What to do about Biases in Survey Research

Gary King

http://GKing.Harvard.Edu

November 15, 2013

- Gary King and Jonathan Wand. "Comparing Incomparable Survey Responses: Evaluating and Selecting Anchoring Vignettes," *Political Analysis*, 15, 1 (Winter, 2007): 46–66.

Readings

- Gary King and Jonathan Wand. "Comparing Incomparable Survey Responses: Evaluating and Selecting Anchoring Vignettes," *Political Analysis*, 15, 1 (Winter, 2007): 46–66.

- Gary King; Christopher J.L. Murray; Joshua A. Salomon; and Ajay Tandon. "Enhancing the Validity and Cross-cultural Comparability of Measurement in Survey Research," *American Political Science Review*, Vol. 98, No. 1 (February, 2004): 191–207.

Readings

- Gary King and Jonathan Wand. "Comparing Incomparable Survey Responses: Evaluating and Selecting Anchoring Vignettes," *Political Analysis*, 15, 1 (Winter, 2007): 46–66.

- Gary King; Christopher J.L. Murray; Joshua A. Salomon; and Ajay Tandon. "Enhancing the Validity and Cross-cultural Comparability of Measurement in Survey Research," *American Political Science Review*, Vol. 98, No. 1 (February, 2004): 191–207.

- Papers, FAQ, examples, software, conferences, videos: http://GKing.Harvard.edu/vign

Examples of the Problem

Examples of the Problem

Presidential Approval: the longest public opinion time series

Examples of the Problem

Presidential Approval: the longest public opinion time series

- On 9/10/2001, 55% of Americans approved of the way George W. Bush was "handling his job as president".

Examples of the Problem

Presidential Approval: the longest public opinion time series

- On 9/10/2001, 55% of Americans approved of the way George W. Bush was "handling his job as president".

- The next day — which the president spent in hiding — 90% approved.

Examples of the Problem

Presidential Approval: the longest public opinion time series

- On 9/10/2001, 55% of Americans approved of the way George W. Bush was "handling his job as president".
- The next day — which the president spent in hiding — 90% approved.
- Was this massive opinion change, or was the same question interpreted differently?

Examples of the Problem

The O.J. Simpson trial: most publicized murder trial in history

Examples of the Problem

The O.J. Simpson trial: most publicized murder trial in history

- The facts of the case seemed clear

Examples of the Problem

The O.J. Simpson trial: most publicized murder trial in history

- The facts of the case seemed clear
- Did he do it? Whites: 62% say "yes". Blacks: 14% say "yes".

Examples of the Problem

The O.J. Simpson trial: most publicized murder trial in history

- The facts of the case seemed clear
- Did he do it? Whites: 62% say "yes". Blacks: 14% say "yes".
- Did black and white Americans have genuinely opposing views about whether Simpson committed murder, or did the two groups interpret the same survey question differently?

Examples of the Problem

Examples of the Problem

The most common measure of the health of populations: "How healthy are you? Excellent, Good, Fair, or Poor"

Examples of the Problem

The most common measure of the health of populations: "How healthy are you? Excellent, Good, Fair, or Poor"

- Suppose an otherwise healthy 25-year-old woman with a cold and a backache answers "fair" and a 90-year-old man just able to get out of bed says "excellent"

Examples of the Problem

The most common measure of the health of populations: "How healthy are you? Excellent, Good, Fair, or Poor"

- Suppose an otherwise healthy 25-year-old woman with a cold and a backache answers "fair" and a 90-year-old man just able to get out of bed says "excellent"
- Is the young woman less healthy or are the two interpreting the same question differently?

Examples of the Problem

The most common measure of the health of populations: "How healthy are you? Excellent, Good, Fair, or Poor"

- Suppose an otherwise healthy 25-year-old woman with a cold and a backache answers "fair" and a 90-year-old man just able to get out of bed says "excellent"

- Is the young woman less healthy or are the two interpreting the same question differently?

- In some countries, responses to this survey question correlate negatively with objective measures of health status (Sen, 2002).
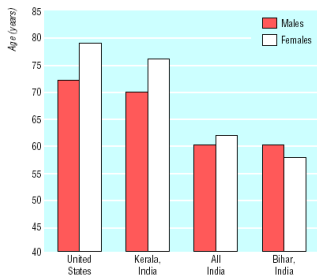
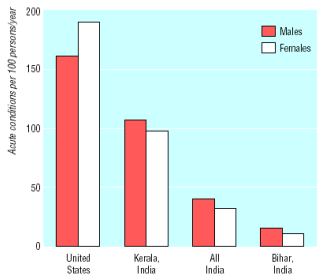Fig 1 Life expectancy among males and females in India compared with United States, mid-1990s[7 a]

Fig 2 Incidence of reported morbidity in India, mid-1970s, compared with United States, mid-1980s[8 9]

Anchoring Vignettes & Self-Assessments: Political Efficacy (about voting)

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

- "[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."

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

- "[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."

