



*La cassetta degli attrezzi. Strumenti per le scienze umane*

Direttore

*Giovanni Di Franco*, Università di Salerno

Comitato editoriale

*Elena Battaglini*, Ires-Cgil

*Sara Bentivegna*, Università di Roma

*Alberto Marradi*, Università di Firenze

*Federica Pintaldi*, Istat

*Luciana Quattrococchi*, Istat

*Marta Simoni*, Iref-Acli

La collana, rivolta a ricercatori accademici e professionisti, studiosi, studenti, e operatori del variegato mondo della ricerca empirica nelle scienze umane, si colloca sul versante dell'alta divulgazione e intende offrire strumenti di riflessione e di intervento per la ricerca.

Obiettivo è consolidare le discipline umane presentando gli strumenti di ricerca empirica, sia di raccolta sia di analisi dei dati, in modo intellegibile e metodologicamente critico così da consentirne l'applicazione proficua rispetto a definiti obiettivi cognitivi.

I testi sono scritti da professionisti della ricerca che, attingendo alla personale esperienza maturata in anni di attività, offrono ai lettori strumenti concettuali e tecnici immediatamente applicabili nella propria attività di ricerca.

Tutti i volumi pubblicati sono sottoposti a referaggio.

I lettori che desiderano informarsi sui libri e le riviste da noi pubblicati possono consultare il nostro sito Internet: [www.francoangeli.it](http://www.francoangeli.it) e iscriversi nella home page al servizio "informazioni" per ricevere via e-mail le segnalazioni delle novità o scrivere, inviando il loro indirizzo, a: "FrancoAngeli, viale Monza 106, 20127 Milano".

Giovanni Di Franco

# **IL CAMPIONAMENTO NELLE SCIENZE UMANE**

Teoria e pratica

La cassetta degli attrezzi  
Strumenti per le scienze umane/4

**FrancoAngeli**

Progetto grafico di copertina di Maria Teresa Pizzetti

Copyright © 2010 by FrancoAngeli s.r.l., Milano, Italy.

*L'opera, comprese tutte le sue parti, è tutelata dalla legge sul diritto d'autore. L'Utente nel momento in cui effettua il download dell'opera accetta tutte le condizioni della licenza d'uso dell'opera previste e comunicate sul sito [www.francoangeli.it](http://www.francoangeli.it).*

119. *La cassetta degli attrezzi. Strumenti per le scienze umane*

Volumi pubblicati:

1. Giovanni Di Franco, *L'analisi dei dati con SPSS. Guida alla programmazione e alla sintassi dei comandi*
2. Silvia Cataldi, *Come si analizzano i focus group*
3. Federica Pintaldi, *Come si analizzano i dati territoriali*
4. Giovanni Di Franco, *Il campionamento nelle scienze umane: dalla teoria alla pratica*
5. Lucia Coppola, *NVivo un programma per l'analisi qualitativa*
6. Simone Gabriellini, *Simulare meccanismi sociali con NetLogo. Una introduzione*

Volumi in preparazione:

7. Alberto Marradi, *Come evitare gli errori tipici in un questionario*



# Indice

<b>1. Introduzione</b>	pag.	9
<b>2. I concetti fondamentali del campionamento</b>	»	19
2.1 Cenni di storia del campionamento	»	20
2.2 Probabilità come frequenza relativa	»	23
2.3 Variabili casuali, loro distribuzioni e cenni di calcolo combinatorio	»	29
2.4 Il campionamento e il Teorema del limite centrale	»	34
2.5 Intervalli di fiducia e inferenza campione/popolazione	»	40
2.6 Cosa leggere per saperne di più	»	49
<b>3. Il campionamento nelle scienze umane</b>	»	51
3.1 Le tecniche probabilistiche	»	53
3.2 Le tecniche non probabilistiche	»	60
3.3 Il problema della copertura	»	73
3.4 Il problema delle mancate risposte	»	77
3.5 Definire la dimensione di un campione casuale	»	83
3.6 Quando e perché ponderare i casi	»	91
3.7 Cosa leggere per saperne di più	»	95
<b>4. Rappresentatività e casualità</b>	»	97
4.1 Caso, casualità ed estrazione casuale	»	98
4.2 Il concetto di rappresentatività nei campioni probabilistici	»	100
4.3 Il concetto di rappresentatività nei campioni non probabilistici	»	106
4.4 Casualità e rappresentatività	»	107
4.5 Casualità e rappresentatività nei campioni usati		

