

*Prima parte*  
*Tecniche d'indagine*

## **INTRODUZIONE**

La statistica aziendale è una disciplina che si occupa di temi diversi da quelli che sono solitamente trattati negli altri insegnamenti di statistica. Si occupa di metodi di analisi di dati interni all'azienda, metodi di statistica industriale e di rilevazione di informazioni esterne all'azienda.

Nell'analisi di dati interni all'azienda si raccolgono e si elaborano dati relativi ad attività svolte all'interno dell'azienda, ad esempio un'analisi di questo tipo può avere ad oggetto i dati di bilancio; nel secondo caso invece oggetto dell'indagine è il controllo statistico della qualità del prodotto, quindi del processo produttivo. Nell'ultimo caso lo statistico ha il compito di acquisire dati difficilmente reperibili, in quanto esterni all'azienda, pertanto si rivela opportuno progettare e organizzare vere e proprie *indagini statistiche*.

Nella progettazione di un'indagine statistica è necessario innanzitutto definire quali sono gli obiettivi conoscitivi dell'indagine (cosa si vuole conoscere attraverso l'indagine), per poi passare alla scelta delle tecniche e degli strumenti che si vogliono utilizzare. Ad esempio:

### **I FASE: Definire**

- Fenomeno d'interesse;
- unità statistica;
- Popolazione di riferimento;
- Dominio di stima;
- periodo di riferimento dell'indagine;
- Tipologia dell'indagine (censuaria, campionaria, panel, non panel);

### **II FASE: Costruzione di un archivio di unità statistiche;**

### **III FASE: Redazione del questionario;**

**IV FASE: Test del questionario** (tramite un campione pilota);

**V FASE: Somministrazione del questionario;**

**VI FASE: Registrazione dei dati;**

**VII FASE: Controllo e correzione dei dati;**

**VIII FASE: Piano di campionamento;**

**IX FASE: Produzione della stima;**

**X FASE: Stima della varianza delle stime;**

**XI FASE: Diffusione dei dati;**

*Osservazione:* la prima fase è quella che richiede maggiore attenzione, perché da questa fase dipendono la qualità e la coerenza delle fasi successive. Mentre la fase più lunga è la fase di controllo e correzione dati (questo spiega perché i censimenti sono pubblicati con molto ritardo rispetto alla somministrazione dei questionari). La fase più costosa è la costruzione dell'archivio; ad esempio è costosissimo costruire un archivio delle imprese della regione Abruzzo; tuttavia in un archivio possiamo trovare tutte le informazioni di cui abbiamo bisogno.

## 1. Obiettivi e definizioni

Il momento della definizione degli obiettivi è uno dei più critici per il successo di un'indagine statistica. È fondamentale avere le idee molto chiare sugli obiettivi della ricerca e interrogarsi a fondo per identificare nei minimi dettagli i traguardi conoscitivi cui si vuole giungere; infatti, modificare costantemente gli obiettivi nel corso della ricerca significa solitamente rendere poco efficace il lavoro dei ricercatori, quindi l'individuazione degli obiettivi deve essere oculata e precisa poiché ogni obiettivo si traduce in un costo (in fase di intervista, in fase di elaborazione, etc.).

Dobbiamo individuare esattamente il fenomeno che vogliamo studiare, definire l'obiettivo e le unità statistiche riferite al fenomeno d'interesse, unità che possono essere differenti ma comunque collegate all'oggetto della nostra indagine.

Per popolazione di riferimento s'intende l'insieme delle unità statistiche rilevanti per l'indagine statistica e che sono portatrici dei caratteri che ci interessano.

Per unità statistica s'intende ogni ente (elemento, individuo...) portatore di un carattere che può avere rilevanza per il raggiungimento del nostro obiettivo. Una delimitazione della popolazione di riferimento è quella rappresentata dal tempo (istante o periodo) e dallo spazio.

Per quanto riguarda il periodo di riferimento dell'indagine è necessario prendere due decisioni di carattere temporale: il periodo relativo alla definizione della popolazione e per quanto tempo portare avanti l'indagine.

Per migliorare la qualità delle nostre stime potrebbe essere necessario definire in questa fase i cosiddetti domini di stima, cioè gli aggregati all'interno dei quali effettuare una stima che sia coerente con gli obiettivi e le definizioni precedenti, che saranno compresi meglio dopo aver trattato "la stratificazione", giacché esiste una corrispondenza di solito biunivoca tra domini di stima e variabili di stratificazione. Per quanto riguarda la tipologia dell'indagine si rimanda ai capitoli successivi.

Altre due considerazioni sono molto importanti: a) la definizione di unità statistica può cambiare secondo la fase della rilevazione, concetto che sarà chiarito successivamente; b) le fasi di un'indagine statistica non sono separate e indipendenti, ma in stretta relazione. In ogni fase dovremo tener conto delle fasi precedenti e di quelle successive,

concetto che si potrà comprendere meglio quando si avranno a disposizione tutti gli strumenti per realizzare un'indagine.

## **1.1 L'unità statistica**

Come prima operazione dobbiamo definire l'unità statistica. Possiamo utilizzare vari criteri, tra cui la capacità informativa e i costi per reperire le informazioni; prenderemo come unità statistiche le famiglie, le imprese o gli individui, che sono quelle più trattate. Se vogliamo conoscere ad esempio la produzione industriale della regione Abruzzo l'impresa è un'unità statistica accettabile? Assolutamente no, perché prenderemmo solo le imprese residenti e quelle non residenti saranno eliminate dal nostro archivio perché non siamo in grado di rilevare le caratteristiche regionali. Una soluzione più idonea potrebbe essere utilizzare come unità statistica il singolo stabilimento ovvero le unità locali, naturalmente come unità di campionamento di secondo stadio se si tratta di un'indagine campionaria, mentre come unità di campionamento di primo stadio prenderemo la ripartizione del territorio.