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

- "[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

- The only reason for vignette assessments to change over respondents is DIF
- Assumption holds because investigator creates the anchors (Alison, Jane, Moses)
- Our simple (nonparametric) method works this way.

A Simple, Nonparametric Method

A Simple, Nonparametric Method

- Define self-assessment answers *relative* to vignettes answers.

A Simple, Nonparametric Method

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

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

A Simple, Nonparametric Method

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

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

- Apportion C equally among tied vignette categories

A Simple, Nonparametric Method

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

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

- Apportion C equally among tied vignette categories
- (This is wrong, but simple; we will improve shortly)

A Simple, Nonparametric Method

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

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

- Apportion C equally among tied vignette categories
- (This is wrong, but simple; we will improve shortly)
- Treat vignette ranking inconsistencies as ties

A Simple, Nonparametric Method

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

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

- 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 respondents

A Simple, Nonparametric Method

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

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

- 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 respondents
- (Our parametric method doesn't)

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

Plenty of political efficacy

China: How much say do you have in getting the government to address issues that interest you?

Nonparametric Estimates of Political Efficacy
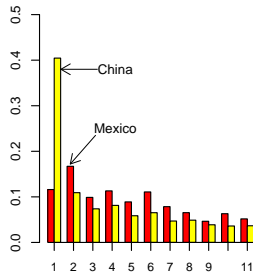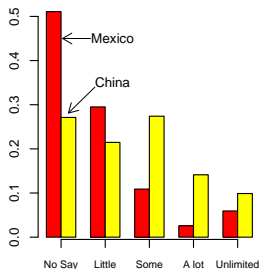
- The left graph is a histogram of the observed categorical self-assessments.
- The right graph is a histogram of C, our nonparametric DIF-corrected estimate of the same distribution.

Key Measurement Assumptions

Key Measurement Assumptions

1. *Response Consistency*: Each respondent uses the self-assessment and vignette categories in approximately the same way across questions.

Key Measurement Assumptions

1. *Response Consistency*: Each respondent uses the self-assessment and vignette categories in approximately the same way across questions. (DIF occurs across respondents, not across questions for any one respondent.)

Key Measurement Assumptions

1. *Response Consistency*: Each respondent uses the self-assessment and vignette categories in approximately the same way across questions. (DIF occurs across respondents, not across questions for any one respondent.)
2. *Vignette Equivalence*:

Key Measurement Assumptions

1. *Response Consistency*: Each respondent uses the self-assessment and vignette categories in approximately the same way across questions. (DIF occurs across respondents, not across questions for any one respondent.)
2. *Vignette Equivalence*:
    (a) The actual level for any vignette is the same for all respondents.

Key Measurement Assumptions

1. *Response Consistency*: Each respondent uses the self-assessment and vignette categories in approximately the same way across questions. (DIF occurs across respondents, not across questions for any one respondent.)

2. *Vignette Equivalence*:
   (a) The actual level for any vignette is the same for all respondents.
   (b) The quantity being estimated exists.

Key Measurement Assumptions

1. *Response Consistency*: Each respondent uses the self-assessment and vignette categories in approximately the same way across questions. (DIF occurs across respondents, not across questions for any one respondent.)

2. *Vignette Equivalence*:
   (a) The actual level for any vignette is the same for all respondents.
   (b) The quantity being estimated exists.
   (c) The scale being tapped is perceived as unidimensional.

Key Measurement Assumptions

1. *Response Consistency*: Each respondent uses the self-assessment and vignette categories in approximately the same way across questions. (DIF occurs across respondents, not across questions for any one respondent.)

2. *Vignette Equivalence*:
   (a) The actual level for any vignette is the same for all respondents.
   (b) The quantity being estimated exists.
   (c) The scale being tapped is perceived as unidimensional.

3. In other words: we allow response-category DIF but assume stem question equivalence.

Ties and Inconsistencies Produce *Ranges*

Example	Survey Responses	1: $y < z_1$	2: $y = z_1$	3: $z_1 < y < z_2$	4: $y = z_2$	5: $y > z_2$	C

Ties and Inconsistencies Produce *Ranges*

Example	Survey Responses	1: $y < z_1$	2: $y = z_1$	3: $z_1 < y < z_2$	4: $y = z_2$	5: $y > z_2$	C

Ties and Inconsistencies Produce *Ranges*

Example	Survey Responses	1: $y < z_1$	2: $y = z_1$	3: $z_1 < y < z_2$	4: $y = z_2$	5: $y > z_2$	C
1	$y < z_1 < z_2$	T					{1}

Ties and Inconsistencies Produce *Ranges*

Example	Survey Responses	1: $y < z_1$	2: $y = z_1$	3: $z_1 < y < z_2$	4: $y = z_2$	5: $y > z_2$	C
1	$y < z_1 < z_2$	T					$\{1\}$
2	$y = z_1 < z_2$		T				$\{2\}$

