

4. La ricerca longitudinale

di Amalia Caputo

Lo studio del cambiamento di stato di specifici fenomeni può avvenire attraverso due approcci opposti, mediante analisi degli andamenti e dei *trend* temporali oppure mediante comparazioni spaziali; nel primo caso parleremo di ricerca longitudinale nel secondo di ricerca trasversale. La sostanziale differenza dei due tipi di ricerca è evidente già a partire dal significato dei termini: l'aggettivo 'longitudinale' deriva da longitudine (lunghezza), letteralmente «che procede, si sviluppa, è disposto nel senso della lunghezza», in contrapposizione a trasversale che, al contrario fa riferimento alla «larghezza». Sembra evidente che ciò che contraddistingue i due tipi di ricerca è la componente tempo: mentre gli studi trasversali rilevano informazioni su un oggetto in un determinato tempo, quelli longitudinali rilevano informazioni dello stesso oggetto in un tempo ripetuto o prolungato. Tralasciamo momentaneamente la ricerca trasversale rinviandola al capitolo successivo e affrontiamo nel dettaglio quella longitudinale.

La ricerca longitudinale è caratterizzata da quattro elementi (Menard, 2002): i dati vengono raccolti per ciascuna variabile in due o più periodi di tempo successivi; nelle differenti rilevazioni la popolazione (campione) è la medesima, se si tratta di diverse popolazioni (campioni) queste devono risultare equivalenti ossia «la popolazione rimane definita negli stessi termini (ad es., confini geografici, limiti di età, ecc.), ma può cambiare composizione nel tempo» (Duncan, Kalton, 1987, p.97); l'esito dell'analisi è una comparazione tra i dati ottenuti nelle differenti rilevazioni; infine, consentono non solo di individuare le cause che hanno dato vita ai fenomeni e, in molti casi, anche di stabilire la direzione del segno e della grandezza delle relazioni causali, questione che in questa sede, diviene particolarmente rilevante. Nelle indagini sociali molto spesso l'obiettivo conoscitivo che muove la ricerca è quello di definire i nessi causali che sottendono il legame di determinati fenomeni nella società. Quando si ritiene che un dato evento che comporta cambiamento sia da considerarsi conseguenza o agente nello sviluppo di determinate configurazioni, il nesso che legherà i fattori intervenienti individuati all'evento oggetto di studio si definirà causale o di causa-effetto (Feinstein, 1979).

Nella determinazione di tale nesso attraverso l'indagine empirica, generalmente, si procederà confrontando il cambiamento osservato tra soggetti sottoposti o non sottoposti ai fattori ritenuti incidenti e intervenienti rispetto alla variazione di status ricercata. Ciò equivale ad assumere un approccio valutativo che compara la variazione delle circostanze nel verificarsi di un cambiamento su un insieme di unità statistiche oggetto di indagine che siano tra loro equivalenti, eccetto che per l'essere state sottoposte/esposte o meno ad uno stimolo volto a introdurre consapevolmente variazioni. Il nesso causale nell'associazione tra fenomeni può essere derivato dal manifestarsi di cinque condizioni:

1. *consistenza*, ovvero l'associazione causale è indipendente dallo spazio e dal metodo usato per rilevarla;
2. *intensità*, il livello dell'effetto di cambiamento rilevato deve variare proporzionalmente alla variazione all'esposizione del fattore ritenuto interveniente o agente;
3. *specificità*, quando l'esistenza di un nesso di causalità tra il fattore e l'effetto è preponderante rispetto alla connessione tra l'effetto e altri fattori potenzialmente agenti o scatenanti;
4. *sequenzialità temporale*, l'effetto deve essere preceduto temporalmente dal manifestarsi del fattore o dei fattori considerati come intervenienti o agenti;
5. *logicità*, ovvero dal punto di vista sostantivo, il nesso causale deve poter essere considerato come plausibile.

Per quanto le prime tre condizioni siano valutabili statisticamente, le restanti due afferiscono all'ambito di valutazioni che possono essere condotte all'interno di uno specifico *framework* teorico e analitico di

riferimento. Non è semplice determinare con precisione i nessi di causa-effetto¹. All'interno di una sequenza temporale l'effetto può riprodursi quasi istantaneamente, cioè sembrano riprodursi nello stesso istante temporale; talvolta, invece, gli effetti si manifestano in tempi più lunghi, cioè tra causa ed effetto si verifica un ritardo (*delay o lag*) a cui bisogna prestare attenzione in fase di analisi. Oltre alla durata dell'intervallo temporale che intercorre tra causa ed effetto, bisogna fare attenzione ai possibili modi con cui un effetto può presentarsi. Nello specifico, un effetto può presentarsi in forma monotonica ovvero assumere un andamento lineare oppure ciclica ovvero in forma di andamento più complesso.

Nel caso della trattazione che seguirà ci si pone nell'ottica per la quale ad intervenire sugli effetti ricercati, oltre a un fattore specifico in relazione con l'evento, si inserisce anche il peso della dimensione temporale. È in questo senso che i disegni di indagine longitudinali diventano utili strumenti nello studio dei nessi di causalità ricercati.

