Alberto Vitalini

L'uso delle reti sociali per la costruzione di campioni probabilistici

www.aracneeditrice.it
info@aracneeditrice.it

via Raffaele Garofalo, 133/A–B
00173 Roma
(06) 93781065

ISBN 978-88-548-4894-8

I edizione: maggio 2012

Indice

Appendici

Premessa

L'idea di questo libro è nata abbastanza casualmente durante un vagabondaggio nella rete. Mentre cercavo articoli che affrontassero in modo scientifico il legame fra creatività e uso di sostanze stupefacenti nel mondo dell'arte mi sono imbattuto in un articolo dal titolo "Finding the beat: Using respondent–driven sampling to study jazz musicians" che descriveva l'uso di una strategia di campionamento "innovativa" (parole dell'autore) per identificare e intervistare musicisti jazz in quattro città americane. La strategia era descritta in questi termini: "Sulla base delle informazioni iniziali fornite da parte del coordinatore e dallo staff cittadino, sarà selezionata una mezza dozzina di musicisti in ogni città come 'semi' a partire dai quali iniziare le catene di reclutamento. Questi musicisti saranno informati riguardo alle finalità dello studio, intervistati faccia a faccia usando un questionario a 114 domande, e verrà data loro, successivamente, l'opportunità di reclutare al massimo quattro musicisti ognuno. I reclutati saranno poi intervistati a loro volta e verrà data loro l'opportunità di reclutare altri colleghi. Questo processo continuerà finché l'obiettivo di 300 musicisti sarà raggiunto in ogni città. I musicisti saranno pagati per le loro interviste e per ognuno delle persone reclutate dopo che quest'ultime avranno completato l'intervista". (Heckathorn–Jeffri, 2001, p. 317).

Nulla di nuovo, già visto si dirà. Il ricercatore propone una strategia di selezione dei casi che va sotto il nome di campionamento a valanga e l'utilizzo di incentivi monetari per assicurarsi una maggiore collaborazione da parte degli intervistati. Dov'è l'aspetto innovativo? Nel fatto che l'autore affermava di aver sviluppato una teoria statistica che rendeva per la prima volta possibile ottenere sia stime non distorte dei parametri della popolazione sia misure della precisione di queste stime a partire da dati raccolti con un campionamento a valanga.

Non c'è bisogno di sottolineare le potenzialità di una simile "scoperta". Il mio interesse si era acceso. Ho iniziato così, non senza una certa dose di scetticismo, a documentarmi sul metodo proposto. Secondo il

vecchio adagio che l'appetito viene mangiando, il mio interesse si è, poi, via via allargato allo studio generale dei disegni di campionamento che sfruttano i legami sociali e alle possibilità del loro utilizzo per la costruzione di campioni di tipo probabilistico.

Il presente lavoro è il frutto di questo studio e può essere considerato il resoconto di un viaggio in un territorio poco esplorato della metodologia delle scienze sociali. Il linguaggio che si è cercato di utilizzare nel testo è stato il più possibile discorsivo, riducendo al minimo l'uso della formalizzazione matematica, per mettere in risalto i fondamenti logici delle tecniche campionarie trattate e le loro possibilità di utilizzo nelle scienze sociali. I dettagli delle procedure statistico matematiche e di calcolo sono stati, invece, rimandati nelle note e nelle appendici.

Le affermazioni contenute in questo libro esprimono il mio personale punto di vista, e non impegnano in alcun modo l'Istituto Nazionale di Statistica, presso il quale lavoro.

Introduzione

Gli individui in società sono legati gli uni con gli altri da relazioni ed interazioni che ne influenzano le percezioni, gli atteggiamenti e i comportamenti. Lo studio dei legami fra i membri di un gruppo si è rivelato così importante che si è sviluppato un approccio teorico e metodologico che va sotto il nome di analisi delle reti e che cerca di identificare, misurare, testare ipotesi riguardo le forme strutturali e ai contenuti sostanziali delle relazioni fra gli attori (Mattioli, 1995; Piselli, 1995; Chiesi, 1996, 1999; Gribaudi, 1996; Mutti, 1996; Knoke–Yang, 2008).

I legami che uniscono le persone possono però essere considerati, in un'accezione maggiormente pragmatica, anche come mezzi per trovare casi "interessanti" da inserire nella ricerca. Considerando questa seconda accezione essi sono applicati in situazioni molto diverse in riferimento sia agli obiettivi della ricerca che alle tecniche usate per la raccolta dei dati e delle informazioni. Alcuni esempi chiariranno il punto.

