme Court voting records

- We don’t typically think about voting records as item response data, but it is!

- There are some key assumptions we make here—what are they?

Supreme Court voting records

- We don't typically think about voting records as item response data, but it is!
- Individuals (Justices, i)

- There are some key assumptions we make here—what are they?

Supreme Court voting records

- We don’t typically think about voting records as item response data, but it is!
- Individuals (Justices, i)
- respond to items (cases, j)

- There are some key assumptions we make here—what are they?

Supreme Court voting records

- We don’t typically think about voting records as item response data, but it is!
- Individuals (Justices, i)
- respond to items (cases, j)
- by endorsing the majority opinion ($X_{ij} = 1$)

- There are some key assumptions we make here—what are they?

Supreme Court voting records

- We don't typically think about voting records as item response data, but it is!
- Individuals (Justices, i)
- respond to items (cases, j)
- by endorsing the majority opinion ($X_{ij} = 1$)
- or dissenting ($X_{ij} = 0$)
- There are some key assumptions we make here—what are they?

Preparing the data

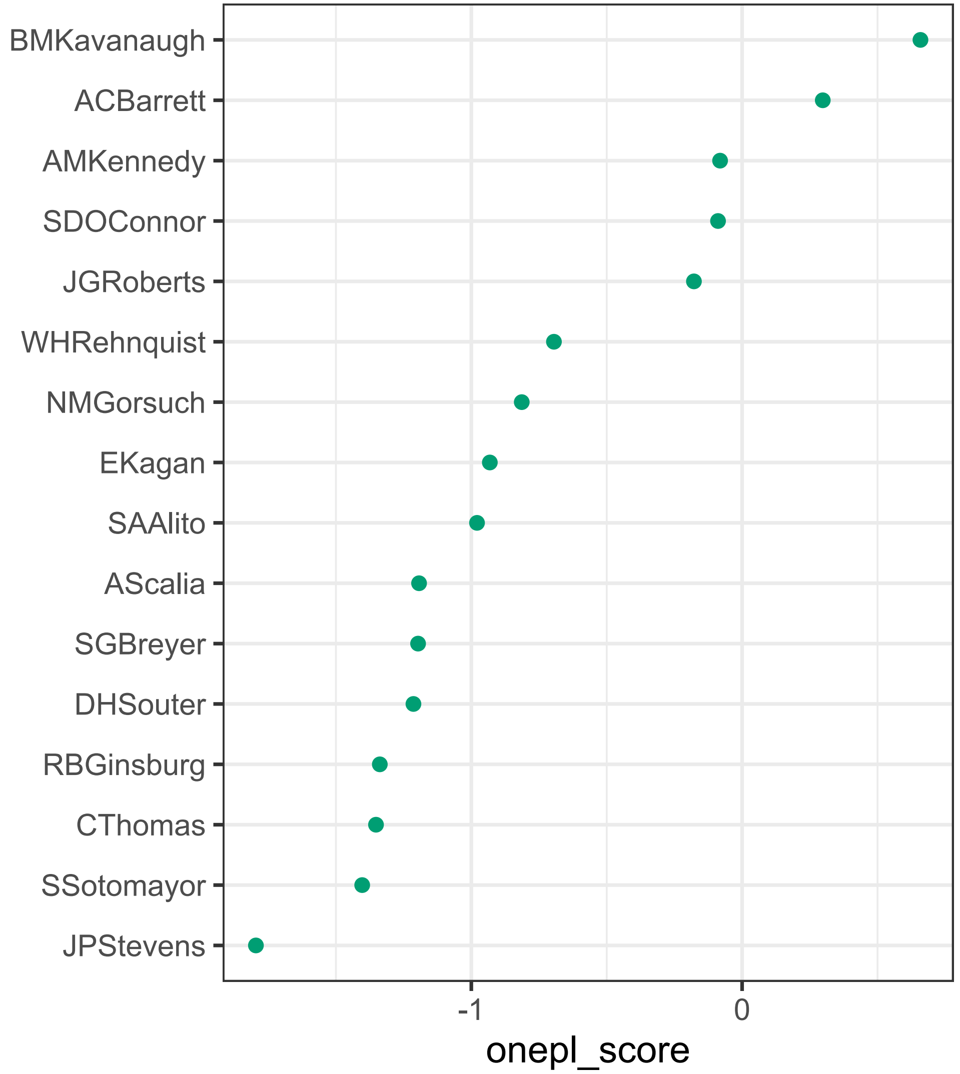
- Only consider cases after 2000
- Recode majority opinion/concurrence as 1 and dissent as 0, dropping everything else
- Drop cases with no disagreement

```
d <- SCDB_2022_01_justiceCentered_Citation |>
  filter(year(dateDecision) >= 2000) |>
  select(caseId, justiceName, vote) |>
  mutate(vote = case_when(vote %in% c(1, 3) ~ 1,
                          vote == 2 ~ 0,
                          TRUE ~ NA_real_)) |>
  na.omit() |>
  group_by(caseId) |>
  mutate(var = var(vote, na.rm=T)) |>
  filter(var != 0) |>
  pivot_wider(id_cols = justiceName, names_from = caseId, values_from = vote)
```

