



ISO 9001:2008
Certificato n. 4651



METODI QUANTITATIVI AL SERVIZIO DELLA QUALITÀ

I sondaggi mentono?

*Riflessioni sul sondaggio Metis Ricerche sulle intenzioni di voto
espresse per il Referendum del 4 Dicembre 2016 sulle modifiche
alla legge costituzionale*

Gennaio 2017

Indice

1. Premessa: il sondaggio	3
2. Le stime	3
3. I pesi	4
4. Considerazioni di merito	6
5. Considerazioni di ordine metodologico: i sondaggi mentono?	9
L'importanza delle liste di campionamento.....	9
L'importanza del metodo di estrazione del campione.....	10
L'importanza dei pesi	10
L'importanza della formulazione delle domande.....	12
L'importanza della sostituzione dei dati mancanti	12
Attribuzione dei dati mancanti con modello probabilistico basato su un modello casuale	12
Attribuzione dei dati mancanti con modello probabilistico basato su un modello causale	13
Attribuzione con ricorso ad altra domanda	14
L'importanza del tempo.....	14
La criticità della scelta.....	14
L'importanza della comunicazione.....	16
6. Conclusione: ma i sondaggi mentono o non mentono?	17
Il questionario.....	19

1. Premessa: il sondaggio

Tra il 13 e il 25 novembre Metis Ricerche ha inviato 1200 mail¹ per stimare i risultati del referendum sulla proposta di modifiche alla costituzione² che si sarebbe svolto il 4 Dicembre. Con la speranza di raggiungere un pubblico maggiore il sondaggio è stato anche postato su Facebook e LinkedIn, che hanno fatto registrare il maggior numero di interviste portate a buon fine.

Le interviste portate a termine sono state 255.

I risultati non sono stati pubblicati a conclusione del sondaggio per via del blackout preelettorale, ma sono stati consegnati a *stakeholder* cui Metis Ricerche normalmente fa riferimento.

Divulghiamo ora in modo esteso i risultati per due motivi:

1. Permette di far luce su alcune delle motivazioni che hanno spinto gli elettori a votare in un certo modo;
2. Permette di cominciare a dare alcune risposte all'interrogativo: i sondaggi mentono?

Presenteremo prima le stime, quindi i risultati di merito (anticipando: hanno votato No i giovani, gli anziani, i residenti nel meridione, i poveri), e successivamente le questioni di ordine metodologico.

Prima però illustriamo il modo in cui i dati sono stati pesati, per la sua importanza nel calcolo delle stime.

2. Le stime

I NO, stimati al 25 Novembre, sono il 59,78%. A quella data circa il 5% non sapeva ancora cosa votare.

¹ Gli indirizzi mail usati sono quelli del panel e degli archivi Metis Ricerche

² Questo il testo sottoposto a referendum: «Approvate voi il testo della legge costituzionale concernente "Disposizioni per il superamento del bicameralismo paritario, la riduzione del numero dei parlamentari, il contenimento dei costi di funzionamento delle istituzioni, la soppressione del CNEL e la revisione del titolo V della parte II della Costituzione" approvato dal Parlamento e pubblicato nella Gazzetta Ufficiale n. 88 del 15 aprile 2016?»

Tab. 1 – Distribuzione delle intenzioni di voto al 25 Novembre 2016

Intenzione di voto	N	%
Sì	91	35,58
No	152	59,78
Non so	12	4,64
Totale	255	100

I dati osservati alle elezioni del 4 Dicembre sono stati 59,28% per il NO e 40,72% per il SI se si considera solo il dato nazionale, 59,11% per il NO e 40,89% per il SI, comprendendo anche l'estero.

3.I pesi

Le persone che hanno partecipato al sondaggio sono state 255 e rappresentano un campione auto selezionato. I recapiti sono stati estratti dalle fonti sopra citate che hanno costituito una sorta di primo stadio di campionamento. L'universo degli indirizzi e-mail in esse contenuti ha costituito la lista di campionamento del secondo stadio. La frazione di campionamento è stata di 1 a 1 (numero di invii su numero di presenti nelle liste di campionamento). Sono cioè stati inviati i link al questionario a tutto gli indirizzi email a disposizione.

La distribuzione per fonte dei questionari compilati è stata la seguente (tab. 2).

