

# Quantitative Social Science Methods, I, Lecture Notes: Anchoring Vignettes for Interpersonally Incomparable Survey Responses

Gary King<sup>1</sup>  
Institute for Quantitative Social Science  
Harvard University

August 17, 2020

---

<sup>1</sup>GaryKing.org



# Readings

- “Enhancing the Validity and Cross-cultural Comparability of Measurement in Survey Research” Gary King; Christopher J.L. Murray; Joshua A. Salomon; and Ajay Tandon, *American Political Science Review*, 2004
- “Comparing Incomparable Survey Responses: Evaluating and Selecting Anchoring Vignettes” Gary King and Jonathan Wand, *Political Analysis*, 2007
- “Improving Anchoring Vignettes: Designing Surveys to Correct Interpersonal Incomparability” Daniel Hopkins and Gary King, *Public Opinion Quarterly*, 2010
- “Engaging Private and Information Technology in a Digital Age” National Research Council, [j.mp/PrivInfoAge](http://j.mp/PrivInfoAge), 2007: p.86ff

[GaryKing.org/vign](http://GaryKing.org/vign)

# Anchoring Vignettes Address Two Problems

- Big concepts we can define only by example
  - **Illustrations:** Freedom, efficacy, pornography, health, compactness, democracy, intelligence, ...
  - **The usual advice:** You don't have a methodological problem. Get a theory and we'll know what to measure.
  - **Result of more concreteness:** more reliability, no more validity
- Incomparability in Survey Responses (“DIF”)
  - Chinese report having more political efficacy than Americans
  - The most common measure of health — “How healthy are you? (Excellent, Good, Fair, Poor)” — often correlates negatively with actual health (Amaryta Sen, 2002)
  - Ethiopians and Danes report being equally healthy
  - “Individuals understand the ‘same’ questions in vastly different ways.” (Brady, 1985)

# Anchoring Vignettes: Political Efficacy Through Voting

- “[Alison] lacks clean drinking water. She and her neighbors are supporting an opposition candidate in the forthcoming elections that has promised to address the issue. It appears that so many people in her area feel the same way that the opposition candidate will defeat the incumbent representative.”
- “[Jane] lacks clean drinking water because the government is pursuing an industrial development plan. In the campaign for an upcoming election, an opposition party has promised to address the issue, but she feels it would be futile to vote for the opposition since the government is certain to win.”
- “[Moses] lacks clean drinking water. He would like to change this, but he can’t vote, and feels that no one in the government cares about this issue. So he suffers in silence, hoping something will be done in the future.”

How much say [does ‘name’ / do you] have in getting the government to address issues that interest [him / her / you]?

(a) Unlimited say, (b) A lot of say, (c) Some say, (d) Little say, (e) No say at all

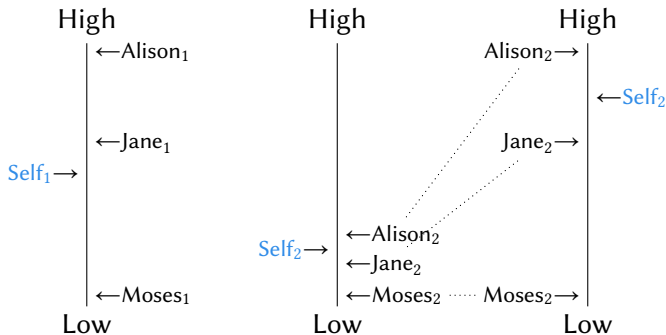
Introduction

A Nonparametric Method

A Parametric Method

Illustrations

# Who has More Political Efficacy: $R_1$ or $R_2$ ?



- Quiz: What does *varying vignette assessments* tell us? DIF
- Quiz: How do we know true vignette assessments are fixed over respondents? We create the anchors (Alison, Jane, Moses)!

