Week 10: Unsupervised Scale Measurement II
Categorical Indicators
POLS0013 Measurement in Data Science

Dr. Julia de Romémont

Academic Year 23-24

UCL Departement of Political Science

A reminder about the use of AI in your assessment

Can I use AI tools in my quantitative methods assignments?

- Yes, but only for certain tasks:
  - To correct errors in your code or to solve specific, common coding problems.
  - To help improve your writing, including greater clarity or more accurate grammar.
  - To support your efforts to resolve conceptual queries, although you should always make use of your classes, support and feedback hours, and moodle forums first.
- This means you cannot use it:
  - To write parts or all of an assessment;
  - To generate outlines, structures and high-level arguments for essays;
  - For rewriting or paraphrasing text from other sources for use in written work.

All use of AI must be acknowledged, described and referenced in your essay.

Last lecture was about...

Unsupervised measurement with

1. Input: ***continuous*** indicators on interval-level scales
2. Output: a ***continuous*** measure in form of a scale

Principle Components Analysis (PCA)

- A method for efficiently summarising the variation in a set of variables
- A method for putting weights on indicators for an index

Exploratory Factor Analysis (EFA)

- A model for describing variation in a set of variables in terms of latent variables
- A measurement model for recovering a latent variable from a set of indicators

This lecture is about...

Unsupervised measurement with

1. Input: ***categorical*** indicators on nominal or ordinal scales
2. Output: a ***continuous*** measure in form of a scale

Item Response Theory (IRT)

- A model for describing variation in a set of *categorical* variables in terms of latent variables
- A model for recovering a latent variable from a set of *categorical* indicators

This Lecture

From Continuous to Categorical

Item Response Theory (IRT)

Implementation of IRT

Interpreting IRT and Further Examples

From Continuous to Categorical

Working Example

How do countries vote in the United Nation General Assembly?

Data from "Clashes in the Assembly" by Erik Voeten (2000) who investigates "the dimensionality and stability of global conflict as well as the substantive content of the voting alignments that have replaced the Cold War East-West dimension". It includes all adopted resolutions put to a roll-call vote at the UNGA between 1946-88 and 1991-96.[1]

- **Observations**: 60 countries during the early period of the United Nations (1946-52)
- **Indicators**: 348 roll call votes, excluding unanimous votes
  - *Binary* response: Yes vs No (Abstentions treated as missing here)[2]

[1] Data is available within the `R` package `unvotes`.

[2] Similar results if using ordinal outcome yes/abstain/no.

Roll call votes as indicators

The patterns of voting in the UN General Assembly *might* tell us something about international politics.

- Which countries tend to vote together?
- Which countries tend to be in conflict?

$\Rightarrow$ Each time there we see a country vote it is an **indicator** of **something** about that country. (but what?)

Supervised measurement?

1. *Conceptualisation*:
   - Pick/define a concept that we want to measure, e.g. support for the USSR; support for the US; or support for Israel
2. *Measurement*:
   - Observe/select the set of relevant votes
3. *Aggregation*:
   - Combine the votes according to how (we determined) they relate to the concept we want to measure, i.e. according to specified functional form and weights

Advantage: supervised measurement would ensure that we measured the particular thing we wanted to measure

Disadvantage: more work!

- Need to pick *which* votes are relevant to the concept
- Need to determine *how* each vote is related to the concept

Unsupervised measurement?

Use all the votes available and put them into a model that aims to provide a simple description of variation all while explaining as much variation as possible.

- See what comes out and try to make sense of it!

Could pretend that the indicators are continuous and use factor analysis...

... but it might make sense to use a different approach that better reflects the binary/ordinal data type

- This is factor model analogue of moving from linear regression to the various limited dependent variable regression models (binary logistic, ordinal logistic, etc)!

Item Response Theory (IRT)

$=$ a *class* of factor-analysis-like models for limited dependent variables[3]

- Today, we will focus on a couple of the most widely used ones:
  - Logistic Item Response Theory model
  - Ordinal Item Response Theory model
- Today, we will just focus on models with a single latent factor θ_i for each unit
  - As opposed to last week where we looked at a factor analysis model where each observation i has q latent factors $\theta_i = (\theta_{i1}, \theta_{i2}, \ldots, \theta_{iq})$.