Tab. 2. – Distribuzione delle interviste per tipo di fonte

Fonte	n.	%
Panel Metis Ricerche 1	11	4,31
Panel Metis Ricerche 2	31	12,16
Panel Metis Ricerche 3	45	17,65
Panel Metis Ricerche 4	80	31,37
Panel Metis Ricerche 5	2	0,78
Social Media	86	33,73
Totale	255	100

Le interviste sono state pesate per genere, età (tre classi di età) e provenienza geografica (cinque provenienze).

Le classi di età considerate sono state fino a 39 anni, tra 40 e 59, 60 e oltre. La provenienza geografica consta di 5 macroregioni, Nord-Ovest, Nord-Est, Centro, Sud, Isole.

Le distribuzioni usate fanno riferimento agli stessi caratteri rilevati per il territorio nazionale nell'anno 2016, che forniscono le seguenti distribuzioni percentuali, rispettivamente per genere, età e ripartizione geografica.

Tab. 3 – Distribuzione delle variabili di stratificazione usate per i pesi

Genere	%
Maschi	48,56
Femmine	51,45
Classi Età	
18-39	30,54
40-59	36,25
60-oltre	33,21
Ripartizione geografica	
Nord-Ovest	26,56
Nord-Est	19,19
Centro	19,89
Sud	23,26
Isole	11,10

Conseguentemente la distribuzione per fonte pesata cambia. La tab.4 è la tab.2 cui sono stati applicati i pesi.

Tab. 4. - Distribuzione delle interviste per tipo di fonte

Fonte	n.	%
Panel Metis Ricerche 1	16	6,20
Panel Metis Ricerche 2	36	14,25
Panel Metis Ricerche 3	26	10,38
Panel Metis Ricerche 4	73	28,61
Panel Metis Ricerche 5	8	3,24
Social Media	95	37,33
Totale	255	100,00

4.Considerazioni di merito

La distribuzione per fonte evidenzia come in tutte le fonti la percentuale più elevata sia per i no, salvo che in due casi. Solo per le mail recuperate nei "Panel Metis Ricerche 1 e 3" la situazione si capovolge. L'entità esigua dei rispondenti per quelle fonti fa sì che ciò non influenzi molto il risultato finale. Considerando soltanto le altre fonti utilizzate nel reperire gli indirizzi email, le variazioni nella propensione espressa al voto continuano ad essere significative, ma meno.

Tab. 5 – Distribuzione delle intenzioni di voto per tipo di fonte

Panel	Sì		No		Non so		Totale	
	n.	%	n.	%	n.	%	n.	%
Panel Metis Ricerche 1	9	54,39	7	45,61	0	0,00	16	100
Panel Metis Ricerche 2	15	40,81	21	57,14	1	2,06	36	100
Panel Metis Ricerche 3	18	67,99	7	27,30	1	4,71	26	100
Panel Metis Ricerche 4	27	36,64	45	61,41	1	1,95	73	100
Panel Metis Ricerche 5	0	0,00	8	100,00	0	0,00	8	100
Social Media	23	23,74	64	67,42	8	8,84	95	100
Totale	91	35,58	152	59,78	12	4,64	255	100

La tabella 6 evidenzia che i giovani avevano intenzione di votare No in massima parte e che anche gli anziani propendevano per tale scelta. I "coetanei" di Renzi erano invece per il Sì.

Tab. 6 – Distribuzione delle intenzioni di voto per classe di età

Età	Sì		No		Non so		Totale	
	n.	%	n.	%	n.	%	n.	%
18-39	5	5,81	71	91,07	2	3,12	78	100
40-59	48	52,68	35	38,52	8	8,80	91	100
60 e oltre	38	44,47	47	53,89	1	1,64	86	100
Totale	91	35,58	152	59,78	12	4,64	255	100

Nella tabella 7 si vede chiaramente che il centro-sud nelle intenzioni di voto propendeva per il NO, mentre il nord per il SI.

Tab. 7 – Distribuzione delle intenzioni di voto per ripartizione geografica

Ripartizione geografica	Sì		No		Non so		Totale	
	n.	%	n.	%	n.	%	n.	%
Nord-ovest	32	47,75	30	44,96	5	7,29	67	100
Nord-est	30	62,24	13	26,95	5	10,81	48	100
Centro	17	34,32	32	62,52	2	3,16	51	100
Sud	0	0,00	59	100,00	0	0,00	59	100
Isole	11	37,21	18	62,79	0	0,00	29	100
Totale	91	35,58	152	59,78	12	4,64	255	100

Dalla successiva tabella 8 è abbastanza chiaro che i benestanti preferivano il Sì.