Ties and Inconsistencies Produce *Ranges*

Example	Survey Responses	1: $y < z_1$	2: $y = z_1$	3: $z_1 < y < z_2$	4: $y = z_2$	5: $y > z_2$	C
1	$y < z_1 < z_2$	T					$\{1\}$
2	$y = z_1 < z_2$		T				$\{2\}$
3	$z_1 < y < z_2$			T			$\{3\}$

Ties and Inconsistencies Produce *Ranges*

Example	Survey Responses	1: $y < z_1$	2: $y = z_1$	3: $z_1 < y < z_2$	4: $y = z_2$	5: $y > z_2$	C
1	$y < z_1 < z_2$	T					$\{1\}$
2	$y = z_1 < z_2$		T				$\{2\}$
3	$z_1 < y < z_2$			T			$\{3\}$
4	$z_1 < y = z_2$				T		$\{4\}$

Ties and Inconsistencies Produce *Ranges*

Example	Survey Responses	1: $y < z_1$	2: $y = z_1$	3: $z_1 < y < z_2$	4: $y = z_2$	5: $y > z_2$	C
1	$y < z_1 < z_2$	T					$\{1\}$
2	$y = z_1 < z_2$		T				$\{2\}$
3	$z_1 < y < z_2$			T			$\{3\}$
4	$z_1 < y = z_2$				T		$\{4\}$
5	$z_1 < z_2 < y$					T	$\{5\}$

Ties and Inconsistencies Produce *Ranges*

Example	Survey Responses	1: $y < z_1$	2: $y = z_1$	3: $z_1 < y < z_2$	4: $y = z_2$	5: $y > z_2$	C
1	$y < z_1 < z_2$	T					$\{1\}$
2	$y = z_1 < z_2$		T				$\{2\}$
3	$z_1 < y < z_2$			T			$\{3\}$
4	$z_1 < y = z_2$				T		$\{4\}$
5	$z_1 < z_2 < y$					T	$\{5\}$

Ties:

Ties and Inconsistencies Produce *Ranges*

Example	Survey Responses	1: $y < z_1$	2: $y = z_1$	3: $z_1 < y < z_2$	4: $y = z_2$	5: $y > z_2$	C
1	$y < z_1 < z_2$	T					$\{1\}$
2	$y = z_1 < z_2$		T				$\{2\}$
3	$z_1 < y < z_2$			T			$\{3\}$
4	$z_1 < y = z_2$				T		$\{4\}$
5	$z_1 < z_2 < y$					T	$\{5\}$
Ties:							
6	$y < z_1 = z_2$	T					$\{1\}$

Ties and Inconsistencies Produce *Ranges*

Example	Survey Responses	1: $y < z_1$	2: $y = z_1$	3: $z_1 < y < z_2$	4: $y = z_2$	5: $y > z_2$	C
1	$y < z_1 < z_2$	T					$\{1\}$
2	$y = z_1 < z_2$		T				$\{2\}$
3	$z_1 < y < z_2$			T			$\{3\}$
4	$z_1 < y = z_2$				T		$\{4\}$
5	$z_1 < z_2 < y$					T	$\{5\}$
Ties:							
6	$y < z_1 = z_2$	T					$\{1\}$
7	$y = z_1 = z_2$		T		T		$\{2,3,4\}$

Ties and Inconsistencies Produce *Ranges*

Example	Survey Responses	1: $y < z_1$	2: $y = z_1$	3: $z_1 < y < z_2$	4: $y = z_2$	5: $y > z_2$	C
1	$y < z_1 < z_2$	T					$\{1\}$
2	$y = z_1 < z_2$		T				$\{2\}$
3	$z_1 < y < z_2$			T			$\{3\}$
4	$z_1 < y = z_2$				T		$\{4\}$
5	$z_1 < z_2 < y$					T	$\{5\}$
Ties:							
6	$y < z_1 = z_2$	T					$\{1\}$
7	$y = z_1 = z_2$		T		T		$\{2,3,4\}$
8	$z_1 = z_2 < y$					T	$\{5\}$

Ties and Inconsistencies Produce *Ranges*

Example	Survey Responses	1: $y < z_1$	2: $y = z_1$	3: $z_1 < y < z_2$	4: $y = z_2$	5: $y > z_2$	C
1	$y < z_1 < z_2$	T					{1}
2	$y = z_1 < z_2$		T				{2}
3	$z_1 < y < z_2$			T			{3}
4	$z_1 < y = z_2$				T		{4}
5	$z_1 < z_2 < y$					T	{5}
Ties:							
6	$y < z_1 = z_2$	T					{1}
7	$y = z_1 = z_2$		T		T		{2,3,4}
8	$z_1 = z_2 < y$					T	{5}
Inconsistencies:							