L'obiettivo di questo tipo di disegni è, dunque, lo studio del cambiamento nel tempo sia individuale sia aggregato (Biolcati-Rinaldi, Vezzoni, 2012), nel primo caso si tratta di studiare il cambiamento dei singoli rispetto ad un determinato evento, nel secondo caso si tratta di studiare un cambiamento intervenuto in un aggregato di individui. Facciamo un esempio. Immaginiamo che il fenomeno sociale da studiare sia la genitorialità tra il 2013 e il 2017: studiare il cambiamento individuale significa rilevare che nel 2013 Anna non è madre e che nel 2017 Anna ha avuto il suo primo figlio, mentre studiare il cambiamento aggregato significa rilevare che in Italia tra il 2013 ed il 2017 il tasso di fertilità è in costante aumento.

I dati prodotti sia a livello individuale sia aggregato hanno per la ricerca sociale un valore conoscitivo molto elevato, quando si parla di studio del cambiamento, infatti, si fa implicito riferimento a molte aree di indagine. Seguendo lo schema della Ruspini (2004), si individuano almeno quattro possibili applicazioni dei dati longitudinali; oltre alla rilevazione dei cambiamenti nei valori di una variabile da un periodo all'altro e all'analisi della durata dei fenomeni sociali, i dati longitudinali consentono di identificare gli *sleeper effects* e le cause dei fenomeni sociali. L'individuazione degli *sleeper effects* ovvero la connessione tra eventi separati da periodi di tempo consistenti, permette di spiegare le condizioni attuali degli individui o degli aggregati sulla base del loro vissuto; ad esempio, la scelta di alcuni giovani di non studiare, di non formarsi e di non cercare un lavoro (NEET- *Not engaged in Education, Employment or Training*) potrebbe essere collegata alla situazione socioeconomica e culturale della famiglia d'origine ma anche, ad esempio, alle esperienze scolastiche negative del giovane.

I dati longitudinali consentono, poi, di individuare le cause dei fenomeni sociali in quanto consentono di specificare gli antecedenti del fenomeno e le conseguenze che ne derivano. Ciò significa che questo tipo di dati permette al ricercatore di esaminare non solo quella parte della popolazione che in differenti momenti si trova in una determinata condizione, ad esempio di disoccupazione, ma i flussi dentro e fuori tale condizione. In altre parole, ad esempio, è possibile ricostruire su un gruppo di disoccupati il rapporto tra la condizione di disoccupazione odierna e ciò che è avvenuto in precedenza (Olagnero, Saraceno 1993). E' implicito – come vedremo a seguire - che ciascun 'tipo di dato longitudinale' è definito dalle differenti modalità di raccolta che, a loro volta, fanno capo a differenti disegni di ricerca longitudinale.

4.1. Tipi di disegni longitudinali

Nell'ambito delle scienze sociali è possibile individuare tre tipi di disegni diacronici, il *repeated cross-sectional*, i *panel* o studi prospettici e gli studi longitudinali retrospettivi². Come anticipato, ciascun disegno si differenzia per le finalità conoscitive e per il tipo di dato prodotto.

Il *Repeated Cross-Sectional*, anche detto *studio trasversale ripetuto* o di *trend*, prevede che «misure simili siano fatte in punti diversi del tempo su campioni di una popolazione equivalente, senza che i casi siano gli

¹ Il rapporto causa-effetto affonda le radici nella sociologia classica, lo stesso Durkheim afferma che le cause dei fatti sociali vanno ricercate nei fatti antecedenti e non negli stati della coscienza individuale e che a ogni fatto corrisponde un'unica causa; per tale motivo, quando ad un fenomeno corrispondono varie cause è perché esistono diversi tipi dello stesso fenomeno (si pensi ai diversi tipi di suicidio individuati da Durkheim). Tuttavia si deve a Mill (1843) una vera e propria formalizzazione in cinque canoni di ricerca dei nessi causali: metodo della concordanza, metodo della differenza, metodo congiunto della concordanza e della differenza, metodo dei residui e, infine, metodo delle variazioni concomitanti.

² Nella pratica di ricerca si fa frequentemente ricorso ai cosiddetti studi longitudinali misti, ottenuti dalla combinazione dei tre tipi di strategie; un esempio è lo studio *cross-sectional* ripetuto della Banca di Italia sui bilanci delle famiglie nel qual è previsto anche un panel (Ruspini, 2004).