# Formalizing A Nonparametric Methodology

- Define self-assessment answers *relative* to vignettes answers.
- For respondents who rank vignettes,  $z_{i1} < z_{i2} < \dots < z_{iJ}$ ,

$$C_i = \begin{cases} 1 & \text{if } y_i < z_{i1} \\ 2 & \text{if } y_i = z_{i1} \\ 3 & \text{if } z_{i1} < y_i < z_{i2} \\ \vdots & \vdots \\ 2J + 1 & \text{if } y_i > z_{iJ} \end{cases}$$

- Apportion  $C$  equally among tied vignette categories
- (This is wrong, but simple; we will improve shortly)
- Treat vignette ranking inconsistencies as ties
- Requires vignettes and self-assessments asked of all
- (A parametric method to come doesn't)



# Measurement Assumptions

- **Response Consistency**
  - Each respondent uses the self-assessment and vignette categories in approximately the same way across questions
  - I.e., DIF occurs across respondents, not across questions for any one respondent
- **Vignette Equivalence**
  - The actual level for any vignette is the same for all respondents
  - The quantity being estimated exists
  - The scale being tapped is perceived as unidimensional
- **In other words:** we allow response-category DIF but assume stem question equivalence

# Comparing Political Efficacy in Mexico and China

## Mexico



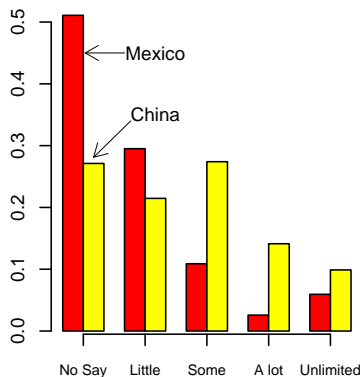
Opposition leader Vicente Fox elected president. 71-year rule of PRI Party ends. Peaceful transition of power begins.

## China

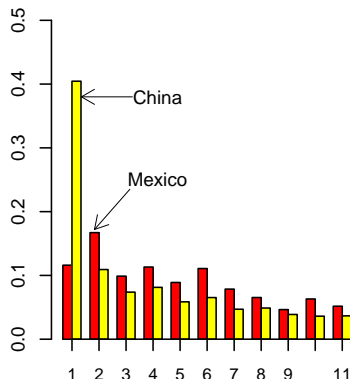


How much say do you have getting government to address issues that interest you?

# Nonparametric Estimates of Political Efficacy



observed categorical  
self-assessments



C, our nonparametric  
DIF-corrected estimate of the  
same distribution

# Ties and Inconsistencies Produce *Ranges*

|                         | Survey Responses | 1:<br>$y < z_1$ | 2:<br>$y = z_1$ | 3:<br>$z_1 < y < z_2$ | 4:<br>$y = z_2$ | 5:<br>$y > z_2$ | C           |
|-------------------------|------------------|-----------------|-----------------|-----------------------|-----------------|-----------------|-------------|
| <b>Scalars:</b>         |                  |                 |                 |                       |                 |                 |             |
|                         | $y < z_1 < z_2$  | T               |                 |                       |                 |                 | {1}         |
|                         | $y = z_1 < z_2$  |                 | T               |                       |                 |                 | {2}         |
|                         | $z_1 < y < z_2$  |                 |                 | T                     |                 |                 | {3}         |
|                         | $z_1 < y = z_2$  |                 |                 |                       | T               |                 | {4}         |
|                         | $z_1 < z_2 < y$  |                 |                 |                       |                 | T               | {5}         |
| <b>Ties:</b>            |                  |                 |                 |                       |                 |                 |             |
|                         | $y < z_1 = z_2$  | T               |                 |                       |                 |                 | {1}         |
|                         | $y = z_1 = z_2$  |                 | T               |                       | T               |                 | {2,3,4}     |
|                         | $z_1 = z_2 < y$  |                 |                 |                       |                 | T               | {5}         |
| <b>Inconsistencies:</b> |                  |                 |                 |                       |                 |                 |             |
|                         | $y < z_2 < z_1$  | T               |                 |                       |                 |                 | {1}         |
|                         | $y = z_2 < z_1$  | T               |                 |                       | T               |                 | {1,2,3,4}   |
|                         | $z_2 < y < z_1$  | T               |                 |                       |                 | T               | {1,2,3,4,5} |
|                         | $z_2 < y = z_1$  |                 | T               |                       |                 | T               | {2,3,4,5}   |
|                         | $z_2 < z_1 < y$  |                 |                 |                       |                 | T               | {5}         |

# A Likelihood for Scalar or Vector Responses

Parametric likelihood used in a nonparametric methodology

- Define an unobserved variable:  $Y_i \sim \text{Normal}(x_i\beta, 1)$
- With observation mechanism, for scalar  $C$ , same as ordered probit:  $C_i = c$  if  $\tau_{c-1} \leq Y_i < \tau_c$
- Probability of observing category  $c$ , for  $X = x_0$ :

$$\Pr(C = c|x_0) = \int_{\tau_{c-1}}^{\tau_c} \text{Normal}(y|x_0\beta, 1)dy$$

- Observation mechanism for vector valued  $C$ :

$$C_i = c \quad \text{if } \tau_{\min(c)-1} \leq Y_i < \tau_{\max(c)}$$

- Quiz: Do you know how to form the likelihood function?