nei sondaggi politico-elettorali	»	114
4.6 Cosa leggere per saperne di più	»	128
<b>5. Il campionamento ecologico</b>	»	131
5.1 Cos'è il campionamento ecologico	»	133
5.2 Scegliere l'unità di analisi territoriale	»	135
5.3 Costruire unità territoriali omogenee	»	138
5.4 Come procedere al campionamento ecologico	»	150
5.5 Cosa leggere per saperne di più	»	159

## 1. Introduzione

*Un'idea un concetto un'idea  
finché resta un'idea  
è solo un'astrazione  
se potessi mangiare un'idea  
avrei fatto la mia  
rivoluzione  
(Giorgio Gaber)*

*Alea jacta est* disse Cesare passando il Rubicone. Fin dai tempi antichi l'uomo acquisì dimestichezza con ciò che chiamò in vari modi: fato, destino, caso, caos, etc. Questa frase esprime la consapevolezza che prima di lanciare un dado sono possibili diversi eventi (ad esempio, può presentarsi una faccia numerata da uno a sei; può uscire un numero pari o un numero dispari); una volta che lo si è lanciato si verifica un solo evento e questo esito non si può cambiare. I giochi basati sul caso, come i dadi, allietarono gli esseri umani fin dall'antico Egitto, e anche da prima.

Pur disponendo di una forte familiarità con il caso, con i giochi basati sul caso, e con altri concetti affini, paradossalmente l'umanità arrivò a formulare il concetto e il relativo termine che più si connette al caso, e cioè 'probabilità' in tempi relativamente recenti (fra il XVI e il XVII secolo). Anzi, a dire il vero, ancora oggi non c'è una definizione completamente soddisfacente di cosa sia la probabilità.

Sul concetto di probabilità, e su altri fondamentali concetti di natura matematica di cui diremo nel secondo capitolo, si fonda la teoria statistica del campionamento e i teoremi che riguardano l'inferenza statistica: ossia i procedimenti che consentono di generalizzare i risultati ottenuti con un campione probabilistico all'intera popolazione di riferimento.

Un campione si dice 'probabilistico' se, e solo se, i casi che ne fanno parte sono scelti con un'estrazione casuale. La metafora più chiara di un'estrazione casuale è l'estrazione dei numeri del Lotto. Un'urna contiene novanta bussolotti ciascuno contrassegnato da un numero che va da uno a novanta. Prima di estrarre un numero si fa

girare l'urna in modo che i bussolotti si mescolino. Poi si apre l'urna e si estrae un bussolotto. Si va avanti così fino all'estrazione dell'ultimo bussolotto previsto. Conoscendo l'intera popolazione dei numeri da uno a novanta e avendoli inclusi tutti nell'urna, sappiamo che ogni numero ha una probabilità su novanta di essere il primo estratto. Per il secondo numero estratto, visto che il primo estratto non viene reinserito nell'urna, la probabilità di estrazione è di uno su ottantanove, e così via per i successivi numeri estratti.

Ci sono diversi tipi di campionamento con estrazione casuale che saranno illustrati nel terzo capitolo. Per ciascuno conosciamo la probabilità che ogni elemento della popolazione ha di essere estratto nel campione. Definiamo 'casuale' un campione che deriva da un'estrazione dove ciascun elemento della popolazione ha *la stessa probabilità* di essere estratto; 'probabilistico' un campione che deriva da un'estrazione casuale di elementi che hanno una *probabilità nota e non nulla*, di essere estratti.