Ties and Inconsistencies Produce *Ranges*

Example	Survey Responses	1: $y < z_1$	2: $y = z_1$	3: $z_1 < y < z_2$	4: $y = z_2$	5: $y > z_2$	C
1	$y < z_1 < z_2$	T					$\{1\}$
2	$y = z_1 < z_2$		T				$\{2\}$
3	$z_1 < y < z_2$			T			$\{3\}$
4	$z_1 < y = z_2$				T		$\{4\}$
5	$z_1 < z_2 < y$					T	$\{5\}$
Ties:							
6	$y < z_1 = z_2$	T					$\{1\}$
7	$y = z_1 = z_2$		T		T		$\{2,3,4\}$
8	$z_1 = z_2 < y$					T	$\{5\}$
Inconsistencies:							
9	$y < z_2 < z_1$	T					$\{1\}$

Ties and Inconsistencies Produce *Ranges*

Example	Survey Responses	1: $y < z_1$	2: $y = z_1$	3: $z_1 < y < z_2$	4: $y = z_2$	5: $y > z_2$	C
1	$y < z_1 < z_2$	T					$\{1\}$
2	$y = z_1 < z_2$		T				$\{2\}$
3	$z_1 < y < z_2$			T			$\{3\}$
4	$z_1 < y = z_2$				T		$\{4\}$
5	$z_1 < z_2 < y$					T	$\{5\}$
Ties:							
6	$y < z_1 = z_2$	T					$\{1\}$
7	$y = z_1 = z_2$		T		T		$\{2,3,4\}$
8	$z_1 = z_2 < y$					T	$\{5\}$
Inconsistencies:							
9	$y < z_2 < z_1$	T					$\{1\}$
10	$y = z_2 < z_1$	T			T		$\{1,2,3,4\}$

Ties and Inconsistencies Produce *Ranges*

Example	Survey Responses	1: $y < z_1$	2: $y = z_1$	3: $z_1 < y < z_2$	4: $y = z_2$	5: $y > z_2$	C
1	$y < z_1 < z_2$	T					$\{1\}$
2	$y = z_1 < z_2$		T				$\{2\}$
3	$z_1 < y < z_2$			T			$\{3\}$
4	$z_1 < y = z_2$				T		$\{4\}$
5	$z_1 < z_2 < y$					T	$\{5\}$
Ties:							
6	$y < z_1 = z_2$	T					$\{1\}$
7	$y = z_1 = z_2$		T		T		$\{2,3,4\}$
8	$z_1 = z_2 < y$					T	$\{5\}$
Inconsistencies:							
9	$y < z_2 < z_1$	T					$\{1\}$
10	$y = z_2 < z_1$	T			T		$\{1,2,3,4\}$
11	$z_2 < y < z_1$	T				T	$\{1,2,3,4,5\}$

Ties and Inconsistencies Produce *Ranges*

Example	Survey Responses	1: $y < z_1$	2: $y = z_1$	3: $z_1 < y < z_2$	4: $y = z_2$	5: $y > z_2$	C
1	$y < z_1 < z_2$	T					$\{1\}$
2	$y = z_1 < z_2$		T				$\{2\}$
3	$z_1 < y < z_2$			T			$\{3\}$
4	$z_1 < y = z_2$				T		$\{4\}$
5	$z_1 < z_2 < y$					T	$\{5\}$
Ties:							
6	$y < z_1 = z_2$	T					$\{1\}$
7	$y = z_1 = z_2$		T		T		$\{2,3,4\}$
8	$z_1 = z_2 < y$					T	$\{5\}$
Inconsistencies:							
9	$y < z_2 < z_1$	T					$\{1\}$
10	$y = z_2 < z_1$	T			T		$\{1,2,3,4\}$
11	$z_2 < y < z_1$	T				T	$\{1,2,3,4,5\}$
12	$z_2 < y = z_1$		T			T	$\{2,3,4,5\}$
13	$z_2 < z_1 < y$					T	$\{5\}$

Analyzing the DIF-Free Variable: More Efficiencies

- How to analyze a variable with scalar and vector responses?

- How to analyze a variable with scalar and vector responses?
- We define a new method (censored ordered probit), a direct extension of ordinal probit allowing for ranges of responses

- How to analyze a variable with scalar and vector responses?
- We define a new method (censored ordered probit), a direct extension of ordinal probit allowing for ranges of responses
- Useful for vignettes; also useful for survey questions that allow ranges of responses

Improved Efficiency in Practice
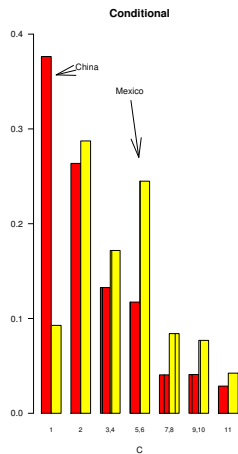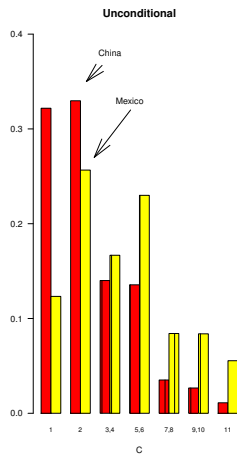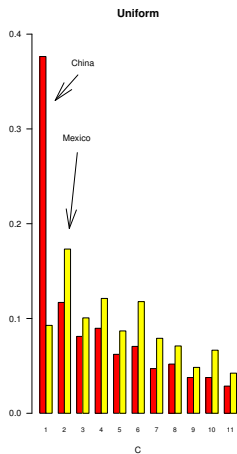
Optimally Choosing Vignettes

Optimally Choosing Vignettes

- **Ultimate Goal:** Define categories with vignettes to learn about a continuous unobserved variable (health, efficacy).