[3] Remember: analogous to moving from linear regression to logistic regression!

For a one factor model, we assume that the observed items/indicators I_{ij} are related to the latent factor θ_i in the following way:

$$I_{ij} = \alpha_j + \beta_j \theta_i + \delta_{ij}$$

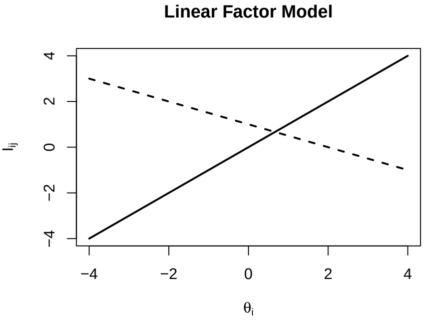
where

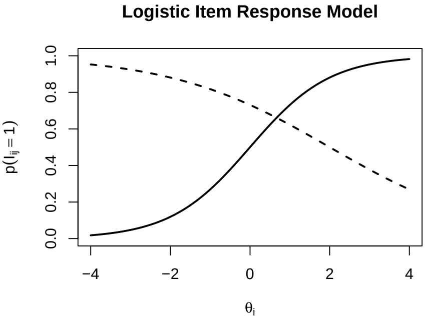
- i refers to different units/individuals/observations
- j refers to different indicators for the units
- no need for third index q, since we only have one latent factor

$\Rightarrow$ The model for each indicator is just a simple, ‘bivariate’ linear model.

Limited Dependent Variables

- If our indicators I_{ij} are *limited* dependent variables, a linear model has the usual limitations:
  - Out of bounds predictions
  - Potentially poor fit for in bounds predictions
- Again, Item Response Theory is to factor analysis what logistic regression (and friends) are to linear regression...





Binary Item Response Model

Take the factor model, and replace the left hand side with the same log-odds transformation that generates a logistic regression from a linear model:

$$I_{ij} = [\ldots] = \alpha_j + \beta_j \theta_i$$

$$\log \left(\frac{p(I_{ij} = 1)}{p(I_{ij} = 0)} \right) = \alpha_j + \beta_j \theta_i$$

These models are sometimes parametrised differently:

$$\log \left(\frac{p(I_{ij} = 1)}{p(I_{ij} = 0)} \right) = \beta_j \left(\theta_i - \alpha_j \right)$$

where α_j can be interpreted as the "*difficulty* parameter" and β_j as the "*discrimination* parameter" for indicator j

Implications of alternative parametrisation

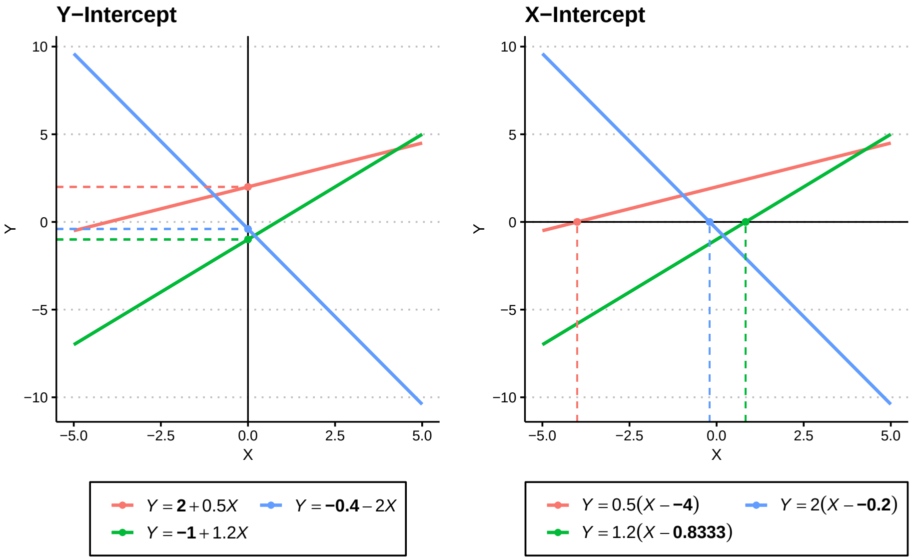
$$\log\left(\frac{p(I_{ij}=1)}{p(I_{ij}=0)}\right) = \beta_j\left(\theta_i - \alpha_j\right)$$