Tab. 8 – Distribuzione delle intenzioni di voto per posizione sociale

Posizione sociale	Sì		No		Non so		Totale	
	n.	%	n.	%	n.	%	n.	%
Bassa	2	15,96	12	79,45	1	4,59	15	100
Media	50	26,87	129	68,95	8	4,19	187	100
Medio alta	33	69,57	11	23,47	3	6,96	47	100
Alta	5	88,58	1	10,91	0	0,51	6	100
Totale	91	35,58	152	59,78	12	4,64	255	100

Infine il genere: le donne più degli uomini propendevano, al momento dell'intervista, per il NO.

Tab. 9 Distribuzione delle intenzioni di voto per genere

Genere	Sì		No		Non so		Totale	
	n.	%	n.	%	n.	%	n.	%
Maschio	53	42,45	70	56,94	1	0,62	124	100
Femmina	38	29,10	82	62,46	11	8,44	131	100
Totale	91	35,58	152	59,78	12	4,64	255	100

5.Considerazioni di ordine metodologico: i sondaggi mentono?

Dal momento che questo sondaggio ha indovinato quello che poi è stato l'esito del referendum, non solo qualitativamente, ma anche quantitativamente (non solo ha predetto che avrebbero vinto i No, ma ha anche detto di quanto) la risposta sarebbe ora facile: i sondaggi non mentono.

Vorrei però approfondire ulteriormente e rispondere in modo un po' più esaustivo alla domanda: i sondaggi mentono o non mentono? Lo farò accennando alle principali cause che influenzano il calcolo delle stime: *le liste di campionamento, il metodo di campionamento, i pesi, la sostituzione delle osservazioni mancanti, le domande del questionario, la tempestività dell'indagine, criteri di scelta della stima corretta*. Farò vedere come ad influenzare la scelta della stima da diffondere continuo anche *ragioni di opportunità di comunicazione*.

L'importanza delle liste di campionamento

Non sempre si hanno a disposizione liste di campionamento che coprono l'universo interessante per l'indagine. In questo sondaggio sarebbe stato per esempio necessario essere in possesso degli indirizzi email di tutti gli elettori da cui estrarre poi il campione degli intervistandi. Impresa chiaramente impossibile. Quasi sempre la ricostruzione di una lista di campionamento esaustiva non è possibile. Di modo che la lista da cui si estrae il campione è essa stessa già un campione, sovente distorto, dell'universo da rappresentare. Per citare un altro esempio pensiamo alle interviste telefoniche: mano a mano che l'uso dei telefoni cellulari si è diffuso le liste ricostruite sulla base degli elenchi telefonici che riportano solo i numeri degli abbonati con telefono fisso risultano incomplete e scarsamente rappresentative.

Nel nostro caso ci siamo serviti, come detto, delle fonti a nostra disposizione, già elencate. In totale il nostro universo di elettori è stato di 1200 elettori a cui vanno sommati coloro che hanno visto l'evento sul sito FB e il post su LinkedIn che potremmo presumere siano altrettanti. Una frazione molto piccola dell'universo reale comunque. Una lista così costruita aumenta a priori il

marginale di incertezza delle stime poiché mette in dubbio la casualità dell'estrazione del campione finale. È stato possibile misurare l'efficienza di questa lista soltanto a posteriori, confrontando le stime da essa provenienti con il dato reale.

L'importanza del metodo di estrazione del campione

È noto che il campione estratto dalle liste di campionamento deve essere casuale per rendere corretti i test statistici (la media campionaria dei parametri stimati è uguale al parametro da stimare nell'universo). Anche questa caratteristica non è quasi mai rispettata nell'esperienza reale. Soprattutto nei sondaggi e nelle surveys in generale. Ai questionari risponde chi vuole rispondere. Dal momento che chi vuole rispondere ha normalmente caratteristiche diverse da coloro che non rispondono, il campione reale risulta spesso (quasi sempre) distorto. Nel nostro sondaggio il link al questionario è stato mandato a tutti i componenti dell'universo rappresentato nelle nostre liste di campionamento. Valutiamo in circa 15/100 la frazione di campionamento risultante (numero di rispondenti su numero di invii. Hanno risposto in 255). Frazione che tende a zero se si considera l'universo reale.