Optimally Choosing Vignettes

- **Ultimate Goal:** Define categories with vignettes to learn about a continuous unobserved variable (health, efficacy).
  - **Worst choice:** All in one category, no discriminatory power

Optimally Choosing Vignettes

- Ultimate Goal: Define categories with vignettes to learn about a continuous unobserved variable (health, efficacy).
  - Worst choice: All in one category, no discriminatory power (E.g., "Bob ran two marathons last week..." does not discriminate among respondents)

Optimally Choosing Vignettes

- Ultimate Goal: Define categories with vignettes to learn about a continuous unobserved variable (health, efficacy).
    - Worst choice: All in one category, no discriminatory power (E.g., "Bob ran two marathons last week..." does not discriminate among respondents)
    - Best choice: Largest number of categories, equal proportions across categories

Optimally Choosing Vignettes

- Ultimate Goal: Define categories with vignettes to learn about a continuous unobserved variable (health, efficacy).
  - Worst choice: All in one category, no discriminatory power (E.g., "Bob ran two marathons last week..." does not discriminate among respondents)
  - Best choice: Largest number of categories, equal proportions across categories
- Immediate Goal: Measure <u>information</u> in a categorization scheme

Optimally Choosing Vignettes

- Ultimate Goal: Define categories with vignettes to learn about a continuous unobserved variable (health, efficacy).
    - Worst choice: All in one category, no discriminatory power (E.g., "Bob ran two marathons last week..." does not discriminate among respondents)
    - Best choice: Largest number of categories, equal proportions across categories
- Immediate Goal: Measure <u>information</u> in a categorization scheme
    - Operational use:

Optimally Choosing Vignettes

- Ultimate Goal: Define categories with vignettes to learn about a continuous unobserved variable (health, efficacy).
    - Worst choice: All in one category, no discriminatory power (E.g., "Bob ran two marathons last week..." does not discriminate among respondents)
    - Best choice: Largest number of categories, equal proportions across categories
- Immediate Goal: Measure <u>information</u> in a categorization scheme
    - Operational use:
        - Run a pretest with lots of vignettes

Optimally Choosing Vignettes

- **Ultimate Goal:** Define categories with vignettes to learn about a continuous unobserved variable (health, efficacy).
  - **Worst choice:** All in one category, no discriminatory power (E.g., "Bob ran two marathons last week..." does not discriminate among respondents)
  - **Best choice:** Largest number of categories, equal proportions across categories
- **Immediate Goal:** Measure <u>information</u> in a categorization scheme
  - **Operational use:**
    - Run a pretest with lots of vignettes
    - Compute C and $H(C)$ for each possible subset,

Optimally Choosing Vignettes

- Ultimate Goal: Define categories with vignettes to learn about a continuous unobserved variable (health, efficacy).
  - Worst choice: All in one category, no discriminatory power (E.g., "Bob ran two marathons last week..." does not discriminate among respondents)
  - Best choice: Largest number of categories, equal proportions across categories
- Immediate Goal: Measure <u>information</u> in a categorization scheme
  - Operational use:
    - Run a pretest with lots of vignettes
    - Compute C and $H(C)$ for each possible subset,
    - Choose vignettes for the main survey based on H and cost of survey questions.

A measure of information $H(C)$?

A measure of information $H(C)$?

- Three Criteria for a measure, $H(C)$:

A measure of information $H(C)$?

- Three Criteria for a measure, $H(C)$:
  1. $H(C)$ should be 0 when all answers are in one category;

A measure of information $H(C)$?

- Three Criteria for a measure, $H(C)$:
    1. $H(C)$ should be 0 when all answers are in one category; at a maximum when proportions are equal across categories

A measure of information $H(C)$?

- Three Criteria for a measure, $H(C)$:
  1. $H(C)$ should be 0 when all answers are in one category; at a maximum when proportions are equal across categories
  2. $H(C)$ should increase monotonically with the number of vignettes (and thus categories)

A measure of information $H(C)$?

- Three Criteria for a measure, $H(C)$:
    1. $H(C)$ should be 0 when all answers are in one category; at a maximum when proportions are equal across categories
    2. $H(C)$ should increase monotonically with the number of vignettes (and thus categories)
    3. Assume consistent decomposition as we add vignettes

A measure of information $H(C)$?