Oltre ai campioni casuali e probabilistici, esistono altri procedimenti nei quali il criterio di estrazione dei casi si dice "a scelta ragionata". In tali procedimenti è il ricercatore che, sulla base delle sue necessità cognitive, determina le regole per stabilire quali casi andranno a far parte del campione. Il fatto che la scelta dei casi non sia casuale definisce questi campioni come non probabilistici; ad essi non è possibile applicare i procedimenti dell'inferenza statistica: i risultati ottenuti con campioni non probabilistici non sono generalizzabili all'intera popolazione di riferimento. Per molti questo limite rende di fatto i campioni non probabilistici praticamente inutilizzabili nella ricerca; come vedremo nei capitoli seguenti, le cose non stanno proprio così.

Infine, ci sono altri campioni – sarebbe meglio chiamarli pseudo-campioni – che potremmo definire "accidentali" e/o "a casaccio" e/o "autoselezionati", per i quali di fatto non esiste una regola in base alla quale si determina quali casi debbano entrare nel campione, oppure si lascia alle persone la decisione di entrare a fare parte di un campione. Ad esempio, intervistando le persone che escono da una stazione ferroviaria un dato giorno in una data ora; intervistando persone fermate per strada; chiedendo di compilare un questionario inserito in un sito web. I risultati ottenuti con questi pseudo campioni non

possono rappresentare niente altro che le caratteristiche delle persone che ne fanno parte. In tutti i manuali si sconsiglia decisamente di adottare campioni simili.

Non è nostra intenzione produrre un ennesimo testo di teoria del campionamento che illustri le formule e i teoremi della statistica inferenziale. Su tali temi esiste una consolidata letteratura sia specialistica sia introduttiva. Quello che manca è un testo che, a partire dall'illustrazione dei concetti chiave, necessari per comprendere l'oggetto della discussione, si prefigga di illustrare come la teoria viene applicata nelle operazioni di ricerca empirica nelle scienze umane.

Sorgono numerosi problemi nella traduzione della teoria statistica del campionamento in operazioni di ricerca – soprattutto, come capita molto spesso, quando l'unità di analisi è l'individuo – per varie ragioni che verranno esposte nel testo. In particolare, occorre sfatare un certo numero di luoghi comuni e stereotipi che si incontrano leggendo i rapporti di ricerca, e in particolar modo nei cosiddetti documenti metodologici dei sondaggi d'opinione che ormai da anni imperversano nei mass media del nostro Paese. Questi documenti delle agenzie si caratterizzano per l'uso mistificatorio delle definizioni e del linguaggio tecnico al fine di presentare i loro campioni sotto una veste di legittimità statistica che in realtà non possiedono affatto. Proclamano con enfasi di aver costruito un campione casuale e probabilistico mentre nella migliore delle ipotesi i loro campioni possono essere classificati come semplici campioni per quote e quindi di tipo non probabilistico, se non addirittura campioni “a casaccio” nei quali i soggetti intervistati sono autoselezionati (vedi capitolo quattro, paragrafo 4.5).

Obiettivo del testo è presentare una serie di buone pratiche da seguire nel rispetto degli obiettivi cognitivi di una ricerca senza trascurare, ignorare o addirittura violentare, la teoria del campionamento così come essa si è sviluppata negli ultimi decenni. In tale prospettiva si propone di mettere al centro gli obiettivi cognitivi del ricercatore e rispetto a questi adattare, quando possibile e quando effettivamente utile, le procedure di campionamento ai criteri imposti dalla teoria statistica. In altri termini, si vuole suggerire un rovesciamento della prassi consolidata che consiste nell'adottare – troppo spesso in modo del tutto acritico – le procedure per così dire canoniche, senza

che queste siano né effettivamente praticabili, né compatibili con gli obiettivi cognitivi del ricercatore.

Per molti ricercatori il campione è un problema che con una certa ansia si vorrebbe lasciare alle spalle. Quello che conta è poter dichiarare di aver rispettato i criteri dettati dalla dottrina statistica senza minimamente essersi posti il problema della compatibilità di tali preceetti con i propri obiettivi di ricerca, per mettere al riparo la propria ricerca da possibili critiche rivolte alla forma di campionamento. Fra queste la più frequente: “il tuo campione non è rappresentativo”, anche se il concetto di rappresentatività statistica è molto più complesso e difficile da valutare rispetto a quanto abitualmente si creda (vedi capitolo quattro).