Introduction

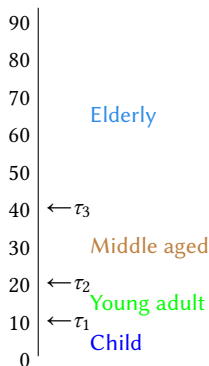
A Nonparametric Method

A Parametric Method

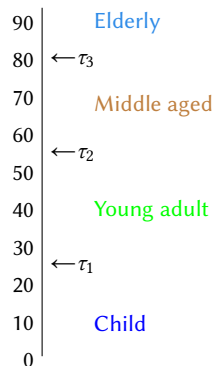
Illustrations

# Categorizing Years of Age

Respondent 1

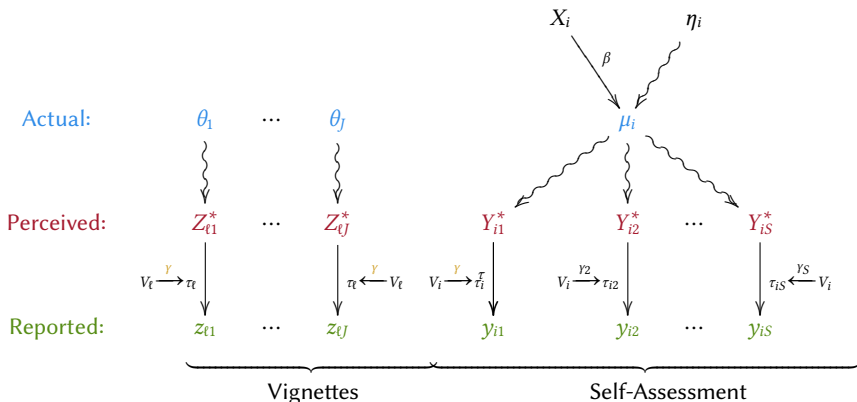


Respondent 2



- If thresholds vary: categorical answers are meaningless
- Quiz: In Ordered Logit, can we subscript  $\tau_i$  by person?
- Quiz: What information would identify the  $\tau_i$ 's?
- A parametric model works by estimating the thresholds

# Statistical Model: Visual Summary



- An ordinal probit model,
- with varying thresholds,
- a vignette for identification (with the same  $\gamma$ ),
- more vignettes for better discrimination,
- optional multiple self-assessment questions for more efficiency,
- and an optional random effect.



## Self-Assessment Model Component ( $i = 1, \dots, n$ )

- **Actual:**  $\mu_i = X_i\beta + \eta_i$ , with random effect  $\eta_i \sim N(0, \omega^2)$
- **Perceived:**  $Y_{i1}^* \sim N(\mu_i, 1) \quad \dots \quad Y_{iS}^* \sim N(\mu_i, 1)$
- **Reported** (observation mechanism):

$$y_{i1} = k \quad \text{if } \tau_{i1}^{k-1} \leq Y_{i1}^* < \tau_{i1}^k$$

$\vdots$

$$y_{iS} = k \quad \text{if } \tau_{iS}^{k-1} \leq Y_{iS}^* < \tau_{iS}^k$$

- where  $\tau_{is}^1 = \gamma_1 V_i$ ,  $\tau_{is}^k = \tau_{is}^{k-1} + e^{\gamma_k V_i}$ , ( $k = 2, \dots, K_s$ )

## Vignette Model Component ( $\ell = 1, \dots, N$ )

- **Actual:**  $\theta_1, \dots, \theta_J$
- **Perceived:**  $Z_{\ell 1}^* \sim N(\theta_1, \sigma^2) \quad \dots \quad Z_{\ell J}^* \sim N(\theta_J, \sigma^2)$
- **Reported:**  $z_{\ell j} = k \quad \text{if } \tau_{\ell 1}^{k-1} \leq Z_{\ell j}^* < \tau_{\ell 1}^k$ ,
- where  $\tau_{\ell s}^1 = \gamma_1 V_\ell$ ,  $\tau_{\ell s}^k = \tau_{\ell s}^{k-1} + e^{\gamma_k V_\ell}$  ( $k = 2, \dots, K_s$ )