L'importanza dei pesi

Se consideriamo l'intenzione di voto sul campione non pesato (cioè non riproporzionato sulla distribuzione reale delle variabili di stratificazione), il sondaggio predice Sì. L'applicazione del peso come si è visto capovolge la situazione (tab.10)

Tab. 10 – Distribuzione delle intenzioni di voto per il campione non pesato

Intenzione di voto	n.	%
Si	124	48,63
No	111	43,53
Non so	20	7,84
Totale	255	100

Andiamo a vedere cosa è successo applicando i pesi ad ogni singola variabile usata nella loro costruzione³ e confrontando le distribuzioni così ottenute con la distribuzione non pesata. Come detto in precedenza le variabili sono Genere, Età, Ripartizione Geografica. A sinistra della tab. 11 riportiamo nuovamente la distribuzione pesata, a destra quella non pesata. Si osserva che mentre la distribuzione per genere e per età è abbastanza equilibrata e le numerosità all'interno di ciascun sottoinsieme sufficientemente rappresentative, quella per ripartizione geografica è fortemente distorta e a rischio. Risulta fortemente sovrarappresentato il nord-ovest e risultano fortemente sottorappresentati il centro, il sud e le isole. Mentre per il centro e le isole il numero può ancora rappresentare un campione affidabile anche se troppo piccolo in valore assoluto (4,7% vuol dire 12 rispondenti), per il sud è davvero troppo basso (1,2% vuol dire 3 rilievi in valore assoluto). Di qui le cautele osservate nel calcolare le stime in questa indagine. Nondimeno il sondaggio si è dimostrato efficace.

Tab. 11 – Distribuzioni delle variabili di stratificazione per il campione pesato e non pesato a confronto

Genere	% Pesate	% Non pesate
Maschi	48,55	54,12
Femmine	51,45	45,88
Classi di età		
18-39	30,54	36,47
40-59	35,57	36,08
60-oltre	33,89	27,45
Ripartizione geografica		
Nord-ovest	26,55	81,96
Nord-est	19,19	7,45
Centro	19,89	4,71
Sud	23,26	1,18
Isole	11,10	4,71

³ I pesi sono stati calcolati iterativamente per fare in modo che le distribuzioni marginali delle tre variabili usate fossero rispettate contemporaneamente. È stata usata una macro SAS costruita appositamente.

L'importanza della formulazione delle domande

Il questionario proposto come si osserva nell'allegato è stato molto scarno: 3 domande, 5 se si considerano doppie le ultime due (di cui la prima rivolta solo a chi non dichiara l'intenzione di voto): la domanda sulla probabilità di votare sì o no. E non considerando le anagrafiche. L'ultima domanda chiede il grado di simpatia verso Renzi e Trump con voti da 1 a 10. L'inserimento della domanda sulla probabilità di votare Sì o No ci ha consentito di ridurre il numero degli incerti e di attribuire ai Sì o ai No coloro che dichiarano una preferenza superiore a 5 nell'un caso o nell'altro⁴.

L'importanza della sostituzione dei dati mancanti

Nel contesto del sondaggio i dati mancanti sono in realtà gli incerti, coloro che al 25 Novembre non sapevano cosa votare. Si poteva operare in tre modi: i) lasciare le cose come stavano considerando incerti coloro che dichiaravano di non sapere cosa avrebbero votato, ii) tentare di prevedere cosa avrebbero votato, attribuendo agli incerti una scelta sulla base di un modello probabilistico, iii) ricorrere ad una ulteriore domanda.

Attribuzione dei dati mancanti con modello probabilistico basato su un modello casuale

In questi casi si fa l'ipotesi che gli incerti al momento del voto si distribuiranno allo stesso modo di coloro che hanno dichiarato la loro intenzione di voto al sondaggio. *In questa ipotesi la distribuzione sarà identica a quella misurata per questi ultimi e sarà calcolata sul totale di quelli che hanno dichiarato l'intenzione di voto sul campione non pesato.*

⁴ Chi ha dichiarato una preferenza superiore o inferiore a 5 sia per il Sì che per il No non è stato attribuito. Comunque la stima senza attribuzione dei "non so" attribuisce ai No il 58,79 % che con attribuzione diventa 59,78%

Le distribuzioni delle intenzioni di voto senza e con attribuzione degli incerti calcolate sul file non pesato assegnano entrambe la vittoria ai Sì (se tutti gli intervistati votassero) (tab. 12).