Nell'ambito delle tecniche non standard i legami sociali sono utilizzati per guadagnare l'accesso al campo studiato in ricerche che fanno uso dell'osservazione partecipante (Whyte, 1955; Patrick, 1973; Sclavi, 1994). Fa parte della storia sociologica il racconto di Whyte (1955) che fu accettato dal gruppo di giovani di Cornerville, solo dopo che Doc, un giovane che godeva di credibilità e prestigio nel quartiere, lo aveva portato in giro e presentato come amico. Il nome dello stesso Doc era stato a sua volta suggerito a Whyte da un'assistente sociale del quartiere.

Oltre che per guadagnare l'accesso al campo, i legami sociali sono spesso utilizzati per individuare soggetti da coinvolgere nella realizzazione di interviste biografiche: ad ogni persona intervistata viene chiesto di suggerire il nome di altre persone da intervistare. Per fare un esempio, con questo sistema in una ricerca sulla criminalità (Barbagli, 1995) si riuscirono ad intervistare oltre sessanta autori di reati individuati a partire dalle conoscenze personali di due intervistatori.

Nell'ambito delle tecniche standard, i legami sociali sono utilizzati per costruire campioni di persone, appartenenti a popolazioni di cui non si possiede la lista di campionamento (es. immigrati clandestini), alle quali somministrare un questionario. La procedura è simile a quella utilizzata per le interviste biografiche: ad ogni persona intervistata viene chiesto di segnalare il nome di altre persone da intervistare che, oltre ad entrare a far parte del campione ed essere intervistate, diventeranno informatori per l'individuazione di ulteriori contatti e così via... La differenza rispetto alle interviste biografiche è nella numerosità campionaria: l'intervista con questionario comporta numerosità campionarie solitamente elevate. In genere i ricercatori che utilizzano procedure di questo tipo si limitano a selezionare un congruo numero di persone, alle quali somministrare il questionario, solo al fine di esplorare il fenomeno studiato senza porsi come obiettivo la generalizzazione dei risultati, spesso per preparare il campo a studi successivi.

Poniamoci la seguente domanda: nel caso i ricercatori non volessero limitare gli obiettivi della ricerca ad un studio esplorativo e intendessero generalizzare i risultati alla popolazione studiata, sarebbero in grado di farlo, partendo da campioni selezionati sulla base di procedure che sfruttano i legami sociali? La riflessione che ha cercato di rispondere a questa domanda ha avuto un vigoroso sviluppo, sia per le applicazioni pratiche sia per gli aspetti metodologici, soprattutto a partire dalla seconda metà degli anni '80 (per una rassegna cfr. Atkinson–Flint, 2001), in concomitanza con l'emergere di una richiesta sempre più consistente di tipi di campionamento "alternativi" a quelli, tradizionalmente usati nella ricerca sociale, basati sull'estrazione casuale a partire da liste di campionamento.

Questa crescente richiesta è attribuibile, in ultima istanza, alle difficoltà che le tecniche di campionamento tradizionali, basate su liste di nominativi, incontrano nel risolvere problemi di reperimento casuale posti dalle grandi trasformazioni che hanno attraversato le società occidentali (ad es. mutamenti delle forme di convivenza familiare, elevata mobilità lavorativa e territoriale e processi migratori); dall'espansione della ricerca sociale in settori molto delicati della vita individuale (ad es. la ricerca sociale nella sanità) (Lanzetti, 2004; Lanzetti *et al.*, 2008) e dalla crescente preoccupazione circa le problematiche della privacy e della riservatezza.

Il presente contributo si inserisce in questo filone di riflessione: più precisamente l'obiettivo di questo lavoro consiste nel chiarire le possibilità e i limiti di utilizzo, nelle indagini campionarie, dei legami sociali per la costruzione di campioni probabilistici per lo studio di popolazioni di cui non si possiede la lista di campionamento e che, di conseguenza, non sono facilmente campionabili utilizzando le strategie di campionamento "classiche" quali, ad esempio, il campionamento casuale semplice, stratificato e a stadi. Alcuni esempi daranno concretamente l'idea del tipo di popolazioni per le quali l'utilizzo delle riflessioni esposte in questo testo potrebbero rivelarsi utili: immigrati irregolari, persone con redditi molto alti o molto bassi, persone colpite da gravi malattie, scienziati disabili, famiglie che possiedono solo il telefonino; bambini dai sei ai dieci anni, tossicodipendenti, membri di sette religiose, sieropositivi, collezionisti di francobolli, membri di community on line, senza tetto, collezionisti di auto d'epoca, veterani dell'Iraq.