Da qui nasce l'uso fuorviante dell'aggettivo ‘scientifico’ che associato al sostantivo ‘campione’ mette – o meglio si pensa che metta – preventivamente al riparo da qualsiasi possibile critica. Ma quale significato fattuale possiede l'espressione ‘campione scientifico’? Il più delle volte il campione effettivamente estratto in una certa ricerca è tutto fuorché un campione probabilistico, ossia quel tipo di campione che la dottrina impone per poter applicare i teoremi dell'inferenza statistica, che permettono di generalizzare i risultati da un campione alla popolazione di riferimento (vedi capitolo due).

L'uso feticistico dell'aggettivo ‘scientifico’ serve – o meglio dovrebbe servire – a fornire credibilità ai risultati di una ricerca. Questo è uno dei tanti esempi di termini usati dai ricercatori nelle scienze umane (altri esempi sono ‘misurazione’, ‘verifica’, ‘esperimento’, etc.) per mostrare di avere le carte in regola con un'immagine istituzionalizzata, che in realtà è uno stereotipo di scienza.

Per spiegare la mancanza di consapevolezza del vano tentativo di rincorrere l'ideale della ‘correttezza statistica’ per il proprio campione si devono chiamare in causa molti manuali di metodologia della ricerca, nei quali i concetti alla base del campionamento sono acriticamente mutuati dalla statistica senza porsi il problema di un raccordo con le effettive situazioni di ricerca nelle scienze umane, che molto spesso non sono compatibili con un campione probabilistico.

Per questi ed altri motivi che illustreremo nel testo, è importante proporre nuovi concetti e nuovi procedimenti che siano più facilmente attuabili nelle operazioni di ricerca e soprattutto più coerenti con

gli obiettivi cognitivi dei ricercatori. In questo testo tenteremo di fornire un primo contributo in questa direzione (capitolo cinque).

Il campionamento è un problema fondamentale per qualsiasi ricerca empirica poiché, quale che sia l'oggetto di studio, quasi sempre è impossibile studiarne tutti i casi per vincoli di natura temporale, economica, etc. Tuttavia, in genere ci si prefigge di scoprire qualcosa che possa applicarsi a tutti i casi di un certo tipo, generalizzando i risultati relativi al campione. I ricercatori devono quindi ritagliare una parte che sia rappresentativa dell'intero. Si tratta di un procedimento che equivale alla figura retorica chiamata "sineddoche" che consiste nel prendere una parte per raffigurare il tutto.

Sulla base di questi assunti possiamo formulare le seguenti tesi che discuteremo nei prossimi capitoli:

1. Il campionamento casuale (e in generale i campioni probabilistici) sono una soluzione efficace solo per *alcuni* problemi cognitivi quando le ricerche adottano un'unità di analisi sia individuale sia non individuale. In effetti c'è una forte tendenza nel considerare l'individuo – o altre unità composte da individui, come la famiglia, la categoria professionale, etc. – l'unità di analisi privilegiata delle scienze umane. A mio parere questa tendenza andrebbe mitigata. Nella storia della ricerca, esistono feconde tradizioni di ricerca che si sono avvalse di altri tipi di unità come le ricerche ecologiche, le ricerche sui mass-media, le ricerche su documenti e altre fonti, le ricerche sulle reti sociali, etc. Quando si adottano unità di analisi di questo genere è molto più semplice procedere con campionamenti casuali o probabilistici. Quando l'unità di analisi è l'individuo, possiamo procedere ad un campione casuale solo se la popolazione è completamente conosciuta e disponiamo di un elenco completo della stessa. Ad esempio, quando vogliamo valutare l'esito professionale dei laureati di una certa Facoltà universitaria, o quando vogliamo conoscere le motivazioni dei soggetti che intendono arruolarsi nell'esercito. In entrambi i casi i nomi di tutti i soggetti della popolazione sono registrati da apposite procedure amministrative e, quindi, sono facilmente reperibili.