Tab. 12 – Distribuzione delle intenzioni di voto con attribuzione degli incerti sul campione non pesato

Intenzione di voto	n. senza attribuzione	% senza attribuzione	n. con attribuzione	% con attribuzione
Si	124	48,63	135	52,94
No	111	43,53	120	47,06
Non so	20	7,84	0	0,0
Totale	255	100,0	255	100,0

Bisogna ora calcolare le stesse distribuzioni pesando il campione. A questo scopo occorre assegnare fisicamente il voto a ciascun intervistato che non ha dichiarato la propria intenzione di voto (sono 20 persone). Consideriamo a questo scopo una distribuzione uniforme generata casualmente e assegniamo a caso al 52,9% il voto Sì al referendum. Ai restanti assegniamo il No. Nella distribuzione non pesata vincono i Sì. Se pesiamo vincono i No e in misura ancor più ampia di quanto in realtà è successo.

Tab. 13 – Distribuzione delle intenzioni di voto con attribuzione degli incerti pesate e non pesate a confronto

Intenzione di voto	n. non pesato	% non pesata	n. pesato	% pesata
Si	135	52,94	98	38,46
No	120	47,06	157	61,54
Totale	255	100,0	255	100,0

Attribuzione dei dati mancanti con modello probabilistico basato su un modello causale

Un altro modo per assegnare i valori mancanti consiste nell'ipotizzare un effetto di altre variabili sulla decisione futura di votare Sì o No. Abbiamo ipotizzato che a governare la scelta avrebbe potuto essere il grado di simpatia/antipatia suscitata da Renzi. Sulla base della considerazione che chi

intende votare SI lo considera simpatico nella misura di 6,26, chi non intende votarlo nella misura di 3,15, e nella misura di 3,36 gli incerti (complessivamente la media è 4,27). Le stime che si producono in questo caso non sono molto lontane dal modello casuale: i no aumentano ancora e arrivano al 62,2, ma la differenza è poca.

Attribuzione con ricorso ad altra domanda

Come detto l'inserimento della domanda sulla probabilità di votare Sì o No ci ha consentito di ridurre il numero degli incerti e di attribuire ai Sì o ai No coloro che dichiarano una preferenza superiore a 5 nell'un caso o nell'altro. In altre parole ce lo siamo fatto dire da loro stessi. Non sempre ce lo hanno detto per cui in questo caso il numero dei valori mancanti è soltanto ridotto e non eliminato. Ne rimangono 12 non attribuiti. Questo è il metodo che abbiamo scelto per presentare le stime e su cui abbiamo costruito i ragionamenti di merito (nel merito i risultati non sarebbero cambiati anche usando le altre distribuzioni).

L'importanza del tempo

Un altro elemento che può giocare un ruolo sulla correttezza o meno delle previsioni è il tempo intercorrente tra la chiusura del sondaggio e il referendum vero e proprio. Noi abbiamo chiuso il 25 Novembre e il referendum si è svolto il 4 Dicembre. In quell'ultima settimana sono successe molte cose: la dichiarazione di Prodi pro-Renzi, la decisione di chiudere il contratto con gli insegnanti, le notizie intorno ad un aumento delle pensioni. Se la propaganda elettorale sale di tono quando il sondaggio è già chiuso gli effetti sull'intenzione di voto non sono registrati. Nel nostro caso potremmo attribuire a tali effetti lo spostamento degli incerti sul Sì.

La criticità della scelta

Ricordo che questi risultati sono stati ottenuti prima del referendum. Quale stima scegliere? Da quanto precedentemente scritto e considerando solo il campione pesato, le stime tra cui scegliere sono quattro (tab. 14).