Una precisazione è necessaria. Nella riflessione metodologica delle scienze sociali a popolazioni di questo tipo sono spesso associati gli aggettivi "elusive", "nascoste" e altri sinonimi che servono a sintetizzare le difficoltà che caratterizzano il loro studio. Questi aggettivi non verranno utilizzati nel testo, dal momento che c'è mancanza di chiarezza circa le definizioni di "nascosto" e "elusivo". Questi termini sono utilizzati non in modo univoco e chiaro. Sono stati definiti, a seconda degli autori, come popolazioni "elusive" o "nascoste" i più svariati tipi di popolazioni: ad es., giovani donne single e disoccupate, donne vittime di violenze domestiche, minoranze etniche, persone che hanno contratto il virus del HIV, prostitute, consumatori di sostanze psicotrope, anziani, disabili, senza tetto, persone con redditi sotto la soglia di povertà, evasori fiscali, membri di piccole comunità religiose, uomini di affari, residenti non ancora iscritti all'anagrafe, omosessuali (Kish, 1991; Atkinson–Flint, 2001; Brackertz, 2007). L'ampia casistica a cui sono associati i termini "elusivo" e "nascosto" e il loro uso non coerente minano la loro utilità nell'ambito di una riflessione metodologica.

Il libro è formato da due parti: la prima teorica e la seconda empirica.

Nella parte teorica si cercherà di chiarire le possibilità e i limiti di utilizzo, nelle indagini campionarie, dei legami sociali per la costru-

zione di campioni probabilistici di popolazioni di cui non si dispone degli elenchi dei membri.

Le diverse strategie di campionamento di questo tipo verranno ricondotte ad un'unica formalizzazione matematica: il campionamento in grafo. Questa formalizzazione, oltre ad aver un valore in sé perché fornisce un quadro concettuale parsimonioso per affrontare diverse problematiche sollevate dalla riflessione metodologica su questi temi, è funzionale alla comprensione di un particolare disegno di campionamento chiamato *respondent–driven sampling* (Heckathorn, 1997). Questo disegno si propone come uno degli approcci più innovativi e promettenti per lo studio delle popolazioni di cui non si dispone degli elenchi degli appartenenti. La prima parte terminerà considerando criticamente gli aspetti del *respondent–driven sampling* che meritano ulteriori approfondimenti e studi.

La seconda parte consiste in una valutazione empirica del *respondent-driven sampling*. Questo tipo di campionamento si è rivelato relativamente semplice, economico e flessibile, ed è stato utilizzato con successo per studiare, in più di una dozzina di nazioni, diversi tipi di popolazioni di cui non si possiede la lista dei membri: in particolare prostitute, tossicodipendenti, omosessuali (per una rassegna Malekinejad *et al.*, 2008).

Il favore con cui è stato accolto evidenzia la capacità del *respondent-driven sampling* di rispondere ad un bisogno diffuso nella comunità scientifica, ma non deve far dimenticare che una sua completa accettazione richiede un'approfondita valutazione dell'accuratezza delle stime.

Il presente lavoro cercherà di valutare le performance del *respondent–driven sampling* sia attraverso simulazioni sia svolgendo un'indagine su una popolazione reale, una community Internet, di cui è stata ricostruita la struttura delle relazioni che legano le persone fra loro. L'utilizzo combinato di simulazione e indagine–test consentirà di sviluppare una comprensione qualitativamente più profonda di questa forma di campionamento.

Capitolo I

Il campionamento probabilistico nelle scienze sociali: alcune precisazioni

Prima di iniziare a considerare forme di campionamento che sfruttano i legami sociali sono necessarie alcune precisazioni sugli aspetti generali del campionamento rispetto ai quali le riflessioni di un ricercatore in scienze sociali si possono rivelare utili e fondate (par. 1); sui principi logici che consentono di fare inferenza statistica a partire da campioni estratti casualmente (par. 2); sulla ragione per cui le tecniche classiche di campionamento probabilistico sono in difficoltà nel trattare popolazioni di cui non si possiede la lista di campionamento (par. 3); e sull'espressione "campionamento a valanga" che si applica comunemente, nella riflessione metodologica delle scienze sociali, alla strategia che sfrutta i legami fra le persone per individuare i soggetti da inserire nel campione (par. 4).