Difficulty parameter α

- Higher values of α then correspond to items with "higher difficulty"
  - Higher values of the latent variable θ_i are required in order to make $I_{ij} = 1$ probable.
- Turns α from a "y intercept" into an "x intercept"
  - α is the value of the the latent variable θ_i at which $p(I_{ij} = 1) =$?[4]

[4] Try to calculate this yourself!

Y vs X intercept



Implications of alternative parametrisation

$$\log\left(\frac{p(I_{ij}=1)}{p(I_{ij}=0)}\right) = \beta_j\left(\theta_i - \alpha_j\right)$$

Discrimination parameter β

- Higher values of β then correspond to items that "discriminate" more between different values of the latent variable θ_i
- This interpretation only makes sense if all/most of the β_j are positive
  - Meaning that higher probabilities of $I_{ij} = 1$ are consistently associated with higher levels of the latent variable θ_i for all/most indicators j.

Binary IRT for Educational Testing

- One application where this is usually true, and where the difficulty/discrimination language comes from, is standardized education testing.
- Standardized test **design** is to have test items (indicators) that cover a range of difficulties, but which all have high, positive discrimination.
- All items should respond positively to the same latent factor ("understanding of the material") but
  - some should be relatively easy (indicating a minimal level of understanding)
  - some should be more difficult (indicating a higher level of understanding).

Ordinal Item Response Model

- We can extend this model to ordinal categorical indicators
  - This model is sometimes called a "graded response model"
  - Just as you can have a test of items that individuals get right $I_{ij} = 1$ or wrong $I_{ij} = 0$, you can also have "graded responses" for any number of ordered levels ($I_{ij} = 1, 2, 3, ...$)
- We now have a model for the log-odds of being above or below *each of the thresholds* in the ordered categorical variable

$$\log\left(\frac{p(I_{ij} > k)}{p(I_{ij} \leq k)}\right) = \beta_j \left(\theta_i - \alpha_{jk}\right)$$

- There are now specific intercepts α_{jk} for each level k
  - If there are only two levels for a given indicator, this reduces to the binary response model
  - Since there is one equation for each indicator, they don't all need to have the same scale/the same number of response options

Implementation of IRT

R implementations

- `ltm` package
  - `ltm()` for binary IRT model
  - `grm()` for ordinal IRT model
  - Uses $\beta_j \left(\theta_i - \alpha_{jk}\right)$ parametrization
  - Estimation by EM algorithm
  - Tends to work well for data sets with few items, many units (not ideal in UN case)
  - This is the one you will learn to use in the seminar
- `MCMCpack` package
  - `MCMCirt1d` for binary IRT model
  - `MCMCordfactanal` for ordinal IRT model
  - Uses $\beta_j \theta_i - \alpha_{jk}$ parametrization
  - Estimation by MCMC algorithm
  - Tends to work well for data sets with many items, many units
  - This is the one I have used for the UN votes (but which you don't need to learn)

The α and β Estimates

$$\log\left(\frac{p(I_{ij}=1)}{p(I_{ij}=0)}\right) = \beta_j \theta_i - \alpha_j$$

- The α_j estimates, of which there are ?, as many as there are ?

```r
summary(alpha_est)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -3.7149 -1.7928 -0.5561 -0.4497  0.7842  3.4500
```

- The β_j estimates, of which there are ?, as many as there are ?

```r
summary(beta_est)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -4.5068 -1.1337  0.5710  0.5012  2.2323  4.5654
```

Predicted Probabilities

$$p(I_{ij} = 1) = \frac{e^{\beta_j \theta_i - \alpha_j}}{1 + e^{\beta_j \theta_i - \alpha_j}}$$