2. Molto spesso gli obiettivi cognitivi delle ricerche non hanno nulla a che fare con l'inferenza statistica. Di solito gli obiettivi riguardano lo studio delle relazioni fra molte variabili di cui non conosciamo le distribuzioni nella popolazione. In questi casi disporre o no di un campione probabilistico non fa una grossa differenza perché non vogliamo effettuare alcuna generalizzazione. Fra l'altro, occorre sottolineare che quando l'analisi dei dati prevede lo studio delle relazioni fra più variabili, anche disponendo di un campione probabilistico – che sostanzialmente è determinato in base alla conoscenza di una sola o comunque di poche caratteristiche note della popolazione – non avremmo comunque maggiori garanzie rispetto alle analisi multivariate. Sarebbe diverso se disponessimo di un campione costruito a partire dalle distribuzioni univariate e multivariate oggetto dell'analisi. Ma, paradossalmente, se disponessimo di queste informazioni sulla popolazione non avremmo alcun bisogno di costruire un campione. Evidentemente, è necessario trovare un ragionevole compromesso fra le esigenze cognitive della ricerca e la necessità di disporre di una base di dati che risulti sufficiente a soddisfare tutte le esigenze che possono emergere in sede di analisi multivariate. In generale, un principio fondamentale è che i casi del campione della ricerca devono essere fra loro il più eterogenei possibile. Più eterogeneo è il campione più proficua e interessante sarà l'analisi dei dati. Estremizzando il ragionamento, se avessimo, come alle volte capita, un campione di casi molto simili, o addirittura del tutto simili (situazione che nel linguaggio dell'analisi dei dati si dice di bassa o nulla eterogeneità – per le variabili categoriali – e di bassa o nulla varianza – per le cardinali) non si potrebbero analizzare le relazioni fra variabili, perché non vi sarebbe varianza da porre in relazione.
3. Abitualmente gli obiettivi cognitivi delle ricerche sociologiche più che essere quelli compatibili con la statistica inferenziale, e quindi attraverso la costruzione di un campione casuale o probabilistico, sono molto simili a quelli di altre discipline come l'archeologia e la paleontologia. Ad esempio, un paleontologo trova solo alcune ossa di una particolare spe-

cie di dinosauro e vuole ricostruire (inferire) l'intero scheletro dell'animale. Nello stesso modo un archeologo trova alcuni resti di una antica città e vuole ricostruire l'intera città (i suoi confini, il tipo di abitazioni, la struttura sociale, il tipo di attività economiche, etc.). In queste fattispecie non è di alcun aiuto il concetto di estrazione casuale in quanto si tratta di ricostruire un insieme a partire da alcune sue tracce che ne costituiscono una piccola parte non casualmente estratta. Molte ricerche affrontano problemi di questo genere. Spesso siamo interessati a ricostruire la rappresentazione sociale di un'istituzione, di un'organizzazione, e così via, partendo da un insieme limitato e circoscritto di suoi elementi. Sotto un certo punto di vista è un procedimento che può richiamare quello con cui rappresentiamo un concetto generale scomponendolo in dimensioni e definendo operativamente i relativi indicatori per raccogliere un insieme di dati dai quali risalire al concetto di partenza. Evidentemente questo problema ha una natura diversa rispetto a chiedersi quale sia la media o la proporzione di una variabile in una popolazione conoscendo la media o la proporzione di quella stessa variabile in un campione.

4. Un ulteriore problema tipico della ricerca sociale che non può essere risolto con il campionamento è conoscere la gamma completa di stati che assume un dato fenomeno. Quali sono tutti i diversi modi in cui le persone hanno organizzato le proprie relazioni interpersonali? Qual è il ventaglio degli stili di vita delle persone di una certa società? Non desideriamo che il nostro campione abbia caratteristiche specifiche di un qualche sottogruppo dell'insieme, caratteristiche che il ricercatore incauto potrebbe attribuire all'intera classe. A pensarci bene capita molto spesso di trovarsi in questa situazione. Effettuare un campione casuale stratificato richiede la conoscenza delle distribuzioni di una o più proprietà importanti nell'intera popolazione. Ma se non conosciamo la distribuzione di alcuna proprietà nella popolazione (caso molto frequente nella ricerca sociologica) non potremo assumere il campione che costruiremo come rappresentativo della popolazione su alcuna proprietà.

