Direct vs Indirect questioning surveys in a cannabis real study

Elvira Pelle

Department of Communication and Economics

University of Modena and Reggio Emilia

ITACOSM 2019

FLORENCE 5-7 June 2019 | 6th ITAlian COnference on Survey Methodology

Outline

1. Background
2. Focus and aims
3. The survey plan
4. Results
5. Concluding remarks

Background - Respondents' cooperation

The decision of survey participants to honestly cooperate greatly depends on the perceived privacy protection.

Background - Respondents' cooperation

The decision of survey participants to honestly cooperate greatly depends on the perceived privacy protection.

To increase respondents' cooperation:

limit the influence of interviewer from the question and answer process

- self-administered questionnaires (SAQs)
- computer-assisted telephone interviewing (CATI)
- computer-assisted Web interviewing (CAWI), ecc.

build a collaborative and non-hierarchical relationship between the interviewer and the survey participants

- interviewer self-disclosure

Background – Role of interviewer

If the interviewer knows most of the members of the stigmatizing group or is himself/herself a member of that group, the respondents might:

- not be completely inhibited by his/her presence
- be more willing to release personal information
- show indifference to interviewer opinion
- do not fear that their personal information is being released to third parties

Background – Indirect questioning techniques (IQTs)

Principle: *“no direct questions is posed to the survey participants”*

Background – Indirect questioning techniques (IQTs)

Principle: *"no direct questions is posed to the survey participants"*

Responses remain confidential to the respondents and their true status remains undisclosed to both the interviewer and the researcher $\Rightarrow$ privacy is protected!

Focus

Two sensitive topics are investigated by means direct and indirect questioning survey modes in the presence of the interviewer

- the **illegal usage of cannabis** for personal and recreational purposes
- people opinion about **cannabis legalization**

Focus

Two sensitive topics are investigated by means direct and indirect questioning survey modes in the presence of the interviewer

- the **illegal usage of cannabis** for personal and recreational purposes
- people opinion about **cannabis legalization**

Why two **correlated** issues?

- *different level of sensitivity*: interviewees are expected to differently behave when responding about them
- evaluation of the *differences in the two questioning survey approaches*

Aims

The aim of the study is twofold:

- **simultaneous estimation** of the prevalence of individuals
    a. who *have used cannabis at least once* in their life and (attribute A)
    b. who *were in favour of its legalization* (attribute B)

- evaluation the impact of **trust interviewer self-disclosure** on the DQ survey mode

The survey plan

A mixed-mode research was conducted in Santa Maria del Cedro, a municipality of about 5,000 inhabitants in the province of Cosenza, in Southern Italy.

The fieldwork was realized by **a single interviewer** who was well-known and who shared with most of fellow citizens personal experiences and stories concerning the investigated topics.

The survey plan

A mixed-mode research was conducted in Santa Maria del Cedro, a municipality of about 5,000 inhabitants in the province of Cosenza, in Southern Italy.

The fieldwork was realized by **a single interviewer** who was well-known and who shared with most of fellow citizens personal experiences and stories concerning the investigated topics.

3 phases:

1. face-to-face interview (short paper-and-pencile questionnaire)
2. Randomized Response Model
3. Direct questioning survey mode (DQ)

Phase 1: Collecting socio-demographics information

Face-to-face interview with a short questionnaire on socio-demographics characteristics:

- gender
- age
- education
- employment status
- marital status
- number of children

Phase 2: Randomized Response Crossed Model (CM)

Lee, C.-S., Sedory, S. A., and Singh, S. *Statistics and Probability Letters*. (2013)

How the Crossed Model works

Phase 2: Randomized Response Crossed Model (CM)

Lee, C.-S., Sedory, S. A., and Singh, S. *Statistics and Probability Letters*. (2013)

How the Crossed Model works

Deck 1

"I have used cannabis at least once in my life"	*"I am not in favour of cannabis legalization"*
p=0.6	**1-p**

Phase 2: Randomized Response Crossed Model (CM)

Lee, C.-S., Sedory, S. A., and Singh, S. *Statistics and Probability Letters*. (2013)

How the Crossed Model works

Deck 1

"I have used cannabis at least once in my life"

p=0.6

"I am not in favour of cannabis legalization"

1-p

Deck 2

"I am in favour of cannabis legalization"

q=0.6

"I have never used cannabis"

1-q

Phase 2: Randomized Response Crossed Model (CM)

Lee, C.-S., Sedory, S. A., and Singh, S. *Statistics and Probability Letters*. (2013)

How the Crossed Model works

Deck 1

"I have used cannabis at least once in my life"

p=0.6

"I am not in favour of cannabis legalization"