- With the α_j's and β_j's, we can now calculate the predicted probability for a country with a given θ value to vote "yes".
  - For example, for vote with Roll Call ID 12 and a θ value of 3

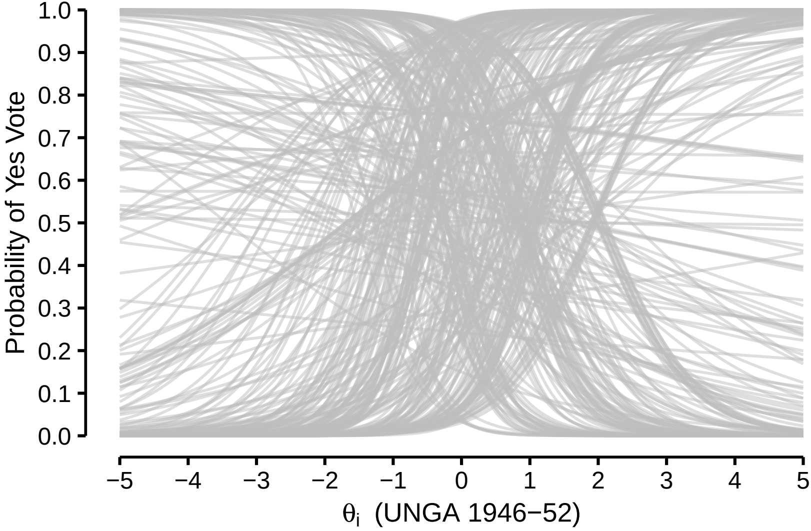
```
exp(-alpha_est["12"] + beta_est["12"]*3)/
  (1+exp(-alpha_est["12"] + beta_est["12"]*3))
```

```
# or
arm::invlogit(-alpha_est["12"] + beta_est["12"]*3)
```

```
##        12
## 0.1497691
```

- In fact, we can calculate the predicted probabilities of voting "yes" for every vote at any value of θ!

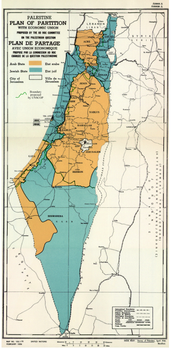
Item Response Curves



Code for Figure

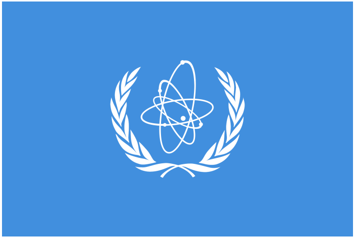
```
# creates a data frame with predicted probabilities for theta values
# from -5 to 5 for each vote
out <- data.frame("voteID"=NA,"theta"=NA,"predprob"=NA)
for (i in 1:length(beta_est)){
  theta <- seq(-5,5,0.01)
  voteID <- rep(names(beta_est)[i], length(theta))
  predprob <- arm::invlogit(-alpha_est[i] + beta_est[i]*(theta))
  out <- rbind(out,data.frame(voteID, theta, predprob))
}
out <- out[-1,]

p <- ggplot(out, aes(x=theta,y=predprob,group=voteID)) +
  geom_line(alpha=.2) +
  scale_x_continuous(expression(theta[i]~" (UNGA 1946-52)"),breaks = -5:5) +
  scale_y_continuous("Probability of Yes Vote",breaks = seq(0,1,0.1)) +
  theme_clean() +
  lemon::coord_capped_cart(bottom = "both", left = "both") +
  theme(plot.background = element_rect(color=NA),
        panel.grid.major = element_blank())
p
```