Tab. 14 – Distribuzione delle intenzioni di voto per tipo di conteggio

	Senza attribuzione dei valori mancanti	Attribuzione parziale dei valori mancanti	Attribuzione totale con metodo caSUale	Attribuzione totale con metodo caUSale	Distribuzione non pesata senza attribuzione dei valori mancanti
Si	32,88	35,58	38,46	37,81	48,63
No	58,79	59,78	61,54	62,19	43,53
Non so	8,33	4,64	0	0,0	7,84
Totale	100	100,0	100	100,0	100,0

Il criterio statistico ci aiuta nello scartare la distribuzione non pesata, confortando una scelta anche intuitivamente corretta (la distribuzione per ripartizione geografica è eccessivamente distorta se non si corregge con il peso). Nella tabella 15 (pesata) si osserva che gli intervalli costruiti intorno al Sì non sono superiori al 50%, mentre quelli costruiti intorno al No non sono inferiori al 51%. Usando questa distribuzione siamo ragionevolmente sicuri (al 95%) che il Sì avrà una percentuale di voti inferiore al 50% e siamo ragionevolmente sicuri che il No avrà una percentuale di voti superiore al 50%. Entrambe le distribuzioni si potranno modificare quando gli indecisi avranno votato. Certamente però (al 95%) il No avrà una percentuale superiore al 50%. Il ragionamento non cambia anche ipotizzando che tutti gli incerti o parte di essi (si vedano le ipotesi già discusse) votino sì.

Tab. 15 – Intervalli di confidenza al 95% per le stime. Distribuzione pesata

	Limite Inferiore al 95%	Valore stimato	Limite superiore al 95%
Si	25,85	32,88	39,91
No	51,42	58,79	66,15
Non so	4,19	8,33	12,47

Nel caso della distribuzione non pesata (tab. 16) le stime dei Sì non consentono di essere ragionevolmente certi (al 95%) che il No non prevalga. Stimare Sì è compatibile con la scelta opposta (si veda in tabella il limite inferiore dell'intervallo di stima per il Sì).

Tab. 16–Intervalli di confidenza al 95% per le stime. Distribuzione non pesata

	Limite Inferiore al 95%	Valore stimato	Limite superiore al 95%
Si	42,46	48,6	54,73
No	37,41	43,5	49,58
Non so	4,51	7,8	11,09

Sempre mettendoci dal punto di vista di chi non sa ancora cosa succederà per davvero, abbiamo ancora quattro opzioni da vagliare. Tutte consentono di essere sufficientemente sicuri nell'affermare che prevarrà il No. Ma di quanto? Il problema non è molto difficile da risolvere in quanto tutte le stime forniscono percentuali abbastanza vicine: si va da un minimo del 58,8% ad un massimo del 62,2% per i No, con una differenza percentuale di 3,4 punti tra minimo e massimo. In questi casi è prassi comune cautelarsi scegliendo per difetto, cioè scegliendo il valore più basso. Questo è quanto abbiamo fatto noi. La stima da noi comunicata ai nostri *stakeholder* è stata 58,8%, con un errore del 3 per mille rispetto a quanto poi effettivamente avvenuto (il No ha vinto con il 59,1%). Per ragioni di comunicazione in questo rapporto abbiamo usato una distribuzione con una stima leggermente superiore e con un errore maggiore, 59,8%. Più sotto nel testo si spiega perché lo abbiamo fatto.

L'importanza della comunicazione

Si sa che i numeri quando entrano nel mondo della comunicazione acquistano altri significati a seconda del contesto in cui vengono collocati. Questo rapporto è stato scritto dopo il referendum. Il contesto attuale è diverso da quello in cui il sondaggio venne realizzato. In particolare il numero su cui i media si sono soffermati e che ha creato comune sentire è stato 59, spesso accompagnato da almeno un decimale. Dal momento che le persone quando si tratta di numeri non arrotondano, ma troncano (si veda quanti prezzi sono presentati con il ,99) se noi ora dicessimo 58,8% daremmo l'impressione di aver sbagliato di più poiché il lettore leggerebbe 58 e confronterebbe con 59. Questa è la ragione per cui in questo rapporto abbiamo scelto di presentare la percentuale dei no con 59,8, una delle stime presentate. Lo stesso lettore, troncando,

legge 59, il risultato osservato. Sebbene l'errore sia in questo caso il 7 per mille, il doppio della stima su cui abbiamo lavorato prima del referendum. Se ci si pensa bene questa è anche la ragione per cui il 59 resta 59 e non diventa 60 nella mente delle persone. Idiosincrasia per gli arrotondamenti.