Estimating a 1PL

```r
resp <- select(d, -justiceName)
m1 <- mirt(resp,
           model = 1,
           itemtype = 'Rasch')

d$onepl_score <- fscores(m1)

ggplot(d,
       aes(x = onepl_score,
           y = reorder(justiceName,
                       onepl_score))) +
  geom_point(color = okabeito_colors(3)) +
  labs(y=NULL) +
  theme_bw()
```

- What have we just measured?
- Why?



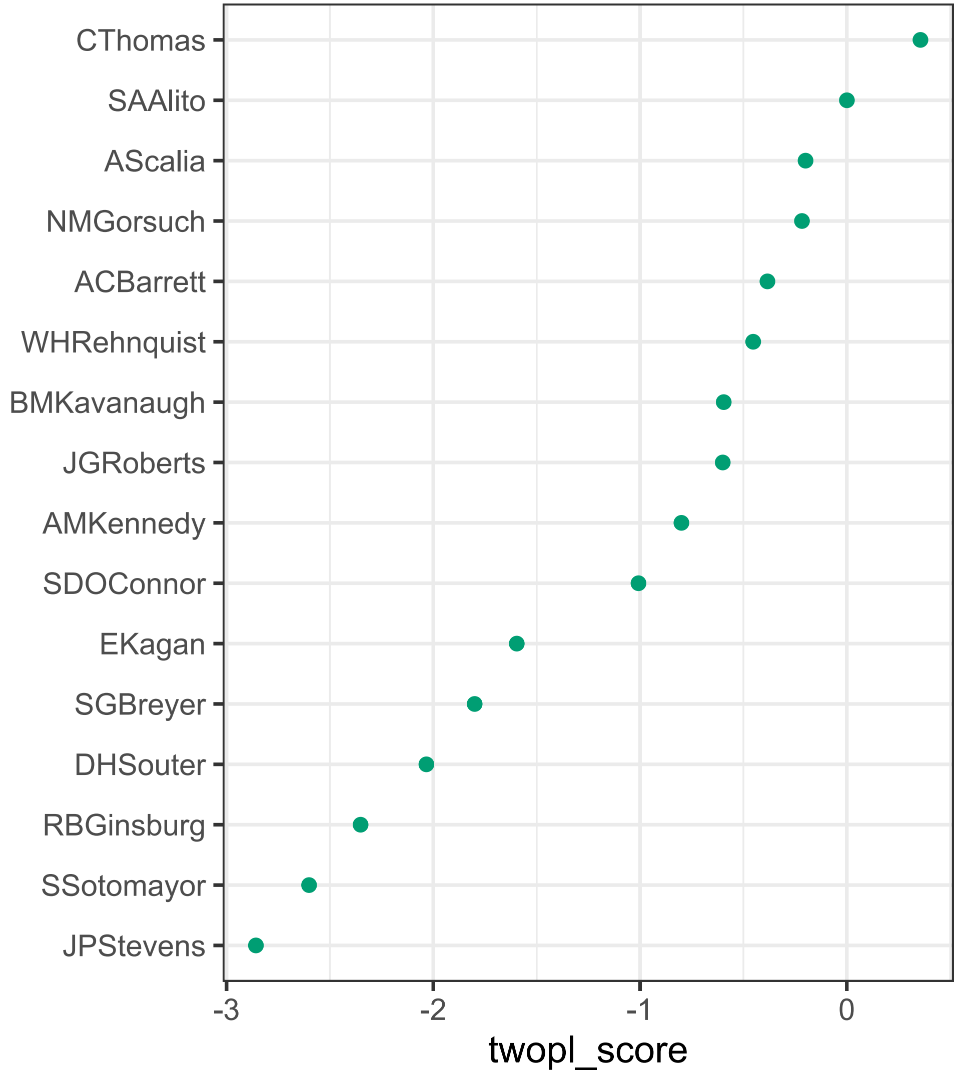
Estimating a 2PL

```
m2 <- mirt(resp,
           model = 1,
           itemtype = '2PL')

anova(m1, m2)
```

```
        AIC    SABIC        HQ       BIC     l
m1 11463.172 9246.408 11501.508 12211.810 -476
m2  6930.605 2501.652  7007.199  8426.337 -152
```

```
d$twopl_score <- fscores(m2)

ggplot(d,
       aes(x = twopl_score,
           y = reorder(justiceName,
                       twopl_score))) +
  geom_point(color = okabeito_colors(3)) +
  labs(y=NULL) +
  theme_bw()
```

- What did we measure now?