stessi nelle diverse occasioni di raccolta dati» (Duncan, Kalton 1987, p.97). Si tratta, in altre parole, di ripetere periodicamente una indagine trasversale estraendo però di volta in volta un nuovo campione avente le stesse caratteristiche del campione precedente (popolazione equivalente). In questo senso, il termine trasversale fa riferimento al campione costituito in modo eterogeneo, ad esempio, per età, istruzione, genere o provenienza. Per queste sue caratteristiche il *Repeated Cross-Sectional* è maggiormente adatto a rilevare il cambiamento aggregato piuttosto che quello individuale, tuttavia il fatto di non poter ricorrere allo stesso campione ma ad uno equivalente consente esclusivamente un'analisi del cambiamento nei tempi considerati, tecnicamente si dice un'analisi del cambiamento netto a livello aggregato; per tornare all'esempio di prima, consente di rilevare la variazione del tasso di fecondità dal 2000 ad oggi, senza che possano essere fatte ulteriori valutazioni. Nell'ambito delle scienze sociali numerosi sono gli esempi di indagini trasversali ripetute nel tempo, a titolo di esempio ricordiamo l'Eurobarometro e le indagini multiscopo sulle famiglie condotte dell'Istat, in entrambi i casi le inchieste vengono ripetute periodicamente ricorrendo allo stesso questionario somministrato a campioni diversi ma equivalenti. L'Eurobarometro fu istituito dalla Commissione Europea nel 1976 con l'obiettivo di rilevare due o cinque volte l'anno, le opinioni, gli atteggiamenti dei cittadini degli Stati membri su molte questioni, l'ampliamento della UE, la situazione sociale, la salute, la cultura, *l'information technology*, l'ambiente, l'Euro e la difesa in un'ottica comparativa. Le indagini multiscopo dell'Istat prevedono, invece, sette differenti ricerche sociali ciascuna condotta con cadenza temporale differente, quella annuale sugli aspetti della vita quotidiana, quella trimestrale sul turismo e le indagini sulle condizioni di salute e ricorso ai servizi sanitari, sui cittadini e il tempo libero, la sicurezza dei cittadini, le famiglie e soggetti sociali e l'uso del tempo che vengono effettuate a rotazione ogni cinque anni. I tipi di analisi che è possibile condurre con dati prodotti dalle indagini ripetute sono essenzialmente le analisi di *trend* e le analisi di *coorte* di cui parleremo a seguire. Le prime descrivono, a livello aggregato, l'andamento nel tempo di un fenomeno mediante specifiche tecniche descrittive ed inferenziali (cfr. cap. VI) che rilevano se l'andamento è stabile, lineare o curvilineo (Fig. 4.1). Immaginiamo di voler studiare il trend della partecipazione politica giovanile in Italia dal 2007 al 2017. Se l'andamento appare lineare, allora vuol dire che l'aumento (o la diminuzione) della partecipazione dei giovani è in costante aumento (o diminuzione) nel tempo; se è, al contrario curvilineo, allora significa che la partecipazione giovanile aumenta e diminuisce a seconda dei periodi.

Fig. 4.1. Trend: Aumento stazionario, lineare e curvilineo



I *panel* o studi longitudinali prospettici raccolgono informazioni sempre sullo stesso campione di soggetti e sempre sugli stessi temi, ma in periodi differenti. In pratica, una volta individuato il campione utile per ottenere informazioni sul fenomeno da studiare, i soggetti che lo costituiscono vengono periodicamente intervistati sullo stesso tema, la specificità è che le informazioni riguardano eventi nel loro svolgersi; il panel permette quindi di studiare il cambiamento a livello individuale consentendo di cogliere quelle specificità che gli studi sui cambiamenti aggregati non riescono di cogliere (Rose 2000, Ruspini, 2004). I panel sono dunque particolarmente adatti ad analizzare il cambiamento a livello micro e individuale e, quindi, ad individuare più facilmente le cause del cambiamento. I panel sono fondamentali per la ricerca sociale, «assolvono alla stessa funzione delle grandi infrastrutture di ricerca delle scienze naturali (...). I gruppi di ricerca che raccolgono dati del panel sviluppano le analisi necessarie a rispondere alle proprie domande di ricerca mettendo i dati a disposizione di altri gruppi e altre domande» (Biolcati-Rinaldi, Vezzoni 2012, 65), per questi motivi i disegni prospettici sono ritenuti i veri studi longitudinali. A fonte dell'enorme bagaglio di informazioni che portano, i panel presentano alcuni considerevoli svantaggi. Innanzitutto, hanno costi elevatissimi sia in termini economici che di forze impegnate per la loro implementazione e conduzione. Un secondo problema riguarda l'*attrition*, logoramento o mortalità del campione ossia l'interruzione della partecipazione del soggetto da una rilevazione all'altra dovuta a differenti fattori, dal semplice rifiuto al decesso dell'intervistato. A questo si aggiunge poi il problema dei *missing occasions* intervistati che abbandonano per un periodo e poi rientrano nel panel e di

coloro che si inseriscono nel panel a rilevazione già iniziata. Per risolvere in parte questi problemi si ricorre al *rotating panel*: ad ogni rilevazione (ondata-waves) si include per più di una rilevazione un sottocampione probabilistico tale da poter sopperire alle distorsioni intervenute tra una rilevazione e l'altra, ciò consente di mantenere invariate le caratteristiche del campione originario adattandolo via via agli inevitabili mutamenti che nel tempo possono intervenire³. Un'altra questione riguarda l'*interferenza* del panel; con il termine interferenza si fa riferimento alle possibili modifiche (o stabilità) degli atteggiamenti dei comportamenti e delle opinioni dovute non ad un reale cambiamento (o ad una persistente stabilità) ma ad un effetto ricordo che porta l'intervistato a modificare volontariamente la risposta fornendone una che dal suo punto di vista è socialmente condivisibile (desiderabilità sociale), oppure a dare sempre la stessa risposta per apparire coerente o, ancora a comportarsi effettivamente in modo diverso per il semplice fatto di sentirsi osservato.