- Three Criteria for a measure, $H(C)$:
    1. $H(C)$ should be 0 when all answers are in one category; at a maximum when proportions are equal across categories
    2. $H(C)$ should increase monotonically with the number of vignettes (and thus categories)
    3. Assume consistent decomposition as we add vignettes
- Lots of candidates exist (all inequality measures): Gini index, variance, absolute deviations, Herfindahl index, etc.

A measure of information $H(C)$?

- Three Criteria for a measure, $H(C)$:
    1. $H(C)$ should be 0 when all answers are in one category; at a maximum when proportions are equal across categories
    2. $H(C)$ should increase monotonically with the number of vignettes (and thus categories)
    3. Assume consistent decomposition as we add vignettes
- Lots of candidates exist (all inequality measures): Gini index, variance, absolute deviations, Herfindahl index, etc.
- Only one measure satisfies all three criteria: entropy.

A measure of information $H(C)$?

- Three Criteria for a measure, $H(C)$:
  1. $H(C)$ should be 0 when all answers are in one category; at a maximum when proportions are equal across categories
  2. $H(C)$ should increase monotonically with the number of vignettes (and thus categories)
  3. Assume consistent decomposition as we add vignettes
- Lots of candidates exist (all inequality measures): Gini index, variance, absolute deviations, Herfindahl index, etc.
- Only one measure satisfies all three criteria: entropy.
- Thus, formally, we set:

$$H(p_1, \ldots, p_{2J+1}) = - \sum_{j=1}^{2J+1} p_j \ln(p_j)$$

A measure of information $H(C)$?

- Three Criteria for a measure, $H(C)$:
  1. $H(C)$ should be 0 when all answers are in one category; at a maximum when proportions are equal across categories
  2. $H(C)$ should increase monotonically with the number of vignettes (and thus categories)
  3. Assume consistent decomposition as we add vignettes
- Lots of candidates exist (all inequality measures): Gini index, variance, absolute deviations, Herfindahl index, etc.
- Only one measure satisfies all three criteria: entropy.
- Thus, formally, we set:

$$H(p_1, \ldots, p_{2J+1}) = -\sum_{j=1}^{2J+1} p_j \ln(p_j)$$

- Only question remaining: How do we calculate entropy when C is not a scalar?

- Without ties or inconsistencies, we simply compute entropy

Step 2: Defining $H(C)$ for scalar and vector C

- Without ties or inconsistencies, we simply compute entropy
- With ties and inconsistencies, 2 options:

- Without ties or inconsistencies, we simply compute entropy
- With ties and inconsistencies, 2 options:
  - Estimated entropy: using the censored ordinal probit model

- Without ties or inconsistencies, we simply compute entropy
- With ties and inconsistencies, 2 options:
  - Estimated entropy: using the censored ordinal probit model
  - Known (minimum) entropy: information in the data we know exists for certain (inferences do not depend on the model)