1-p

Deck 2

"I am in favour of cannabis legalization"

q=0.6

"I have never used cannabis"

1-q

Hence, each units provides just one of these possible couples of responses (Yes, Yes), (Yes, No), (No, Yes) or (No, No).

elvira.pelle@unimore.it

Let $\hat{\theta}_{11}, \hat{\theta}_{10}, \hat{\theta}_{01}$ and $\hat{\theta}_{00}$ be the proportion of observed responses *"Yes, Yes"*, *"Yes, No"*, *"No, Yes"* and *"No, No"*
The parameters of interest can be estimated by:

$$\widehat{\pi}_A = \frac{1}{2} + \frac{(q-p+1)(\widehat{\theta}_{11} - \widehat{\theta}_{00}) + (p+p-1)(\widehat{\theta}_{10} - \widehat{\theta}_{01})}{2(p+q-1)} \quad (1)$$

$$\widehat{\pi}_B = \frac{1}{2} + \frac{(p-q+1)(\widehat{\theta}_{11} - \widehat{\theta}_{00}) + (p+p-1)(\widehat{\theta}_{01} - \widehat{\theta}_{10})}{2(p+q-1)} \quad (2)$$

$$\widehat{\pi}_{A\cap B} = \frac{pq(1-p)(1-q)\widehat{\theta}_{00}}{(p+q-1)[pq+(1-p)(1-q)]} \quad (3)$$

$$\widehat{\pi}_{A\cup B} = \widehat{\pi}_A + \widehat{\pi}_B - \widehat{\pi}_{A\cap B} \quad (4)$$

Phase 3: Direct questioning survey mode (DQ)

The interviewer posed directly to the respondents the two sensitive questions:

D1: *“Had you ever used cannabis at least once in your life?"*

D2: *“Are you in favour of cannabis legalization?"*

Phase 3: Direct questioning survey mode (DQ)

The interviewer posed directly to the respondents the two sensitive questions:

D1: *“Had you ever used cannabis at least once in your life?"*

D2: *“Are you in favour of cannabis legalization?"*

Idea: here, the presence of the interviewer (*well-known and who shared experiences*) may encourages respondent to cooperate and reduce the embarrassment to answer truthfully
(this step is denoted as DQ1).

Phase 3: Direct questioning survey mode (DQ)

To verify this working hypothesis, DQ was repeated a second time, collecting new responses after posing the following request:

“What would have been your answer to my previous two questions if you had not known me and/or you not had a trusting connection with me? Now, please, imagine you never have known me and/or you don’t trust in me, and answer again to my questions D1 and D2"

This step is denoted as DQ2.

Table: Point estimates through the CM and DQ modes.
a_1 and b_1 denote that the difference $\widehat{\pi}_{S,CM} - \widehat{\pi}_{S,DQ1}$ is significant at the 5% and 1% level.
a_2 and b_2 refer to the significance of the difference $\widehat{\pi}_{S,CM} - \widehat{\pi}_{S,DQ2}$.
c and d denote significance at 5% and 1% levels for the difference $\widehat{\pi}_{S,DQ1} - \widehat{\pi}_{S,DQ2}$

	$\widehat{\pi}_A$	$\widehat{\pi}_B$	$\widehat{\pi}_{A\cap B}$	$\widehat{\pi}_{A\cup B}$
Sample (n=289)				
CM	0.4706^{b_2}	$0.6851^{a_1,a_2}$	0.3806^{b_2}	$0.7751^{b_1,b_2}$
DQ1	0.4637^{d}	0.6644	0.3910^{d}	0.7370^{d}
DQ2	0.2803	0.6540	0.2526	0.6817

Results

Whole sample

Table: Point estimates through the CM and DQ modes.

a_1 and b_1 denote that the difference $\widehat{\pi}_{S,CM} - \widehat{\pi}_{S,DQ1}$ is significant at the 5% and 1% level.

a_2 and b_2 refer to the significance of the difference $\widehat{\pi}_{S,CM} - \widehat{\pi}_{S,DQ2}$.

c and d denote significance at 5% and 1% levels for the difference $\widehat{\pi}_{S,DQ1} - \widehat{\pi}_{S,DQ2}$

	$\widehat{\pi}_A$	$\widehat{\pi}_B$	$\widehat{\pi}_{A\cap B}$	$\widehat{\pi}_{A\cup B}$
Sample (n=289)				
CM	0.4706^{b_2}	$0.6851^{a_1,a_2}$	0.3806^{b_2}	$0.7751^{b_1,b_2}$
DQ1	0.4637^{d}	0.6644	0.3910^{d}	0.7370^{d}
DQ2	0.2803	0.6540	0.2526	0.6817