Examining Cases

```r
case_names <- SCDB_2022_01_justiceCentered_Citation |>
  select(caseId, caseName) |>
  distinct()

items <- data.frame(coef(m2, IRTpars=TRUE, simplify=TRUE)$items) |>
  select(-g, -u) |>
  rownames_to_column('caseId') |>
  left_join(case_names, by='caseId')

items |>
  arrange(-b) |>
  head(5) |>
  pull(caseName)
```

```
[1] "A. ELLIOTT ARCHER, ET UX. v. ARLENE L. WARNER"
[2] "VOLVO TRUCKS NORTH AMERICA, INC. v. REEDER-SIMCO GMC, INC."
[3] "NIKE, INC., et al. v. MARC KASKY"
[4] "WESTERNGECO LLC v. ION GEOPHYSICAL CORP."
[5] "CSX TRANSPORTATION, INC., PETITIONER v. ALABAMA DEPARTMENT OF REVENUE et al."
```

Examining Cases

```r
items |>
  arrange(b) |>
  head(5) |>
  pull(caseName)
```

```
[1] "PETRELLA v. METRO-GOLDWYN-MAYER, INC."
[2] "CHARLES ANDREW FOWLER, AKA MAN, PETITIONER v. UNITED STATES"
[3] "EMPIRE HEALTHCHOICE ASSURANCE, INC., DBA EMPIRE BLUE CROSS BLUE SHIELD v. DENISE F. MCVEIGH, AS ADMINISTRATRI
[4] "SUPAP KIRTSAENG, DBA BLUECHRISTINE99, PETITIONER v. JOHN WILEY & SONS, INC."
[5] "CUOZZO SPEED TECHNOLOGIES, LLC v. LEE"
```

```r
items |>
  filter(abs(b) < 1) |>
  arrange(b) |>
  head(5) |>
  pull(caseName)
```

```
[1] "BEN CHAVEZ v. OLIVERIO MARTINEZ"
[2] "UTAH, et al. v. DONALD L. EVANS, SECRETARY OF COMMERCE, et al."
[3] "NATIONAL RAILROAD PASSENGER CORPORATION v. ABNER MORGAN, JR."
[4] "UNITED STATES v. ARTHREX INC."
[5] "TRANSUNION LLC v. RAMIREZ"
```

Questions

With a 3 ± 1 people, answer the following:

1. How do the item parameters look to you? Are there any potential issues?
2. Because of these parameters, how must the IRFs look?
3. What were the major assumptions we made in doing this? How did we try to address them? If we didn't try to address them, how could we check them?

Extensions of IRT Models

Multidimensional IRT Models

- What happens when we think that our items are measuring more than one thing?
- Multiple factors in an instrument imply multidimensionality
- How might we modify an IRT model to include multidimensionality?

Multidimensional IRT Models

- The 2F2PL:

$$P(X_{ij} = 1) = \frac{1}{1 + e^{-(a_{1j}\theta_{1i} + a_{2j}\theta_{2i} + b_j)}}$$

- Two factors, two *types* of item parameters (discriminations and difficulty), logistic link function
- If we have 10 two factor items and 100 respondents, how many total parameters will we estimate?

Multidimensional IRT Models

- The 2F2PL:

$$P(X_{ij} = 1) = \frac{1}{1 + e^{-(a_{1j}\theta_{1i} + a_{2j}\theta_{2i} + b_j)}}$$

- Two factors, two *types* of item parameters (discriminations and difficulty), logistic link function
- If we have 10 two factor items and 100 respondents, how many total parameters will we estimate?
- 30 total item parameters

Multidimensional IRT Models

- The 2F2PL:

$$P(X_{ij} = 1) = \frac{1}{1 + e^{-(a_{1j}\theta_{1i} + a_{2j}\theta_{2i} + b_j)}}$$

- Two factors, two *types* of item parameters (discriminations and difficulty), logistic link function
- If we have 10 two factor items and 100 respondents, how many total parameters will we estimate?
- 30 total item parameters
- Actually fewer (27) because of identification assumptions