La statistica inferenziale ha adottato il concetto di rappresentatività per connotare il campione costruito in una ricerca come adeguato a rappresentare approssimativamente l'intera popolazione di cui è parte. Questo significa che il campione in piccolo riproduce le stesse proporzioni presenti nell'intera popolazione per tutte le distribuzioni prese in considerazione, con lievi scostamenti dovuti al caso. Per cui si può parlare di rappresentatività solo se conosciamo sia la popolazione sia il campione che da questa è stato estratto. Ma se fossimo in grado di controllare totalmente quanto un campione è rappresentativo di una popolazione vorrebbe dire che non avremmo nulla di nuovo da conoscere della popolazione stessa, e pertanto, non avremmo neanche bisogno di fare la ricerca. Questo è il paradosso della rappresentatività statistica.

Da questa breve discussione emerge, o dovrebbe emergere, una qualche forma di diffidenza verso il concetto di rappresentatività statistica quando è esteso oltre l'estendibile (vedi il capitolo quattro).

Diviene quindi opportuno sostituire o quanto meno affiancare al concetto di rappresentatività statistica altri concetti di rappresentatività che potremmo definire "teorica-tipologica" e "pragmatica".

Per rappresentatività teorica o tipologica intendiamo quei campioni che sono costruiti sulla base di una schema tipologico e di una teoria che consentano di enucleare le caratteristiche di una popolazione che devono essere riprodotte nel campione. Ad esempio, poniamo di voler condurre una ricerca su un campione di elettori italiani. Nella costruzione del campione, piuttosto che considerare le solite proprietà socio-demografiche, potremmo adottare la classica tipologia degli elettori proposta nel 1977 da Parisi e Pasquino che individua i seguenti tre tipi: l'elettore di appartenenza, l'elettore d'opinione e l'elettore di scambio. Ciascun tipo individua una diversa motivazione nel comportamento e nelle scelte degli elettori. L'elettore di appartenenza ribadisce con il voto la sua identificazione con un partito. Spesso si tratta di un legame di lunga data ed è condiviso all'interno del gruppo familiare e/o amicale del soggetto. L'elettore d'opinione prima di decidere come votare valuta l'offerta proposta dai partiti in competizione per un data elezione. Dopo aver comparato i diversi programmi e le diverse proposte sceglie sulla base delle sue convinzioni e delle sue considerazioni circa la qualità dei candidati in cam-

po. Infine, l'elettore di scambio baratta il suo voto per conseguire un interesse personale, il più delle volte a carattere privatistico e/o familistico.

Sulla base di ricerche precedenti e di analisi politologiche possiamo stimare la composizione nel corpo elettorale italiano di questi tre tipi di elettori. Sappiamo anche che la diffusione sul territorio è piuttosto eterogenea: gli elettori di opinione prevalgono nei grandi centri urbani, sono persone mediamente più istruite, che occupano livelli professionali medio alti, e così via. Gli elettori di appartenenza sono concentrati in prevalenza nelle cosiddette zone 'rosse' e nelle aree di primo insediamento della Lega Nord; gli elettori di scambio sono molto presenti nel Sud.

Vista la rilevanza della tipologia degli elettori ai fini di una ricerca che abbia come oggetto l'analisi dei comportamenti elettorali, è evidente che sarebbe molto conveniente costruire un campione che sia rappresentativo della tipologia indicata, ossia che includa approssimativamente i tre tipi di elettorato così come essi sono diffusi nell'intera popolazione.

Naturalmente, queste conoscenze devono essere in possesso del ricercatore che deve essere in grado di esplicitarle prima della formazione del campione. Possono derivare: dalla conoscenza del fenomeno e del contesto nel quale si svolge la ricerca; da ricerche precedenti; da teorie di riferimento; etc.