UN Partition plan for Palestine

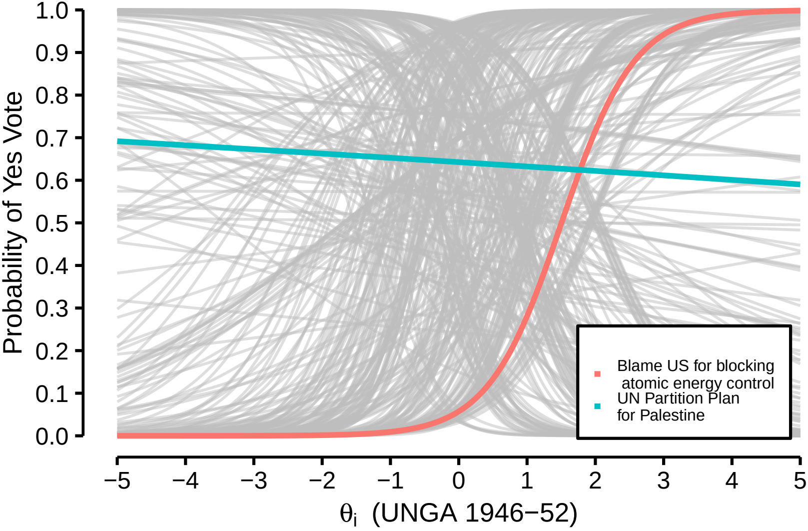
- Roll call ID 77, November 1947
- *"To adopt the report and resolutions A and B (A/516) of the ad hoc comm. on the Palestine question, providing and approving a plan of partition with economic union, for Palestine."*



UN International Atomic Energy Commission

- Roll call ID 198, November 1949
- *"To adopt Paragraph 1 of the USSR draft resolution (A/1120) on the international control of atomic energy."*
- *"...that full responsibility for failure to give effect to the aforesaid resolutions of the General Assembly rests entirely with the Governments of the United States of America and the United Kingdom."*

Selected UN General Assembly Votes



Code for Figure

```r
out$sel <- ifelse(
  out$voteID == "77","UN Partition Plan \nfor Palestine",
  ifelse(
    out$voteID == "198", "Blame US for blocking\n atomic energy control",
    NA))

p + geom_line(aes(color=sel), data = out[!is.na(out$sel),],size=1) +
  theme(legend.title = element_blank(),
        legend.text = element_text(size = 6),
        legend.position = c(0.8,0.16),
        legend.key.size = unit(1,"mm"))
```

- Equation for Roll call ID 77 - Partition of Palestine
  - Small magnitude discrimination parameter β

$$\log\left(\frac{p(I_i = 1)}{p(I_i = 0)}\right) = 0.59 + -0.04 \cdot \theta_i$$

- Equation for Roll call ID 198 - USSR blaming US for blocking control of atomic energy
  - Large magnitude discrimination parameter β

$$\log\left(\frac{p(I_i = 1)}{p(I_i = 0)}\right) = -2.8 + 1.86 \cdot \theta_i$$

The Latent Factor θ

$$\log\left(\frac{p(I_{ij}=1)}{p(I_{ij}=0)}\right) = \beta_j \boldsymbol{\theta_i} - \alpha_j$$

- The θ_i estimates, of which there are ?, as many as there are ?

```r
summary(theta_est)
```