Multidimensional IRT Models

- The 2F2PL:

$$P(X_{ij} = 1) = \frac{1}{1 + e^{-(a_{1j}\theta_{1i} + a_{2j}\theta_{2i} + b_j)}}$$

- Two factors, two *types* of item parameters (discriminations and difficulty), logistic link function
- If we have 10 two factor items and 100 respondents, how many total parameters will we estimate?
- 30 total item parameters
- Actually fewer (27) because of identification assumptions
- 200 total person parameters

Multidimensional IRT Models

- This is the general form of a multidimensional IRT model with K factors $\{\theta_1, \ldots, \theta_K\}$

$$P(X_{ij} = 1) = \frac{1}{1 + e^{-(b_j + \sum_{k=1}^{K} a_{kj}\theta_{ki})}}$$

- These models can be tricky to estimate and have convergence issues!
  - `mirt` will fit an exploratory IRT model if you ask it to
  - In practice, fit an exploratory factor analysis first
  - Treat the subsequent IRT model as *confirmatory* (i.e., evaluate the fit of our model on real data)
  - Constrain some of the $a_{ki} = 0$ based on your preferred solution
  - You specify this in the `mirt.model` syntax
  - Adding priors on item parameters, increasing `NCYCLES` to allow more estimation iterations, and increasing `quadpts` to get better numerical precision can also help here

Compensatory and Non-Compensatory IRT Models

- This model is *compensatory*; Higher levels of one θ can compensate for lower levels of another θ

$$P(X_{ij} = 1) = \frac{1}{1 + e^{-(b_j + \sum_{k=1}^{K} a_{kj}\theta_{ki})}}$$

Compensatory and Non-Compensatory IRT Models

- This model is *compensatory*; Higher levels of one θ can compensate for lower levels of another θ

$$P(X_{ij} = 1) = \frac{1}{1 + e^{-(b_j + \sum_{k=1}^{K} a_{kj}\theta_{ki})}}$$

- This is the most straightforward *non-compensatory* model; How does it work?

$$P(X_{ij} = 1) = \prod_{k=1}^{K} \frac{1}{1 + e^{-a_{kj}(\theta_{ki} - b_{kj})}}$$

Compensatory and Non-Compensatory IRT Models

- This model is *compensatory*; Higher levels of one θ can compensate for lower levels of another θ

$$P(X_{ij} = 1) = \frac{1}{1 + e^{-(b_j + \sum_{k=1}^{K} a_{kj}\theta_{ki})}}$$

- This is the most straightforward *non-compensatory* model; How does it work?

$$P(X_{ij} = 1) = \prod_{k=1}^{K} \frac{1}{1 + e^{-a_{kj}(\theta_{ki} - b_{kj})}}$$

- The total response probability is always constrained by the weakest latent trait!

Estimation with `lme4`

- The `lme4` package can estimate Rasch-type models really easily!
  - No models with discrimination parameters, however!
- `m <- glmer(resp ~ (1|id) + item, data=d, family='binomial')`
  - Requires long form data
  - Of the type we get in the IRW!
- What do we consider to be fixed and random effects here?
  - What does having a random effect do?

Explanatory item response models

- Sometimes we have other features that we think matter for the item response!
  - Often these are features of the item or situation
  - These could be person-level covariates
- We think these features *explain* something about the response process
- These models are *very* easy to fit with `lme4`
- These models are less easy to fit using `mirt`, but it works
  - `mirt` handles this through the `covdata` and `formula` argument to relate covariates to the latent traits (latent regression; not explanatory)
  - You can also use the `item.formula` argument to decompose item difficulties as a function of covariates (explanatory model)
- Let's say we think log response time, $\log t_{ij}$ is related to the observed item response and we want to fit a 2PL, what would the $P(X_{ij} = 1|\theta_i, t_{ij})$ function look like?

Explanatory item response models

- An explanatory model with response time

$$P(X_{ij} = 1|\theta_i, t_{ij}) = \frac{1}{1 + e^{-(a_j\theta_i + b_j + \beta_j \log t_{ij})}}$$

- What would it mean if $\beta_j > 0$? Or $\beta_j < 0$?
- How does this change our interpretation of b_j?

Wrap Up

Recap

- IRT is more flexible than it seems on face
  - We can use it with data that isn't from tests and surveys
  - We can model multiple latent traits simultaneously
  - We can include covariates that we believe are related to the response process
- Assignment reminders:
  - PS1 due tonight
  - PS2 posted! It contains multidimensional IRT and explanatory modeling

Check-Out

- PollEv.com/klintkanopka