Lo studio dell'evoluzione dei fenomeni sociali richiede che gli eventi siano rilevati a mano a mano che si verificano, in una sorta di continuità temporale. I panel e gli studi trasversali ripetuti, come abbiamo visto, rilevano il cambiamento solo in determinati momenti tralasciando, per forza di cose, quanto avviene da un tempo all'altro. Gli *studi longitudinali retrospettivi* rispondono proprio a questa esigenza. Al contrario di quanto accade negli studi prospettici in senso stretto, lo studio del cambiamento avviene attraverso due modalità: la raccolta di informazioni fornite da un campione di soggetti basate sulla ricostruzione del corso della vita individuale e familiare (storie di vita) oppure di storie di eventi specifici (*event histories*); sulla individuazione degli eventi, delle traiettorie e delle transizioni del vissuto ritenuti rilevanti nei processi di cambiamento⁴. Immaginiamo di voler studiare l'evoluzione lavorativa di un gruppo di soggetti, possiamo ricorrere ad una *event history* individuando uno specifico episodio che sia esplicativo del passaggio da uno stato di origine a quello di destinazione (Bossfeld, Rohwer 1995), ad esempio, il momento dell'assunzione segna il passaggio dallo stato di 'in cerca di prima occupazione' a quello di 'occupato'; nell'ambito di questo passaggio si pongono specifiche domande al campione di soggetti riguardanti il proprio vissuto e il contesto sociale ed economico entro cui gli eventi si sono svolti (come ha trovato lavoro?, il mercato in quel momento era favorevole? Che tipo di azienda che l'ha assunta la prima volta?). Sembra chiaro anche gli studi prospettici analizzano le evoluzioni dei fenomeni sociali a livello micro, ma «i corsi di vita individuali sono comprensibili solo se inseriti nel contesto delle traiettorie di vita: poiché gli eventi e i mutamenti che si collocano ad un livello 'micro' incidono potenzialmente sui corsi di vita individuali, il corso di vita non va isolato dalla 'situazione' in cui è immerso» (Ruspini 2004 54; Mayer 1991). Ciò implica che i dati prospettici consentono di analizzare non solo l'evoluzione dei fenomeni a livello micro ma anche a livello macro e quindi del contesto (sociale, economico e culturale) in cui l'individuo vive.

Il principale vantaggio questo tipo di studi longitudinali risiede proprio nella possibilità di raccogliere dati di durata (*event oriented observations*) o relativi a storie di eventi, ossia dati raccolti retrospettivamente e in grado di registrare il momento esatto in cui si è verificato un evento, tenendo conto però anche del tempo antecedente e successivo all'evento stesso. Gli studi longitudinali retrospettivi sono ovviamente soggetti a distorsioni nella ricostruzione degli eventi passati, distorsioni attribuibili – come vedremo nel prossimo paragrafo - quasi esclusivamente al fattore ricordo. Le distorsioni prodotte pongono quesiti non solo sulla qualità del dato ma sulla reale rilevazione dei processi di cambiamento che intervengono tra uno studio e l'altro tanto da far attribuire a questo tipo di disegno l'appellativo di studi quasi-longitudinali (Dex, 1995)⁵.

A conclusione di questo *excursus* sui disegni longitudinali riprendiamo la trattazione dello studio di *coorte* la cui peculiarità risiede nel fatto di potersi collocare sulla linea di confine tra i trend e i panel. L'obiettivo generale è di analizzare l'andamento di un evento vissuto da un gruppo di soggetti nello stesso ambito temporale «... il ricercatore sceglie una o più coorti di nascita e somministra un questionario ad un campione o all'intero gruppo: l'analisi longitudinale è dunque applicata a gruppi omogenei per età e una quantità di generazioni è seguita nel tempo» (Ruspini, 2004, 47). Se ogni generazione di soggetti viene seguita in un lasso

³ Nelle indagini classiche si ricorre agli *split panel*; come nel *rotating panel* una parte del campione definita ruotante viene intervistata per sopperire alle distorsioni intervenute, ma in questo caso i soggetti inclusi in questa parte del campione non vengono intervistati nella successiva rilevazione.

⁴ I corsi di vita sono definiti da *traiettorie*, *transizioni* ed *eventi*; le traiettorie sono i percorsi seguiti nell'ambito di una specifica esperienza con l'aumentare dell'età; le transizioni sono i cambiamenti di stato che intervengono entro una traiettoria; mentre, l'evento è il momento in cui si dà inizio ad una transizione (Olangero, Saraceno 1993; Ruspini 2004). Ad esempio, la condizione occupazionale è una traiettoria perché muta all'aumentare dell'età; nell'ambito della traiettoria sono identificabili dei momenti di transizione, come un cambiamento di ruolo, un aumento in termini economici; il momento del licenziamento o il pensionamento rappresentano gli eventi specifici che danno vita alle transizioni.

⁵ Rientrano tra i panel le *household panel survey*, i *linked panels* o panel amministrativi. Per una più ampia trattazione si veda Ruspini 2004.

temporale ampio lo studio rientra tra quelli panel; al contrario è definibile come insieme di studi trend se in ogni generazione e in ogni periodo di osservazione il campione viene scelto *ex novo*.