- $\widehat{\pi}_A < \widehat{\pi}_B$ and that $\widehat{\pi}_B$ doesn't significantly change $\Rightarrow$ A and B show different levels of sensitivity

Results

Whole sample

Table: Point estimates through the CM and DQ modes.

a_1 and b_1 denote that the difference $\widehat{\pi}_{S,CM} - \widehat{\pi}_{S,DQ1}$ is significant at the 5% and 1% level.

a_2 and b_2 refer to the significance of the difference $\widehat{\pi}_{S,CM} - \widehat{\pi}_{S,DQ2}$.

c and d denote significance at 5% and 1% levels for the difference $\widehat{\pi}_{S,DQ1} - \widehat{\pi}_{S,DQ2}$

	$\widehat{\pi}_A$	$\widehat{\pi}_B$	$\widehat{\pi}_{A\cap B}$	$\widehat{\pi}_{A\cup B}$
Sample (n=289)				
CM	**0.4706^{b_2}**	$0.6851^{a_1,a_2}$	0.3806^{b_2}	$0.7751^{b_1,b_2}$
DQ1	**0.4637^{d}**	0.6644	0.3910^{d}	0.7370^{d}
DQ2	0.2803	0.6540	0.2526	0.6817

- No significant differences between $\widehat{\pi}_{A,CM}$ and $\widehat{\pi}_{A,DQ1}$ $\Rightarrow$ estimates are not affected by the presence of interwiever

Results

Whole sample

Table: Point estimates through the CM and DQ modes.
a_1 and b_1 denote that the difference $\widehat{\pi}_{S,CM} - \widehat{\pi}_{S,DQ1}$ is significant at the 5% and 1% level.
a_2 and b_2 refer to the significance of the difference $\widehat{\pi}_{S,CM} - \widehat{\pi}_{S,DQ2}$.
c and d denote significance at 5% and 1% levels for the difference $\widehat{\pi}_{S,DQ1} - \widehat{\pi}_{S,DQ2}$

	$\widehat{\pi}_A$	$\widehat{\pi}_B$	$\widehat{\pi}_{A\cap B}$	$\widehat{\pi}_{A\cup B}$
Sample (n=289)				
CM	**0.4706^{b_2}**	$0.6851^{a_1,a_2}$	0.3806^{b_2}	$0.7751^{b_1,b_2}$
DQ1	0.4637^{d}	0.6644	0.3910^{d}	0.7370^{d}
DQ2	**0.2803**	0.6540	0.2526	0.6817

- Very large gap between $\widehat{\pi}_{A,CM}$ and $\widehat{\pi}_{A,DQ2}$ $\Rightarrow$ CM works better than DQ2 (“more is better” principle)

Results

Whole sample

Table: Point estimates through the CM and DQ modes.
a_1 and b_1 denote that the difference $\widehat{\pi}_{S,CM} - \widehat{\pi}_{S,DQ1}$ is significant at the 5% and 1% level.
a_2 and b_2 refer to the significance of the difference $\widehat{\pi}_{S,CM} - \widehat{\pi}_{S,DQ2}$.
c and d denote significance at 5% and 1% levels for the difference $\widehat{\pi}_{S,DQ1} - \widehat{\pi}_{S,DQ2}$

	$\widehat{\pi}_A$	$\widehat{\pi}_B$	$\widehat{\pi}_{A\cap B}$	$\widehat{\pi}_{A\cup B}$
Sample (n=289)				
CM	0.4706^{b_2}	$0.6851^{a_1,a_2}$	0.3806^{b_2}	$0.7751^{b_1,b_2}$
DQ1	0.4637^{d}	0.6644	0.3910^{d}	0.7370^{d}
DQ2	0.2803	0.6540	0.2526	0.6817

- Estimates of $\widehat{\pi}_B$ in line with opinon polls in Italy $\Rightarrow$ CM, DQ1 and DQ2 produce reliable results for cannabis legalization