- Without ties or inconsistencies, we simply compute entropy
- With ties and inconsistencies, 2 options:
    - Estimated entropy: using the censored ordinal probit model
    - Known (minimum) entropy: information in the data we know exists for certain (inferences do not depend on the model)
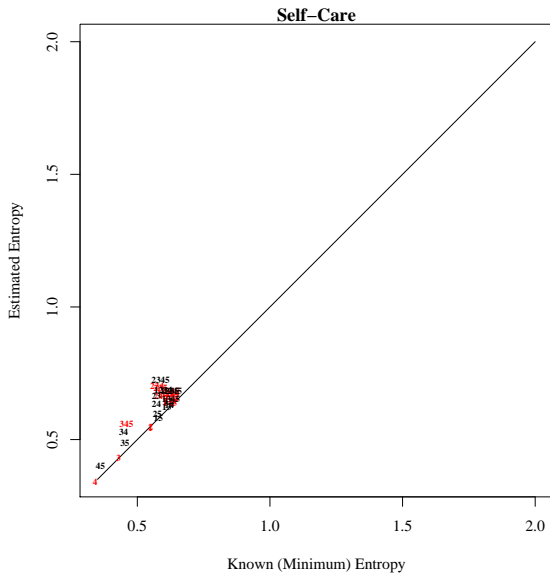- Result is easy to use: one measure indicating information in survey question and vignettes
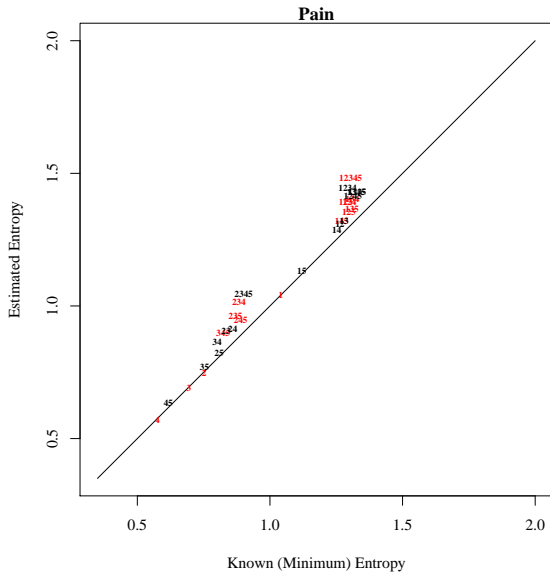
Political Efficacy (Mex & China)
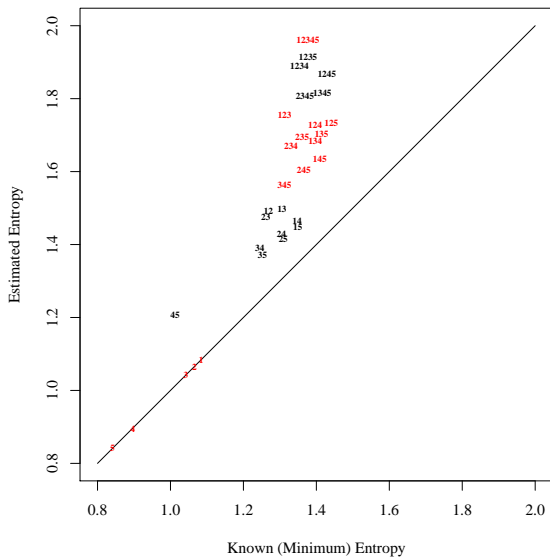
One vignette can be better than three: Sleep (China)
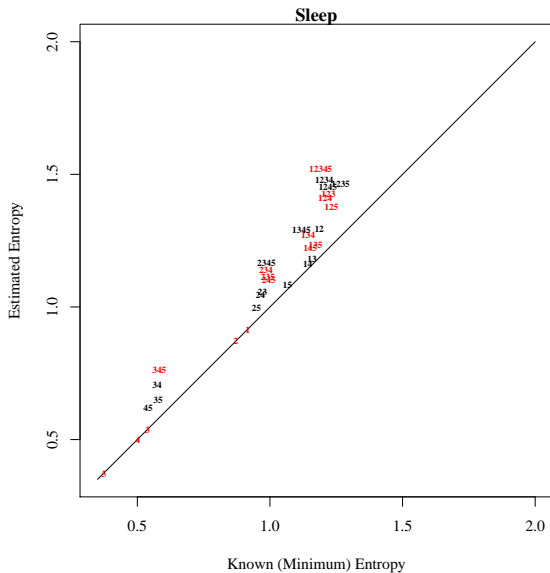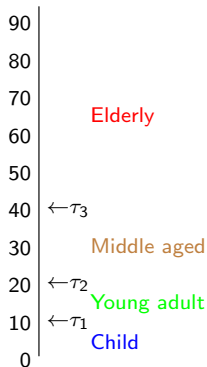
Some vignette sets are uninformative: Self-Care (China)

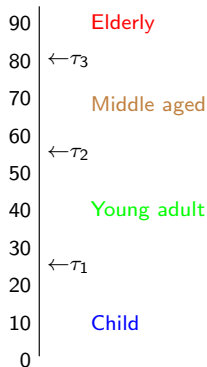# Some covariates are unhelpful: Pain (China)

Categorizing Years of Age

- If thresholds vary, categorical answers are meaningless.
- Our parametric model works by estimating the thresholds.
- Vignettes provide identifying information for the τ'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)?

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

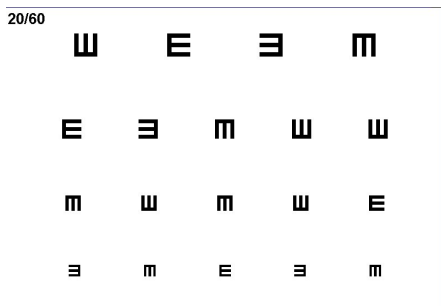## Self-Assessments v. Medical Tests

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:

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

	Snellen Eye Chart		Ordinal Probit		Chopit	
	Mean	(s.e.)	μ	(s.e.)	μ	(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)

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

| | Snellen Eye Chart | | Ordinal Probit | | Chopit | |
	Mean	(s.e.)	μ	(s.e.)	μ	(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)

- The medical test shows Slovakians see much better than the Chinese

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

	Snellen Eye Chart		Ordinal Probit		Chopit	
	Mean	(s.e.)	μ	(s.e.)	μ	(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)

- The medical test shows Slovakians see much better than the Chinese
- Ordinal probit finds no difference

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

	Snellen Eye Chart		Ordinal Probit		Chopit	
	Mean	(s.e.)	μ	(s.e.)	μ	(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)

- The medical test shows Slovakians see much better than the Chinese
- Ordinal probit finds no difference
- Chopit reproduces the same result as the medical test (though on different scale)

Conclusions

- Our approach can fix DIF, if response consistency and vignette equivalence hold — and the survey questions are good

Conclusions

- Our approach can fix DIF, if response consistency and vignette equivalence hold — and the survey questions are good
- Anchoring vignettes will not eliminate all DIF, but problems would have to occur at unrealistically extreme levels to make the unadjusted measures better than the adjusted ones.