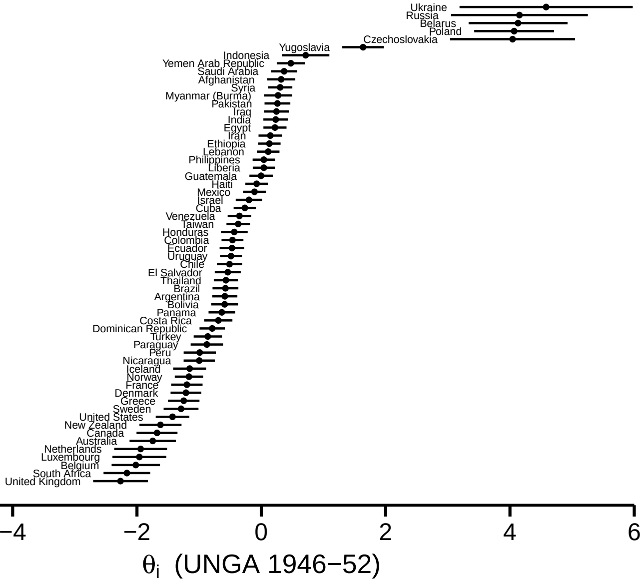
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -2.2652 -1.0372 -0.4681 -0.1675  0.2229  4.5823
```

- The standard errors for each θ_i

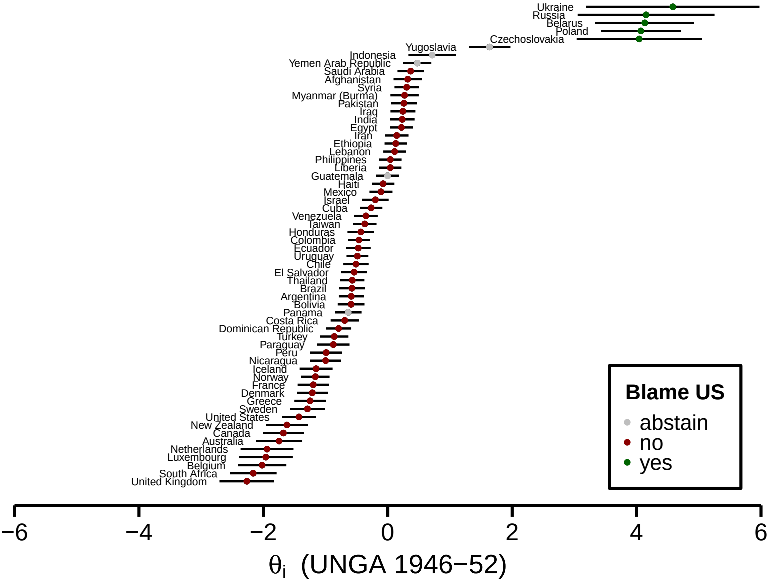
```r
summary(theta_se)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.08979 0.09980 0.11268 0.15574 0.14937 0.71089
```

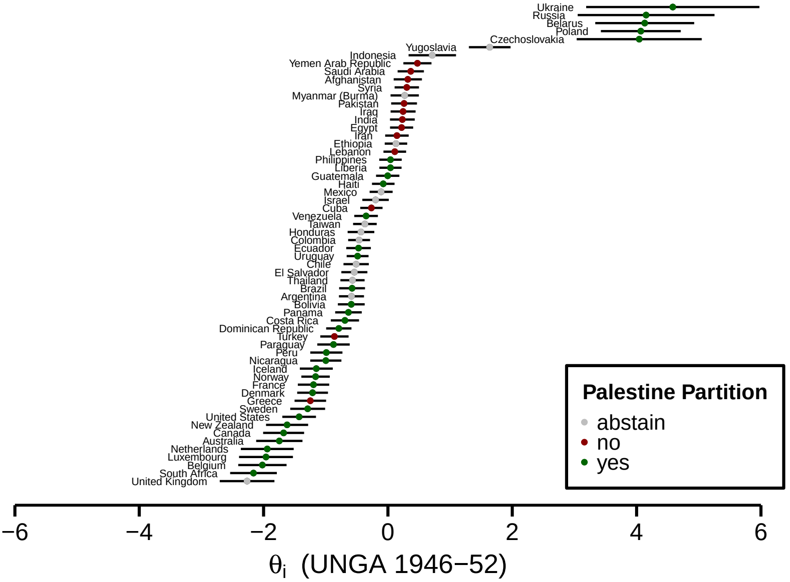
Which countries are where?



Which countries voted to blame the US?



Which countries voted for Palestine partition plan?



Code for Figures

```
thetas <- data.frame(
  "est"=theta_est,"se"=theta_se,
  "hi" = theta_est+1.96*theta_se,
  "lo" = theta_est-1.96*theta_se,
  "cntry"=names(theta_est),
  "rcid77" = roll_call_matrix$`77`,
  "rcid198" = roll_call_matrix$`198`)

p <- ggplot(thetas, aes(x=est, y=fct_reorder(cntry,est))) +
  geom_point() +
  geom_linerange(aes(xmin=lo,xmax=hi)) +
  geom_text(aes(label=cntry),hjust = 1) +
  xlab(expression(theta[i]~" (UNGA 1946-52)")) +
  ylab(NULL) +
  theme_clean() +
  lemon::coord_capped_cart(bottom = "both") +
  theme(plot.background = element_rect(color=NA),
        panel.grid.major.y = element_blank(),
        axis.line.y = element_blank(),
        axis.text.y = element_blank(),
        axis.ticks.y = element_blank(),
        legend.position = c(0.85,0.15))