6. Conclusione: ma i sondaggi mentono o non mentono?

In questo lavoro ci siamo messi in una situazione pessima per l'effettuazione di un sondaggio: scarsità di fonti e fonti non convenzionali, campione bias, presenza di osservazioni mancanti, scarsa numerosità del campione. Campione auto selezionato. In questa situazione il campione ha mentito: ha pronosticato la vittoria del Sì, infatti. Con opportuni interventi, innanzitutto con l'introduzione dei pesi, il campione non ha mentito. Ne traiamo il seguente insegnamento. In una situazione di carenza di liste di campionamento adeguate che producono campioni poco numerosi e distorti si può supplire con un corretto uso delle tecniche di correzione statistica dei risultati e raggiungere previsioni affidabili. Abbiamo visto che ci sono almeno 8 punti da considerare, ciascuno dei quali può generare errori:

- 1. le liste di campionamento,*
- 2. il metodo di campionamento*
- 3. i pesi*
- 4. la sostituzione delle osservazioni mancanti*
- 5. le domande del questionario*
- 6. la tempestività dell'indagine*
- 7. la criticità dei criteri di scelta tra stime diverse*
- 8. la comunicazione*

Su ciascuno dei punti elencati ci si può sbagliare. In questo caso i risultati del sondaggio possono essere errati. Ma se su tutti si lavora correttamente i risultati del sondaggio non mentono. Ci può essere un caso critico: quando nel tempo intercorrente tra il termine del sondaggio e l'evento di interesse succedono cose non prevedibili, come intensificazione abnorme della propaganda nel caso più semplice, fino al verificarsi di vere e proprie catastrofi che mettono in forse l'accadimento stesso dell'evento. Oppure se vi sono brogli

elettorali, nel caso di elezioni. In questi casi a mio giudizio non bisogna parlare di sbagli del sondaggio, piuttosto di distorsione dell'evento. Anche il modo di presentare i risultati può rispondere ad esigenze particolari, per cui può non essere conveniente pubblicare il dato più vicino alla realtà dei fatti, come si è visto per il nostro caso particolare.

Il questionario

D01	Cosa ha intenzione di votare al referendum del 4 Dicembre 2016?	
	1	<i>Si</i>
	2	<i>No</i>
	3	<i>Non so</i>

D02	Quanto ritiene probabile da 1 (per nulla) a 10 (moltissimo)...												
			Per nulla										Moltissimo
	01	Votare sì	1	2	3	4	5	6	7	8	9	10	
	02	Votare no	1	2	3	4	5	6	7	8	9	10	

D03	Quanto le stanno simpatici i seguenti personaggi a 1 (per nulla) a 10 (moltissimo)...												
			Per nulla									Moltissimo	Non so chi sia
	01	Matteo Renzi	1	2	3	4	5	6	7	8	9	10	
	02	Donald Trump	1	2	3	4	5	6	7	8	9	10	

A001	Lei è	
	1	Maschio
	2	Femmina

A002	Quale è il suo anno di nascita?

A003	Attualmente lei è	
	1	Occupato
	2	Cassa integrato / Mobilità
	3	Disoccupato
	4	In cerca di prima occupazione
	5	Pensionato
	6	Casalinga
	7	Studente
	8	Inabile al lavoro
	9	Altro, specificare

A004	Mi potrebbe dire il tipo di contratto che ha o l'ultimo che ha avuto?	
	1	Occupato
	2	Contratto dipendente
	3	Autonomo
	4	Contratto atipico
	5	Altro

A005	Mi potrebbe indicare la sua professione o l'ultima svolta?	
	1	Dirigente
	2	Direttivo, Quadro
	3	Impiegato
	4	Capo operaio, operaio, subalterno
	5	Lavoratore a domicilio
	6	Altro dipendente
	7	Imprenditore con dipendenti
	8	Imprenditore senza dipendenti
	9	Lavoratore in proprio (artigiano, commerciante...)
	10	Socio di cooperativa
	11	Coadiuvante familiare
	12	Altro autonomo

A006	Quale è il suo titolo di studio?	
	1	Nessun titolo
	2	Licenza elementare
	3	Licenza scuola media inferiore
	4	Diploma o qualifica
	5	Laurea di primo livello
	6	Laurea di secondo livello
	7	Laurea vecchio ordinamento
	8	Dottorato, master post laurea

A007	A quale classe sociale pensa di appartenere?	
	1	Bassa
	2	Media
	3	Medio alta
	4	Alta

A008	Quale è il suo comune di residenza?	