La rappresentatività pragmatica può essere considerata una forma debole della rappresentatività teorico-tipologica. Quando si affrontano per la prima volta fenomeni o oggetti di studio non si è in grado di avvalersi di conoscenze relative a precedenti ricerche. In questi casi il ricercatore può e deve comunque pianificare un campione che tenga conto della distribuzione di alcuni caratteri della popolazione di riferimento ritenuti importanti in sede di progetto della ricerca.

In conclusione, piuttosto che pretendere di raggiungere una rappresentatività globale (ossia che il campione sia una miniatura di tutte le innumerevoli proprietà della popolazione, cosa ovviamente irraggiungibile), è molto più utile che il ricercatore concentri la sua attenzione su quelle proprietà che ritiene essenziali per ragioni teoriche rispetto agli obiettivi della sua ricerca. In questa prospettiva si deve

adottare un procedimento di estrazione che non si basi su formule matematiche, ma su conoscenze.

Estendere il significato di rappresentatività dal piano meramente statistico a quello teorico/tipologico e pragmatico può essere in qualche modo assimilato ai vari significati assunti dal concetto di validità.

In genere si usa distinguere fra la validità a vista, validità per criterio e validità per costruito.

La prima forma è quella basata su un giudizio semantico del ricercatore o dei ricercatori.

La validità per criterio riconosce valido un indicatore se è statisticamente associato a un altro indicatore ritenuto valido.

La validità secondo costruito riconosce valido un indicatore se la sua relazione con una o più altre variabili soddisfa le aspettative che deriviamo da una teoria. Ad esempio, vogliamo valutare la validità di un indicatore del concetto di autostima e la teoria di riferimento postula una relazione positiva tra questa proprietà, la partecipazione alle attività scolastiche e i risultati scolastici. Se in base ai risultati della ricerca emerge effettivamente una relazione significativamente positiva fra l'indicatore di autostima con gli indicatori di impegno scolastico e i risultati scolastici, otteniamo una validazione per costruito.

La rappresentatività pragmatica può essere considerata analoga alla validità a vista; la rappresentatività teorico-tipologica può essere considerata analoga alla validità mediante criterio.

Riteniamo che l'adozione di questi concetti di rappresentatività in molti casi possa orientare il ricercatore, nella scelta del campione della sua ricerca, senza costringerlo a rincorrere criteri che molto spesso non è possibile soddisfare alla luce delle risorse disponibili e/o delle caratteristiche della popolazione di riferimento.

Desidero ringraziare l'amico e maestro Alberto Marradi per aver letto alcune parti del testo e criticato le mie tesi sul campionamento.

## 2. I concetti fondamentali del campionamento

Di seguito illustro i concetti di base della statistica inferenziale e del campionamento per consentire al lettore di seguire consapevolmente il contenuto dell'intero testo. Sebbene l'indagine campionaria sia una pratica di ricerca assolutamente consolidata nel panorama delle scienze umane, solo di recente la riflessione sul campionamento costituisce un tema di interesse nella letteratura metodologica nazionale. Preliminarmente è utile definire alcuni termini di base: 'popolazione' è l'insieme di persone (o cose) rispetto al quale siamo interessati a generalizzare (estendere) i risultati della ricerca, di solito si rappresenta con  $N$ ; 'parametro della popolazione' è il valore incognito della popolazione (di solito si tratta di una proporzione o percentuale, o media) che sarà oggetto dell'inferenza, di solito si rappresenta con una lettera dell'alfabeto greco; 'lista di campionamento' è la lista delle persone (o cose) a partire dalla quale è estratto il campione (in effetti, nella pratica la popolazione coincide con la lista di campionamento); 'campione' è l'insieme di alcuni casi estratti dalla popolazione sui quali si effettua la rilevazione, di solito si rappresenta con  $n$ ; 'stima' è il valore di una proporzione o di una media calcolato nel campione che si pone come base per inferire i valori del parametro della popolazione, di solito si rappresenta con una lettera dell'alfabeto latino.

In molti manuali di metodologia della ricerca i concetti alla base del campionamento sono acriticamente mutuati dalla statistica inferenziale nell'illusione che la loro formalizzazione costituisca di per sé garanzia di solidità scientifica e di indiscutibile consenso da parte della comunità scientifica. Tale concezione viene però facilmente