Per concludere, l'eterogeneità dei disegni longitudinali è dettata dalle grandi possibilità che questo tipo di studi ha nel rispondere a molti interrogativi conoscitivi. A questa eterogeneità corrisponde, come si avrà modo di esporre nelle pagine seguenti, una complessità sul piano delle procedure di analisi che variano a seconda del tipo di disegno a cui si fa riferimento.

4.1. Serie temporali e serie storiche

Una serie è un insieme di dati ordinato sulla base di un qualche criterio ed è sulla base di questo criterio che l'ordine delle unità statistiche assume una qualche rilevanza. A seconda del tipo di approccio e del tipo di disegno si ricorre ad un differente tipo di criterio di ordinamento. La serie che utilizza quello territoriale viene detta spaziale (cfr. par.fo 5.3) mentre se il criterio di ordinamento delle unità è temporale la serie prende il nome di storica e temporale (*time series*). Tecnicamente, le serie storiche e temporali sono la sequenza dei valori assunti da una variabile cardinale nello stesso aggregato territoriale in tempi differenti, mentre le serie territoriali sono la sequenza dei valori assunti da una variabile con categorie non ordinate⁶ nello stesso momento in aggregati territoriali differenti. Così, parleremo di serie storica se ci riferiamo al numero di iscritti alle scuole superiori dal 2007 al 2017 e di serie territoriale se ci riferiamo al numero di iscritti alle scuole superiori nei vari ambiti territoriali italiani

Indipendentemente dal criterio di ordinamento, è possibile però riscontrare almeno quattro caratteristiche comuni ai dati prodotti dai due tipi di serie. In entrambi i casi vengono associati i valori di una variabile ai valori della variabile *tempo*, in un caso, e della variabile *zona geografica*, nell'altro e quindi le due serie informano su quali valori assume una determinata variabile nel tempo o nello spazio (seconda caratteristica); ciò significa che, ad esempio all'anno 2017 viene associato il numero di iscritti alle scuole superiori in quell'anno (tab. 4.1). La terza caratteristica comune è che è possibile avvalersi delle operazioni e dei grafici propri delle distribuzioni di frequenza⁷, nonostante le rappresentazioni dei dati non siano effettive distribuzioni di frequenza (quarta caratteristica) in quanto non associano i valori di una variabile alle rispettive frequenze. Ciò comporta che, almeno nella fase preliminare, nell'analisi delle serie è possibile ricorrere alle stesse tecniche ossia allo *studio delle variazioni di un fenomeno* sulla base del confronto temporale o territoriale delle variabili mediante le differenze (assolute e relative) e i numeri (indici a base fissa e mobile); le differenze rilevano le variazioni effettive da un tempo/area all'altra, mentre i numeri indice valutano l'intensità di un fenomeno temporale/territoriale rispetto ad un tempo/un luogo assunti come riferimento oppure rispetto a tempi o luoghi differenti (Corbetta 1999; cfr. cap 6).

Tab. 4.1 *Serie storica e territoriale sugli iscritti alla scuola secondaria di II grado*

In Italia per anno		Nel 2107 per macroarea di residenza	
Serie storica (temporale)		Serie territoriale	
Anno	Iscritti	Macroarea	Iscritti
2015	165.775	Nord ovest	516.033
2016	168.901	Nord est	349.407
2017	169.878	Centro	471.766
Media	168185	Sud	780.036
Dev. St	2143	Isole	348.174
		Totale	2.465.416

Fonte: Nostra elaborazione su dati esemplificativi

Rinviando ai prossimi paragrafi per una argomentazione più dettagliata sulle serie territoriali e sulle tecniche di analisi (cfr. cap.li V e VI), trattiamo nello specifico le serie storiche.

⁶ La tipologia di variabili a cui si fa riferimento è costituita dalle variabili categoriali e cardinali. Nelle prime, come è noto, rientrano le variabili categoriali non ordinate e ordinate, mentre nel secondo tipo rientrano quelle cardinali e quasi cardinali. Per una trattazione specifica si veda Marradi (1992) e Caputo (2012)

⁷ E quindi, i valori di tendenza centrale e di variabilità per variabili categoriali non ordinate (per le serie territoriali) e cardinali (per le serie temporali)

Per serie storica o temporale si intende una successione di dati osservati su un determinato fenomeno Y ordinati secondo la dimensione tempo t (per $t = 1, 2, \dots, N$). Si tratta in definitiva di una successione di dati o valori osservati su un determinato fenomeno in tempi diversi. La cadenza temporale di osservazione è mutevole e dipende dagli obiettivi della ricerca, una serie può essere quindi oraria, giornaliera, settimanale, mensile, trimestrale, annuale o anche quinquennale o decennale, inoltre, la cadenza temporale può riferirsi ad una successione di istanti temporali o ad intervalli temporali equispaziati o meno. Nel primo caso si parla di *serie di stato o posizionali* (gli studenti di un istituto superiore alla fine dell'anno scolastico), nel secondo di *serie di flusso*, come ad esempio accade nello studio della produzione giornaliera di giocattoli nell'azienda X (ad es. la produzione giornaliera). Nelle serie storiche il legame tra variabile e il suo criterio di ordinamento è quindi fittizio e mutabile, ciò che è può essere ritenuta una costante è che le rilevazioni effettuate devono essere condotte sempre ad intervalli di tempo regolari.