p + geom_point(aes(color = rcid198),size=.5) +
  scale_color_manual("Blame US",values = c("gray","darkred","darkgreen"))

p + geom_point(aes(color = rcid77),size=.5) +
  scale_color_manual("Palestine Partition",values = c("gray","darkred","darkgreen"))
```

What have we measured?

- Generally speaking, we measure a *Cold War Alignment* scale that varies from aligned with the US (negative values) to aligned with the USSR (positive values).
- The USSR's blame the US amendment was only supported by the core Soviet bloc
  - This vote is a strong indicator of the general scale, and distinguishes between the most strongly USSR aligned and everyone else.
  - There are other votes which provide discrimination at other points on the scale.
- The UN's plan to partition Palestine is a weak indicator of the concept of Cold War Alignment.
  - The countries that voted against, most of which are countries with Muslim majority populations, are in the middle of the *Cold War Alignment* scale.

Interpreting IRT and Further Examples

Summary

- Like factor analysis models, item response models...
  - ...are unsupervised, they measure whatever dimension(s) best predict variation in the indicators.
  - ...are latent variable models, which hypothesize that the indicators reflect an underlying unknown that we want to measure.
- In this application...
  - We can put a label like *Cold War Alignment* on the measured dimension, but remember that is a label we made up.
  - Not all votes are highly predictive indicators of this dimension.
- From binary to ordinal...
  - The analysis above was using a binary item response model, treating abstentions as missing data.
  - You can also apply a "graded response model" to ordinal data, see the textbook and assignment.

General strategies for interpreting IRT models

- Check extremes of unit-level latent variable θ_i
  - Which units are at the ends of the scale?
- Check extreme and near zero cases of the discrimination parameters.
  - Which items have responses that are most strongly (positively or negatively) related to where a unit is on the scale?
  - Which items have responses that are mostly (linearly) unrelated to where a unit is on the scale?
- These models "automatically" determine which items are positively versus negatively associated with the latent variable, so it does not matter if you code all your indicators in the same direction.
  - But the overall sign of the scale is not identified by the data, it is arbitrary and you can reverse it if you want to.

Comparison to other applications

- Lecture
  - Votes in the UNGA (1946-52)
  - Roll call data set, many respondents, many items
- Textbook
  - Battery of 12 UK political ideology questions
  - Survey data set, many respondents, few items
- Assignment
  - Battery of 6 UK political knowledge questions
  - Survey data set, many respondents, few items

Some more complexities

- Multinomial Item Response Model
  - Models exist for using unordered categorical indicators
  - Somewhat rare to have data of this form though.
- Mixed Indicator Item Response Models
  - It is possible to specify "mixed response" factor models which combine both continuous and categorical indicators.
  - The factors (latent variables) remain continuous, and predict both types of indicators

Frequent Applications

- Psychological Testing
  - See textbook example
- Educational Testing
  - Item response models are used to *design* education tests, but tests are usually scored by counting correct answers
  - When testing potential test questions, you want items that are similarly responsive to the latent dimension, but vary in difficulty
- Political Preferences
  - Item response models are used to model how different kinds of political responses reflect underlying political preference dimensions
  - Votes in legislatures (Clinton et al 2004), decisions by judges (Martin and Quinn 2002), survey responses of citizens (Bafumi et al 2010), etc
  - House of Commons voting in the UK is poorly approximated by these models because of very strong party discipline (Spirling and McLean 2007)

Remember!

"Unsupervised" measurement methods will discover whatever latent factors explain the most variation in your indicators, not necessarily the latent factors that you want them to discover.

```r
print("Computer Lab Session 10...")
```

```
## [1] "Computer Lab Session 10..."
```

- In this week's computer lab, we will analyse political knowledge questions from the British Election Study.