Table: Point estimates through the CM and DQ modes.
a_1 and b_1 denote that the difference $\widehat{\pi}_{S,CM} - \widehat{\pi}_{S,DQ1}$ is significant at the 5% and 1% level.
a_2 and b_2 refer to the significance of the difference $\widehat{\pi}_{S,CM} - \widehat{\pi}_{S,DQ2}$.
c and d denote significance at 5% and 1% levels for the difference $\widehat{\pi}_{S,DQ1} - \widehat{\pi}_{S,DQ2}$

	n	%	$\widehat{\pi}_A$	$\widehat{\pi}_B$	$\widehat{\pi}_{A\cap B}$	$\widehat{\pi}_{A\cup B}$
Gender						
Male	154	53.29%				
CM			$0.5065^{b_1,b_2}$	$0.5909^{b_1,b_2}$	0.3671^{b_1}	0.7303^{b_1}
DQ1			0.5844^{d}	0.7208	0.4870^{d}	0.8182^{d}
DQ2			0.3766	0.7143	0.3312	0.7597
Female	135	46.71%				
CM			$0.4296^{b_1,b_2}$	$0.7926^{b_1,b_2}$	$0.3960^{b_1,b_2}$	$0.8262^{b_1,b_2}$
DQ1			0.3259^{d}	0.6000	0.2815^{d}	0.6444^{c}
DQ2			0.1704	0.5852	0.1630	0.5926

Results

By subgroups

Table: Point estimates through the CM and DQ modes.

a_1 and b_1 denote that the difference $\widehat{\pi}_{S,CM} - \widehat{\pi}_{S,DQ1}$ is significant at the 5% and 1% level.

a_2 and b_2 refer to the significance of the difference $\widehat{\pi}_{S,CM} - \widehat{\pi}_{S,DQ2}$.

c and d denote significance at 5% and 1% levels for the difference $\widehat{\pi}_{S,DQ1} - \widehat{\pi}_{S,DQ2}$

	n	%	$\widehat{\pi}_A$	$\widehat{\pi}_B$	$\widehat{\pi}_{A\cap B}$	$\widehat{\pi}_{A\cup B}$
Age						
16-30	114	39.45%				
CM			0.5439^{b_2}	$0.7632^{b_1,b_2}$	$0.5162^{a_1,b_2}$	0.7908^{b_2}
DQ1			0.5439^{d}	0.6842	0.4561^{d}	0.7719^{d}
DQ2			0.2632	0.6667	0.2368	0.6930
31-60	175	60.55%				
CM			0.4229^{b_2}	0.6343	0.2923^{b_1}	$0.7648^{b_1,b_2}$
DQ1			0.4114^{d}	0.6514	0.3486^{d}	0.7143^{c}
DQ2			0.2914	0.6457	0.2629	0.6743

Table: Point estimates through the CM and DQ modes.

a_1 and b_1 denote that the difference $\widehat{\pi}_{S,CM} - \widehat{\pi}_{S,DQ1}$ is significant at the 5% and 1% level.

a_2 and b_2 refer to the significance of the difference $\widehat{\pi}_{S,CM} - \widehat{\pi}_{S,DQ2}$.

c and d denote significance at 5% and 1% levels for the difference $\widehat{\pi}_{S,DQ1} - \widehat{\pi}_{S,DQ2}$

	n	%	$\widehat{\pi}_A$	$\widehat{\pi}_B$	$\widehat{\pi}_{A\cap B}$	$\widehat{\pi}_{A\cup B}$
Employment status						
Working	144	49.83%				
CM			**$0.6111^{b_1,b_2}$**	**$0.8056^{b_1,b_2}$**	$0.5021^{b_1,b_2}$	$0.9145^{b_1,b_2}$
DQ1			0.5347^{d}	0.7083	0.4375^{d}	0.8056^{c}
DQ2			0.3750	0.7083	0.3333	0.7500
Other	145	50.17%				
CM			$0.3310^{a_1,b_2}$	0.5655^{a_1}	$0.2599^{b_1,a_2}$	0.6366
DQ1			0.3986^{d}	0.6294	0.3497^{d}	0.6783^{c}
DQ2			0.1888	0.6084	0.1748	0.6224

- Among workers: highest % of cannabis users and supporters of its legalization

Results

By subgroups

	n	%	$\widehat{\pi}_A$	$\widehat{\pi}_B$	$\widehat{\pi}_{A\cap B}$	$\widehat{\pi}_{A\cup B}$
Marital status						
Married/Cohabiting	151	52.25%				
CM			0.3510^{a_2}	0.6821	0.2776	$0.7555^{b_1,b_2}$
DQ1			0.3709^{d}	0.6490	0.3245^{d}	0.6954^{c}
DQ2			0.2517	0.6424	0.2450	0.6490
Other	138	47.75%				
CM			0.6014^{b_2}	0.6884	0.4933^{b_2}	0.7965^{b_2}
DQ1			0.5652^{d}	0.6812	0.4638^{d}	0.7826^{d}
DQ2			0.3116	0.6667	0.2609	0.7174
Children						
Yes	137	47.40%				
CM			0.3577^{b_2}	$0.6788^{a_1,a_2}$	0.2920^{a_2}	$0.7445^{b_1,b_2}$
DQ1			0.3285^{d}	0.6277	0.2920^{d}	0.6642
DQ2			0.2336	0.6204	0.2263	0.6277
No	152	52.60%				
CM			0.5724^{b_2}	0.6908	0.4605^{b_2}	0.8026^{b_2}
DQ1			0.5855^{d}	0.6974	0.4803^{d}	0.8026^{d}
DQ2			0.3224	0.6842	0.2763	0.7303