Fig.4.2 Rappresentazione grafica di una serie storica

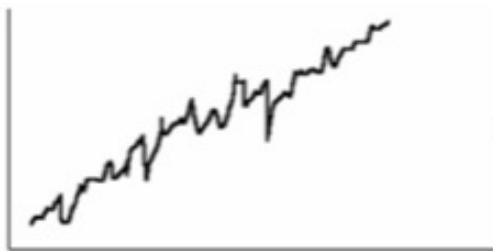
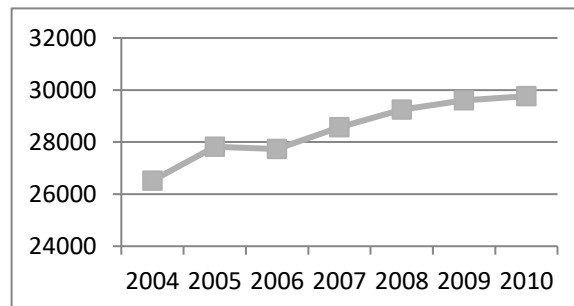


Fig.4.3 Esempio di serie storica tabellare e grafica sulle medie del reddito netto familiare italiano

Anno	Reddito familiare Netto
2004	26521
2005	27815
2006	27731
2007	28575
2008	29243
2009	29606
2010	29766



Fonte: Nostra elaborazione su dati esemplificativi

Riguardo l'analisi di serie storiche, gli obiettivi possibili sono essenzialmente tre, descrivere l'andamento del fenomeno (trend), analizzare la relazione tra il fenomeno oggetto di studio e l'impatto di determinati elementi ed, infine predire l'evoluzione (o involuzione) del fenomeno osservato. Ci si chiede, in definitiva, il numero di iscritti all'università è aumentato negli ultimi 10 anni? L'ipotetico aumento è attribuibile alle riforme universitarie intervenute in questi anni? Ed, infine, sulla base di quanto evinto, in futuro, il numero degli iscritti continuerà ad aumentare?

Generalmente, osservando le tendenze si rileva che la maggior parte delle serie storiche presenta oscillazioni intorno ad un andamento di lungo periodo definite componenti elementari virtuali della serie. Nello specifico, si distinguono quattro tipi di componenti (Henry, 1995). La prima componente è definita *trend (T) o movimento tendenziale monotono di fondo*, è un andamento di lungo periodo che indica una evoluzione strutturale del fenomeno attribuibile a cause che agiscono in modo sistematico sul fenomeno stesso; il trend può essere crescente o decrescente o, molto raramente, stabile nel tempo, in questi casi viene definito stazionario.

Fig. 4.4 Tendenza di lungo periodo



La seconda componente della serie è detta *stagionale* (S) indica le variazioni che intervengono a intervalli regolari nella serie dei dati. Le variazioni stagionali sono quindi sistematiche e riguardano la periodicità della serie, un esempio in tal senso è dato dall'andamento delle vendite di prodotti da spiaggia che registra picchi nei mesi estivi e una forte contrazione negli altri mesi dell'anno oppure al tasso di disoccupazione. E' implicito che il termine stagionale non fa riferimento alle stagioni, piuttosto indica una variazione della serie legata ad una periodicità ossia alla ripartizione temporale considerata; se ad esempio si studiano le variazioni settimanali della frequenza in chiesa dei giovani allora le stagionalità saranno sette (giorni), se si rileva il tasso di disoccupazione ogni quattro mesi, allora le stagionalità saranno quattro

Fig. 4.5 Stagionalità



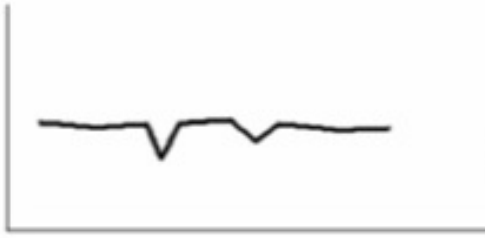
La componente *ciclica o di movimento o di oscillazione congiunturale* (C) individua le oscillazioni di lunghezza ed intensità variabili e di durata più ampia rispetto alle variazioni stagionali; il ciclo ha generalmente effetti sul medio e lungo periodo e le oscillazioni sono legate a circostanze contingenti. Un esempio è offerto dalle oscillazioni cicliche economiche caratterizzate dall'alternanza di fasi di boom economico contrapposte a fasi di recessione.