Queste precisazioni non sono oziose o frutto di un ragionamento bizantino, servono per inquadrare le riflessioni successive evitando possibili fraintendimenti e confusioni. Quando si affronta in ambito sociologico una riflessione sul campionamento non è raro imbattersi in diverse definizioni degli stessi concetti. Basti pensare ai diversi significati attribuiti ai termini "rappresentativo" e "casuale" utilizzati in frasi molto comuni del tipo «il mio campione è rappresentativo perché è stato estratto casualmente» (Marradi, 1997; Palumbo–Garbarino, 2004).

1.1. Il campionamento probabilistico: sintassi e semantica

È bene, quando si parla di campionamento probabilistico nelle scienze sociali, tenere distinti due piani: quello della formalizzazione mate-

matica, che si potrebbe definire grammaticale (o sintattico), e quello della sua applicazione nella realtà, che si potrebbe definire semantico.

Il piano grammaticale è quello del modello matematico cioè della rappresentazione formale, espressa in linguaggio matematico, di un fenomeno (Israel, 2002). In generale, una descrizione "completa" di un fenomeno sarebbe un interminabile (quanto impossibile) discorso aderente a tutte le pieghe dei fatti, nessuna esclusa. Per descrivere un fenomeno è necessario fare delle scelte, selezionarne degli aspetti, in una parola semplificare. Anche il modello matematico risente di questo limite: esso, infatti, è una rappresentazione semplificata di un fenomeno. Il modello e il fenomeno studiato sono, di conseguenza, in un rapporto di analogia, non di identità: questo va ben tenuto presente quando dal piano sintattico si passa a quello semantico.

Il piano semantico è quello dell'applicazione al fenomeno studiato del modello matematico e delle deduzioni che si possono trarre da esso. Affinché quest'operazione sia sensata, il modello deve essere in grado di descrivere adeguatamente il sistema.

Riportando questa riflessione generale al tema specifico del campionamento, il dibattito metodologico nelle scienze sociali si muove essenzialmente all'interno del piano semantico: cioè analizza, discute, problematizza se, e quanto, il modello matematico del campionamento possa applicarsi nella reale pratica della ricerca; se il modello matematico (sintassi) è in grado di descrivere con l'approssimazione richiesta il sistema reale che si desidera studiare. Ad esempio, il modello matematico alla base del campionamento di tipo probabilistico richiede che tutte le persone estratte siano intervistate e che rispondano in modo accurato. Il dibattito metodologico nelle scienze sociali valuta se queste condizioni si riscontrano nella reale pratica delle indagini campionarie e quali sono le conseguenze sull'accuratezza delle stime di una loro eventuale deviazione.

Il contributo di questo libro si muove nell'ambito della semantica: fin dove, con quali limiti si possono utilizzare nella ricerca le riflessioni sviluppate nell'ambito della teoria dei campioni. Come già anticipato nella premessa, il linguaggio che si è cercato di utilizzare nel testo è stato il più possibile discorsivo, riducendo al minimo l'uso della formalizzazione matematica, per mettere in luce i fondamenti logici delle tecniche campionarie trattate (la sintassi). Sono stati considerati gli aspetti sintattico–formali strettamente necessari per capire le appli-

cazioni pratiche a problemi di ricerca che sono "sociologicamente" rilevanti come ad esempio lo studio delle popolazioni devianti o delle minoranze etniche. I dettagli delle procedure statistico matematiche e di calcolo sono stati, invece, rimandati nelle appendici e nelle note a piè di pagina.

1.2. Il campionamento probabilistico e la probabilità di inclusione: un connubio indivisibile

Il campione è una parte selezionata di un tutto dalla cui analisi si traggono informazioni sull'insieme. Accettata questa definizione di campione siamo di fronte ad un tipico problema di inferenza induttiva: da una proposizione particolare (ad es., la maggioranza delle persone nel campione ritiene che non dovrebbero esserci discriminazioni sul lavoro nel caso di donne con figli) si vuole giungere ad asserzioni generali con un certo livello di fiducia (ad es., si è ragionevolmente sicuri che la maggioranza delle persone nella popolazione ritiene che non dovrebbero esserci discriminazioni sul lavoro nel caso di donne con figli). Per rispondere a questo tipo di problema una parte degli studiosi di scienze sociali, che aderisce al paradigma positivistico, fa ricorso agli strumenti e alle procedure matematiche dell'inferenza statistica, i quali sono stati messi a punto nell'ambito di una branca della matematica che va sotto il nome di teoria delle probabilità, o più specificatamente, teoria dei campioni. Per poter applicare in un'indagine campionaria reale gli strumenti dell'inferenza statistica è necessario attribuire ad ogni individuo una probabilità nota, non nulla di venire a far parte del campione. La probabilità non deve essere uguale per tutte le unità; al limite ogni unità della popolazione può avere una probabilità diversa, purché sia conosciuta e diversa da zero.