Conclusions

- Our approach can fix DIF, if response consistency and vignette equivalence hold — and the survey questions are good

- Anchoring vignettes will not eliminate all DIF, but problems would have to occur at unrealistically extreme levels to make the unadjusted measures better than the adjusted ones.

- Expense can be held down to a minimum by assigning each vignette to a smaller subsample. E.g., 4 vignettes asked for 1/4 of the sample each adds only one question/respondent.

Conclusions

- Our approach can fix DIF, if response consistency and vignette equivalence hold — and the survey questions are good

- Anchoring vignettes will not eliminate all DIF, but problems would have to occur at unrealistically extreme levels to make the unadjusted measures better than the adjusted ones.

- Expense can be held down to a minimum by assigning each vignette to a smaller subsample. E.g., 4 vignettes asked for 1/4 of the sample each adds only one question/respondent.

- Writing vignettes aids in the clarification and discovery of additional domains of the concept of interest — even if you do not do a survey.

Conclusions

- Our approach can fix DIF, if response consistency and vignette equivalence hold — and the survey questions are good

- Anchoring vignettes will not eliminate all DIF, but problems would have to occur at unrealistically extreme levels to make the unadjusted measures better than the adjusted ones.

- Expense can be held down to a minimum by assigning each vignette to a smaller subsample. E.g., 4 vignettes asked for 1/4 of the sample each adds only one question/respondent.

- Writing vignettes aids in the clarification and discovery of additional domains of the concept of interest — even if you do not do a survey.

- We do not provide a solution for other common survey problems: Question wording,

Conclusions

- Our approach can fix DIF, if response consistency and vignette equivalence hold — and the survey questions are good

- Anchoring vignettes will not eliminate all DIF, but problems would have to occur at unrealistically extreme levels to make the unadjusted measures better than the adjusted ones.

- Expense can be held down to a minimum by assigning each vignette to a smaller subsample. E.g., 4 vignettes asked for 1/4 of the sample each adds only one question/respondent.

- Writing vignettes aids in the clarification and discovery of additional domains of the concept of interest — even if you do not do a survey.

- We do not provide a solution for other common survey problems: Question wording, Accurate translation,

Conclusions

- Our approach can fix DIF, if response consistency and vignette equivalence hold — and the survey questions are good

- Anchoring vignettes will not eliminate all DIF, but problems would have to occur at unrealistically extreme levels to make the unadjusted measures better than the adjusted ones.

- Expense can be held down to a minimum by assigning each vignette to a smaller subsample. E.g., 4 vignettes asked for 1/4 of the sample each adds only one question/respondent.

- Writing vignettes aids in the clarification and discovery of additional domains of the concept of interest — even if you do not do a survey.

- We do not provide a solution for other common survey problems: Question wording, Accurate translation, Question order,

Conclusions

- Our approach can fix DIF, if response consistency and vignette equivalence hold — and the survey questions are good

- Anchoring vignettes will not eliminate all DIF, but problems would have to occur at unrealistically extreme levels to make the unadjusted measures better than the adjusted ones.

- Expense can be held down to a minimum by assigning each vignette to a smaller subsample. E.g., 4 vignettes asked for 1/4 of the sample each adds only one question/respondent.

- Writing vignettes aids in the clarification and discovery of additional domains of the concept of interest — even if you do not do a survey.

- We do not provide a solution for other common survey problems: Question wording, Accurate translation, Question order, Sampling design,

Conclusions

- Our approach can fix DIF, if response consistency and vignette equivalence hold — and the survey questions are good

- Anchoring vignettes will not eliminate all DIF, but problems would have to occur at unrealistically extreme levels to make the unadjusted measures better than the adjusted ones.

- Expense can be held down to a minimum by assigning each vignette to a smaller subsample. E.g., 4 vignettes asked for 1/4 of the sample each adds only one question/respondent.

- Writing vignettes aids in the clarification and discovery of additional domains of the concept of interest — even if you do not do a survey.

- We do not provide a solution for other common survey problems: Question wording, Accurate translation, Question order, Sampling design, Interview length,

Conclusions

- Our approach can fix DIF, if response consistency and vignette equivalence hold — and the survey questions are good

- Anchoring vignettes will not eliminate all DIF, but problems would have to occur at unrealistically extreme levels to make the unadjusted measures better than the adjusted ones.

- Expense can be held down to a minimum by assigning each vignette to a smaller subsample. E.g., 4 vignettes asked for $1/4$ of the sample each adds only one question/respondent.

- Writing vignettes aids in the clarification and discovery of additional domains of the concept of interest — even if you do not do a survey.

- We do not provide a solution for other common survey problems: Question wording, Accurate translation, Question order, Sampling design, Interview length, Social backgrounds of interviewer and respondent, etc.

http://GKing.Harvard.edu/vign

Includes:

- Academic papers
- Anchoring vignette examples by researchers in many fields,
- Frequently asked questions,
- Videos
- Conferences
- Statistical software