- People engaged in a stable relation or with children: smallest % of cannabis use

Concluding remarks

- Differences in terms of prevalence estimates (especially comparing CM *vs* DQ2)

Concluding remarks

- Differences in terms of prevalence estimates (especially comparing CM *vs* DQ2)
- Different behavior of sensitive attribute, according to the *level of sensitivity*

Concluding remarks

- Differences in terms of prevalence estimates (especially comparing CM *vs* DQ2)
- Different behavior of sensitive attribute, according to the *level of sensitivity*
- *Interviewer self disclosure* plays an important role to improve respondents' cooperation

Thank you
for your attention

When the CM is used, it is not possible to disentangle responses related to attribute A or B $\Rightarrow$ problems arises in obtaining the 2×2 contingency table needed to perform the test.

Testing the difference (CM *vs* DQ)

"Adjusted version" of the McNemar test for paired data

When the CM is used, it is not possible to disentangle responses related to attribute A or B $\Rightarrow$ problems arises in obtaining the 2×2 contingency table needed to perform the test.

Idea: To use available responses under DQ1 and DQ2 on the same units, to obtain a 2×2 contingency table of of prevalence of the sensitive attribute under the CM and DQ* modes, with DQ$^* =$ DQ1, DQ2.

Testing the difference (CM *vs* DQ)

"Adjusted version" of the McNemar test for paired data

The procedure that we carried out can be summarized in the following steps:

- Staring from the estimated prevalence $\widehat{\pi}_S$ on the sample of size n, compute the marginal totals, $n_{i.}$ and $n_{.j}$ $(i, j = 1, 2)$:

	CM		
DQ*	S	$\bar{S}$	Total
S			$n \times \hat{\pi}_{S,\text{DQ}^*}$
$\bar{S}$			$n \times (1 - \hat{\pi}_{S,\text{DQ}^*})$
Total	$n \times \widehat{\pi}_{S,\text{CM}}$	$n \times (1 - \hat{\pi}_{S,\text{CM}})$	n

Testing the difference (CM *vs* DQ)

“Adjusted version” of the McNemar test for paired data

- Fill the above table with the frequencies n_{ij}
  - first derive the contingency table for the CM responses (conditioned to the sub-sample of respondents who declare to possess attribute S under DQ*)
  - Hence, compute again the prevalence estimates on the conditioned table, say $\widehat{\pi}_{S|\mathrm{DQ}^*=\mathrm{Yes}}$

Testing the difference (CM *vs* DQ)

“Adjusted version” of the McNemar test for paired data

- Fill the above table with the frequencies n_{ij}
  - first derive the contingency table for the CM responses (conditioned to the sub-sample of respondents who declare to possess attribute S under DQ*)
  - Hence, compute again the prevalence estimates on the conditioned table, say $\widehat{\pi}_{S|\mathrm{DQ}^*=\mathrm{Yes}}$

- Compute the number of respondents who declare to posses attribute S under both CM and DQ*, say n_{11}, as:

$$n_{11} = n \times \widehat{\pi}_{S,(\mathrm{DQ}^*)} \times \widehat{\pi}_{S|\mathrm{DQ}^*=\mathrm{Yes}}.$$

Testing the difference (CM *vs* DQ)

“Adjusted version” of the McNemar test for paired data

- Complete the contingency table for the McNemar test as follows:

DQ*	CM: S	CM: $\bar{S}$	Total
S	$n \times \widehat{\pi}_{S,\text{CM}}$	$n \times (1 - \widehat{\pi}_{S,\text{DQ}^*} - \widehat{\pi}_{S,\text{CM}^*}) + n_{11}$	$n \times \widehat{\pi}_{S,\text{DQ}^*}$
$\bar{S}$	$n \times \widehat{\pi}_{S,\text{CM}} - n_{11}$	$n \times (1 - \widehat{\pi}_{S,\text{DQ}^*} - \widehat{\pi}_{S,\text{CM}^*}) + n_{11}$	$n \times (1 - \widehat{\pi}_{S,\text{DQ}^*})$
Total	$n \times \widehat{\pi}_{S,\text{CM}}$	$n \times (1 - \widehat{\pi}_{S,\text{CM}^*})$	n