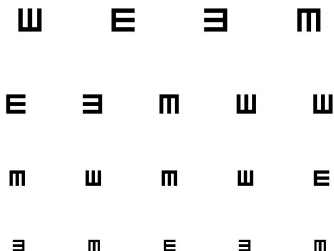


# Self-Assessments v. Medical Tests

**Self-Assessment:** In the last 30 days, how much difficulty did [you/name] have in seeing and recognizing a person you know across the road (i.e. from a distance of about 20 meters)? (A) none, (B) mild, (C) moderate, (D) severe, (E) extreme/cannot do

The Snellen Eye Chart Test:

20/60



## Fixing DIF in Self-Assessments of Visual (Non)acuity

|            | Snellen Eye Chart |        | Ordinal Probit |        | Chopit |        |
|------------|-------------------|--------|----------------|--------|--------|--------|
|            | Mean              | (s.e.) | $\mu$          | (s.e.) | $\mu$  | (s.e.) |
| Slovakia   | 8.006             | (.272) | .660           | (.127) | .286   | (.129) |
| China      | 10.780            | (.148) | .673           | (.073) | .749   | (.081) |
| Difference | -2.774            | (.452) | -.013          | (.053) | -.463  | (.053) |

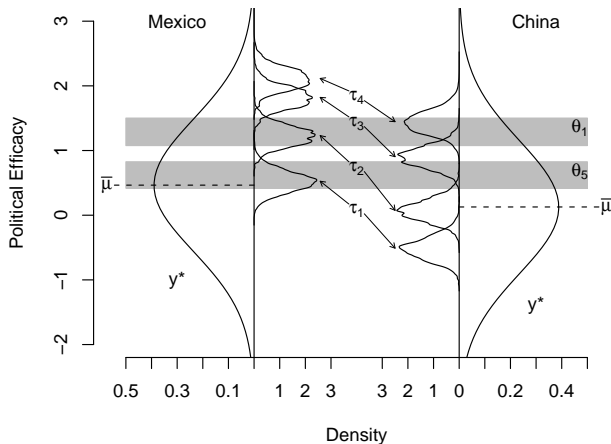
- **Medical test:** Slovaks see much better than the Chinese
- **Ordinal probit:** no difference
- **Chopit:** same result as the medical test (on a different scale)
- (Better to calculate QOIs! The articles explain how.)

## Fixing DIF in China and Mexico

| Eqn.      | Variable     | Ordinal Probit |        | Chopit |        |
|-----------|--------------|----------------|--------|--------|--------|
|           |              | Coeff.         | (s.e.) | Coeff. | (s.e.) |
| $\mu$     | China        | .670           | (.081) | -.362  | (.090) |
|           | age          | .004           | (.003) | .006   | (.003) |
|           | male         | .087           | (.076) | .113   | (.081) |
|           | education    | .020           | (.008) | .019   | (.008) |
| Vignettes | $\theta_1$   |                |        | 1.393  | (.190) |
|           | $\theta_2$   |                |        | 1.304  | (.190) |
|           | $\theta_3$   |                |        | .953   | (.189) |
|           | $\theta_4$   |                |        | .902   | (.188) |
|           | $\theta_5$   |                |        | .729   | (.188) |
|           | $\ln \sigma$ |                |        | -.238  | (.042) |

- **Ordered Probit:** Chinese people have *more* political efficacy than Mexicans
- **Chopit:** Chinese people have *less* political efficacy than Mexicans

# DIF as Threshold Variation: Mexico v. China



- **Self-assessments:** Actual (mean),  $\mu$ ; Perceived  $y^*$
- **Vignette Assessments:**  $\theta_1, \theta_2$
- **Thresholds:** Chinese have lower standards for satisfying  $\tau_j$

## Summary: Anchoring Vignettes to fix DIF

- **Assumptions:** response consistency, vignette equivalence
- **Probably won't eliminate all DIF:** but problems would have to be unrealistically big to make unadjusted measures better than the adjusted ones
- **Reduce expense:** assign each vignette to subsample. E.g., 4 vignettes asked for 1/4 of the sample
- **If you think you have DIF-free questions:** you now have the first real opportunity to test that hypothesis.
- **Whether or not you have DIF:** vignettes can help us follow standard advice of making questions concrete
- **Writing vignettes:** aids in clarification and discovery of domains of concept of interest — even if you do not do a survey.
- **Other common survey problems:** Question wording, Accurate translation, Question order, Sampling design, Interview length, Social backgrounds of interviewer and respondent
- **Other uses:** Defining concepts you only know when you see