Differenti probabilità possono risultare da alcune caratteristiche della procedura di campionamento (ad es., campionamento a stadi) oppure possono essere imposte deliberatamente dal ricercatore per ottenere migliori stime, includendo unità con particolari caratteristiche con una maggiore probabilità (ad es., campionamento stratificato in cui il ricercatore seleziona le donne con una probabilità doppia rispetto agli uomini).

Perché è così importante conoscere la probabilità di un'unità della popolazione di entrare a far parte del campione? L'importanza è dovuta al fatto che, nel modello matematico alla base del campionamento probabilistico, le probabilità sono indispensabili per il calcolo di stimatori non distorti[1] del parametro studiato e della sua varianza campionaria (Kish, 1992; Stuart, 1996) (cfr. appendice 1). In un linguaggio più discorsivo si può comprendere la necessità della conoscenza delle probabilità, se si riflette sul fatto che qualsiasi metodo di stima campionaria si fonda sul seguente principio: le unità comprese nel campione rispondono al questionario in rappresentanza delle rimanenti unità della popolazione che non sono entrate a far parte del campione. Tale principio si realizza praticamente attribuendo a ciascuna unità inclusa nel campione un peso che può essere visto come il numero di persone della popolazione "rappresentate" dalla persona che risponde al questionario. Se, ad esempio, ad un'unità campionaria di sesso femminile di età inferiore ai quarant'anni viene attribuito un peso pari a 32 , questo indica che essa risponde ad una domanda per se stessa e per altre 31 donne con meno di quarant'anni che fanno parte della popolazione, ma che non sono state selezionate per partecipare all'indagine; è come se la rispondente avesse ricevuto la delega a rispondere per altre 31 donne. Il peso da attribuire ad ogni soggetto viene calcolato a partire dalla probabilità dell'unità di entrare a far parte del campione[2] (senza la probabilità non è possibile

1. «Per evitare malintesi è bene ricordare che le affermazioni della teoria dei campioni acquistano significato solo se riferiti all'universo dei campioni e non riguardano, se non indirettamente, il singolo campione osservato o la stima da esso ricavata» (Herzel, 1991, p. 628). In statistica con il termini di stimatore "non distorto" o "corretto" si intende che se noi potessimo estrarre da una popolazione data, ma del tutto arbitraria, con un dato piano di campionamento casuale un milione di campioni diversi di una data numerosità e calcolassimo un milione di percentuali (ad es. della variabile genere) lo stimatore percentuale è corretto se la media del milione di percentuali del genere (una per ogni campione) è uguale al valore della percentuale nella popolazione studiata. In altre parole il concetto di "corretto" o "non distorto" non è associato ad una singola percentuale (ad esempio quella calcolata nel cinquantesimo campione), il cui valore potrebbe essere anche molto differente dal valore della popolazione.

2. Esso non è altro che il reciproco della probabilità di entrare a far parte del campione. Va sottolineato che nelle indagini effettive il peso da attribuire a ciascuna unità è ottenuto non solo in base ad una procedura di calcolo che corregge il peso determinato a partire dalla probabilità di entrare a far parte del campione, ma anche per attenuare l'effetto distorsivo sulle stime dovuto ad eventuali errori di mancata risposta e di non copertura.

calcolare il peso e senza quest'ultimo non si possono calcolare stime non distorte).

1.3. Popolazioni senza lista di campionamento: il meccanismo si inceppa

Stabilita l'importanza di conoscere la probabilità di essere estratti, diventa importante capire come calcolarla. In tutte le forme di campionamento che potremmo definire "classiche", per poter calcolare la probabilità di un'unità di essere inclusa nel campione è necessario disporre della lista di tutte le unità che fanno parte della popolazione studiata[3]. La lista di campionamento dovrebbe essere, nei limiti del possibile, completa cioè contenere tutte le unità di analisi che compongono la popolazione oggetto di studio[4].