Fig. 4.6 Ciclo



Infine, la componente *stocastica o accidentale o di disturbo* (e) si riferisce ai movimenti irregolari, erratici o casuali provocati in determinate di circostanze che provocano cambiamenti improvvisi nella serie di dati; specie in riferimento allo studio dei fenomeni sociali, è difficilmente riscontrabile una relazione perfetta tra il fenomeno osservato e le diverse componenti, il fattore accidentale tiene conto della imprevedibilità dei comportamenti

Fig. 4.7 Accidentale



Una ipotetica relazione per una serie storica può essere sintetizzata nel seguente modo

$$Y_t = f(T_t, C_t, S_t, \epsilon_t) \text{ dove } t=1, \dots, N$$

Nei modelli di scomposizione classica il trend ed il ciclo formano una sola componente, detta componente sistematica (o anche trend-ciclo)⁸. La stagionalità, insieme agli errori accidentali, forma invece una componente non sistematica dalla quale la serie deve essere depurata. L'obiettivo dell'analisi delle serie storiche è principalmente quello di analizzare il trend, pertanto diventa necessario eliminare dalla serie le fluttuazioni stagionali mediante il calcolo, per ogni elemento della serie, delle medie, dette medie mobili (cfr. paragrafo 6). Analizzare il trend significa anche indagare sui possibili effetti che eventi esterni possono avere sul fenomeno rilevato; ad esempio, il numero delle assunzioni dei giovani nelle aziende private italiane è aumentato dopo l'introduzione dell'ultima riforma del lavoro "Job Act"? Ma non solo.

Gli aspetti di stagionalità e ciclo sono sempre presenti ma, al contrario del trend che è più stabile, sono allo stesso tempo transitori. Una serie può cambiare trend, può appiattirsi o impennarsi, tuttavia per essere certi che dell'effettivo cambiamento è necessario che sia trascorso un tempo sufficientemente ampio in modo da esser certi che il cambiamento rilevato non sia invece attribuibile, ad esempio, da un movimento ciclico particolarmente violento.

La maggiore stabilità del trend consente al ricercatore di fare delle previsioni circa i valori che un fenomeno può assumere nel futuro (Ruspini, 2004). Qualsiasi studio di una serie temporale parte, quindi, dall'individuazione del trend; per realizzare una previsione attraverso questo tipo di dati diventa necessario ottenere informazioni sul comportamento evolutivo che in genere esso presenta: ipotizzando di voler studiare il tasso di occupazione in Italia, la prima cosa da fare è capire come esso si comporta nel tempo, cioè se il suo andamento è costante, crescente o decrescente. Dopo un'approfondita analisi della tendenza di fondo di un dato fenomeno e dei problemi di previsione da affrontare, la decisione successiva riguarda la possibilità di raccogliere direttamente i dati oppure di utilizzare quelli già disponibili dalle varie fonti dati.

4.3.1 La qualità del dato negli studi longitudinali

I dati rilevati mediante studi longitudinali, come si è visto, combinano le informazioni relative alle caratteristiche di N individui nello stesso istante temporale (dati *cross section* o cross-sezionali, ovvero dati in formato trasversale) con quelle rilevate per gli stessi individui in T diversi periodi di tempo (dati *time series* o serie storiche o temporali). Ma, ciascun tipo di disegno longitudinale da vita ad un differente tipo di dato ciascuno dei quali presenta livelli di 'qualità' differenti.

I dati trend. Rispetto agli studi diacronici, le indagini trend presentano certamente numerosi vantaggi in termini di tempo e denaro, ma anche, come si è visto, numerosi svantaggi. D'altro canto, però, il fatto che gli stessi individui non siano intervistati nel corso del tempo è più difficile rilevare con precisione i meccanismi che governano il mutamento sociale; in altri termini, gli studi dei trend, basandosi su campioni differenti, si limitano allo studio del livello macro di un fenomeno.

I dati prospettici. I dati longitudinali prospettici non sono esenti da limiti. Innanzitutto, un qualsiasi ricercatore che segue negli anni l'andamento nel tempo di uno specifico gruppo di soggetti si accorgerà del rischio che uno o più individui possano decidere di uscire dall'indagine per varie ragioni già dopo la prima ondata di rilevazione comportando un ridimensionamento del campione, fenomeno conosciuto come *attrito*. Le conseguenze dell'attrito non solo sono circostanziali - vanno ad influire sulla rilevazione all'istante in cui

⁸ E' utile ricordare che esistono due approcci all'analisi delle serie storiche (vedi i testi citati), l'approccio classico o tradizionale e quello moderno. Il primo assume che quanto mostrato dalla serie sia costituito da una parte deterministica che consente di stimare le singole componenti e una parte di disturbo casuale; L'approccio cosiddetto moderno assume, invece che la serie sia stata generata da un processo stocastico a componenti correlate descrivibile con appositi modelli probabilistici.

avviene - ma anche cumulative, nel senso che si ripercuotono per tutta la durata dello studio. Tale problema comporta l'esigenza di trattare le risposte mancanti in un'ottica diacronica in quanto l'abbandono di alcuni soggetti può comportare distorsioni nel campione con il rischio che la sua forma finale presenti caratteristiche molto distanti dal campione originale. E non solo. Distorsioni del campione potrebbero compromettere le inferenze costruite sulla base delle informazioni fornite dal campione stesso: ad esempio, se in un'indagine panel sui consumi televisivi al tempo t_1 partecipa un numero di soggetti minore rispetto al tempo iniziale t , è possibile che il mutamento registrato possa essere l'effetto, in parte o in tutto, di questo tipo di distorsione.