La situazione ideale sarebbe quella di avere una lista di campionamento in cui sono elencati tutti i membri della popolazione oggetto di studio. Se rappresentiamo come due quadrati di uguali dimensioni la lista di campionamento e la popolazione di riferimento, la situazione ideale è quella di perfetta sovrapposizione. In realtà nella maggioranza dei casi non esiste una perfetta sovrapposizione, la situazione nella pratica della ricerca è rappresentabile come in figura 1.1. In ogni caso, se in una ricerca reale si può ragionevolmente sostenere che la copertura è elevata e le differenza fra membri "non coperti" e "coperti" è limitata, si è legittimati ad utilizzare la lista di campionamento, altrimenti si cerca di costruirne una completa (magari unendo due o più liste parziali) oppure si ridefinisce la popolazione oggetto di studio per adattarla alle caratteristiche della lista di campionamento disponibile (Caselli, 2005).

3. Nel campionamento a stadi è necessario avere la lista delle unità che possono essere selezionate ad ogni stadio.

4. In termini teorici è possibile estrarre casualmente i membri di una popolazione anche «in assenza di una lista preventiva, purché esista un luogo dove tutta la popolazione sia localizzata ed il ricercatore possa passare in rassegna, nel corso della selezione, tutti i soggetti» (Corbetta, 1999, pp. 332–333) Se, ad esempio, tutte le persone devono passare da un punto preciso (ingresso di un museo, sportello di un ufficio pubblico) e io ne intervisto, ad es. una ogni cinque, il campione finale risulta essere di tipo casuale. La rarità di queste applicazioni conferma comunque l'importanza di una lista di campionamento preventiva.

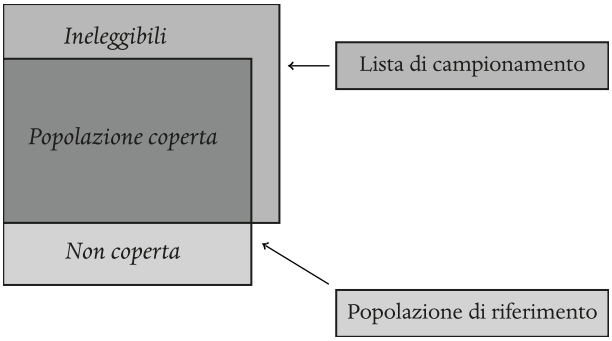

Figura 1.1. Copertura della lista di campionamento della popolazione di riferimento.

Oltre alla lista è necessario utilizzare una procedura di selezione che garantisca a tutte le unità la possibilità (anche molto piccola, ma diversa da zero) di essere estratte. La procedura migliore per garantire questa possibilità consiste nel selezionare casualmente un certo numero di unità dalla lista[5]. Per chiarire questi due punti (lista di campionamento e procedura di selezione casuale) si può ricorrere ad un esempio molto utilizzato nei manuali di statistica: quello dell'estrazione da un contenitore. Si crea una lista degli elementi della popolazione oggetto di indagine (es. famiglie residenti in una provincia, persone maggiorenni residenti in un comune). Ad ogni elemento della popolazione si associa una pallina con un numero progressivo.

5. « La teoria dei campioni, nella sua accezione tradizionale, si fonda comunque esclusivamente sul campionamento casuale, nelle sue svariate forme. Più precisamente, questa teoria studia le proprietà dell'insieme costituito da tutti i campioni che possono essere estratti, ossia osservati, da una popolazione data, ma del tutto arbitraria, con un dato piano di campionamento casuale » (Herzel, 1991, p. 626). A patto di aver estratto un campione casuale (e solo a questa condizione), la teoria dei campioni (piano sintattico) permette di valutare, sulla base dei soli dati campionari, il grado di attendibilità delle stime. Praticamente permette di calcolare un intervallo nella forma:

$$\text{stima} \pm \text{margine di errore campionario}$$

in modo da avere "un'elevata sicurezza" che il valore del parametro della popolazione è da qualche parte tra i due valori, inferiore e superiore, dell'intervallo. Si sottolinea che questa affermazione è veritiera, in una reale indagine campionaria (piano semantico), solo nel caso siano di entità trascurabile altri tipi di errori, chiamati "non campionari" (ad es.: errori dovuti alla mancata partecipazione delle persone campionate, a formulazioni poco chiare delle domande, allo scarso impegno o cattiva volontà degli intervistatori e degli intervistati).