Un'altra questione, non meno importante, è legata all'influenza dello strumento sui soggetti; la natura delle risposte è in qualche modo influenzata dalla partecipazione a più riprese al panel. Può accadere che tra una prima ondata e la successiva alcuni soggetti possono sentirsi più a proprio agio a farsi intervistare oppure prendono più confidenza con la situazione di ricerca, o, ancora, semplicemente perché possono aver acquisito più informazioni o accumulato più esperienza sul fenomeno indagato. Conosciuto con il termine *panel conditioning* (Bailar, 1989), questo fenomeno implica anche che i panel, in virtù del proprio ripetersi, tendono a condizionare il fenomeno studiato.

Gli studi diacronici prospettici generalmente fanno riferimento ad un tempo limitato e danno indicazioni su determinati punti nel tempo, mentre sono poco informativi circa l'evoluzione degli eventi tra i due punti temporali presi come riferimento. In generale, quindi, in un dato istante temporale coincidente con una ondata di rilevazione possono verificarsi diversi fenomeni in grado di incidere sulle risposte dei partecipanti, distorsione conosciuta come *fallacy of historical period*. È difficile, dunque, fissare il cambiamento soprattutto perché le unità su cui si fonda la rilevazione sono dinamiche e mutevoli, i soggetti possono cambiare il proprio status su una proprietà nel tempo ma, allo stesso tempo, i cambiamenti sono anche reversibili, ad esempio, un soggetto può acquisire maggiore esperienza, avanzare nella carriera lavorativa, ma nel corso negli intervalli di tempo che intercorre la rilevazione può passare dallo status di occupato a disoccupato oppure fare e disfare relazioni sociali. Tali studi sono più affidabili nel controllo degli effetti specifici età, periodo e coorte (Rose, 1993), tuttavia, anche in questi casi, si può incorrere nell'errore, definito come *fallacy of cohort centrism*, che si commette quando si vuole inferire dai dati su specifici percorsi di vita di individui, appartenenti ad un determinato periodo storico, sull'andamento dei corsi di vita in generale.

I dati retrospettivi Il disegno di ricerca che è alla base dei dati prospettici prevede che la rilevazione avvenga una sola volta, ciò comporta innumerevoli vantaggi rispetto ai panel in termini di costi sia economici sia di tempo. Il fatto però che la componente longitudinale risieda nella ricostruzione *ex post* del fenomeno oggetto di osservazione e si basi, quindi esclusivamente sulla memoria degli intervistati porta con sé significativi svantaggi. Agli inizi degli anni '90 del secolo scorso, furono individuate sei cause che possono influenzare la capacità di ricordare (Eisenhower, 1991; Berington 1995), oltre alla desiderabilità sociale, *l'effetto decadimento* (tempo trascorso dall'evento), *la salienza* (importanza dell'evento), *la difficoltà dell'intervista* (ammontare delle informazioni richieste), *conflitti tra informazioni* (incapacità di distinguere tra eventi simili) ed, infine, lo *stato psicologico* del soggetto. In generale, quanto più l'evento da ricordare è lontano nel tempo, tanto più quanto viene raccontato potrebbe non corrispondere al dato effettivo, in questo caso l'effetto decadimento è accompagnato da un *forward telescoping* - il soggetto potrebbe collocare l'evento passato in un tempo più recente - o da un *backward telescoping*, l'evento, al contrario è collocato in un tempo più lontano da quello effettivo. A questi problemi va aggiunto poi che alcune informazioni sono di per sé complesse da ricordare a distanza di tempo, si pensi a questioni che riguardano atteggiamenti passati, stati affettivi o cognitivi o, semplicemente, situazioni specifiche come la spesa mensile per beni di seconda necessità di cinque anni prima, e così via.

Sono state individuate strategie per limitare le distorsioni insite nelle procedure dei disegni longitudinali prospettici. A titolo esemplificativo, riportiamo le due più rilevanti. La *aided recall* e la *recall clues* sono i supporti che agevolano il ricordo fornendo all'intervistato, rispettivamente, immagini o indizi. Per limitare gli effetti del *forward telescoping* e del *backward telescoping* si ricorre, invece, all'*event markers*, una strategia che consiste nel «contestualizzare, temporalmente, il quesito facendo esplicito riferimento ad uno o più eventi significativi nella vita dell'intervistato» (Ruspini, 2004 94), definiti *landmark event*, in questo modo è come se nella ricostruzione si creasse un 'ancoraggio', uno 'spartiacque' tra ciò che era prima e ciò che è stato dopo. Se si sta indagando sul lavoro giovanile, un *event markers* potrebbe essere costituito, ad esempio, dal conseguimento della laurea.

Come si è avuto modo di mostrare in queste pagine, gli studi longitudinali ricoprono un ruolo rilevante nell'ambito della ricerca sociale, soprattutto quando si intende rilevare il cambiamento dei fenomeni. Oltre

agli evidenti vantaggi in termini di obiettivi conoscitivi, a parere di chi scrive, la ricerca longitudinale rappresenta lo spazio interdisciplinare, uno spazio in cui interagiscono, la sociologia, la demografia, la statistica. E' però sempre opportuno tener presente quelli che sono i limiti di questo tipo di ricerca che, però possono essere in gran parte superati individuando, con sapienza, il disegno più adatto agli obiettivi che guidano la ricerca.