Secondo le variabili che dobbiamo indagare alcune informazioni potrebbero esserci fornite dallo stabilimento e altre dall'impresa. Uno stabilimento della Fiat, ad esempio, certamente non conosce alcune variabili che sono di stretta competenza dell'impresa, come il costo del lavoro, mentre ce ne sono altre che possono essere più di competenza dello stabilimento.

Se vogliamo sapere ad esempio quante olive sono prodotte in Italia, potremmo utilizzare l'azienda agricola e chiedere a ciascuna quante olive produce. E' una soluzione non molto indicata, come quella di chiedere alle associazioni di coltivatori di olive, poiché qualcuno potrebbe non essere iscritto. Forse la soluzione migliore potrebbe essere quella di dividere l'Italia in aree territoriali, (quadrati, rettangoli, etc...), e vedere in ognuna quante olive sono prodotte, tramite il reperimento delle informazioni presso i frantoi.

Ma qual è il metodo migliore? Utilizziamo fonti amministrative oppure utilizziamo delle partizioni territoriali per costruire il nostro archivio?

Vediamole entrambe: supponiamo di occuparci di ricerche di mercato per la Parmalat e vogliamo sapere quanto latte è munto in Italia, quale unità statistica utilizziamo? La prima possibilità che ci si prospetta, che è anche la più intuitiva, è quella di chiedere all'azienda agricola; la seconda possibilità invece è la più intelligente, ovvero chiedere

alle aziende che trasformano il latte; commentando queste due possibilità sappiamo che in Italia le imprese agricole che producono latte sono circa 4 milioni, si capisce quindi l'inefficienza nel definire un'unità statistica così numerosa e soprattutto costosa, al contrario la seconda possibilità risulta molto più efficiente in quanto a livello italiano abbiamo solo 2000 unità locali che fanno trasformazione di latte, inoltre queste non hanno alcun motivo di mentire in merito alle variabili che c'interessano, gli agricoltori invece potrebbero mentire, ad esempio, in merito alle restrizioni riguardanti le quote latte.

In generale per quanto riguarda la capacità informativa delle unità statistiche utilizziamo la famiglia per lo studio di variabili economiche, l'individuo per variabili psicologiche o sociali, mentre utilizziamo le unità locali per variabili economiche; naturalmente questa non è una scelta obbligata in quanto possiamo chiedere qualsiasi variabile a qualsiasi tipologia di unità statistica, ma certamente la convenienza sia nei costi che nella capacità informativa non è la stessa per tutte le unità statistiche. Il criterio principale allora è la "convenienza".

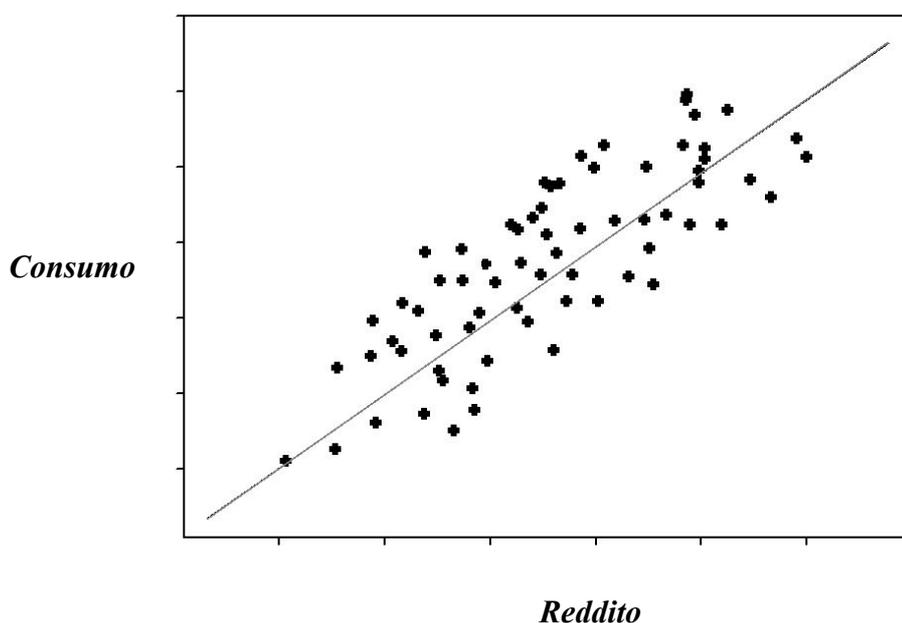
L'unità statistica "partizione del territorio" la utilizziamo sia per variabili ambientali e territoriali che per indagini che riguardano variabili relative a famiglie e imprese. Ad esempio, sul consumo dei cioccolatini, nessuno ci vieta di partizionare il territorio al primo stadio, prendere un campione di parti, e poi per ciascuna parte (ad esempio le partizioni del territorio cittadino), suonare in ogni palazzo e chiedere a tutti ciò che ci interessa. Tutti lo possono fare, costa poco e non si sbaglia mai, come nel caso delle imprese, perché una di queste potrebbe non essere presente nella fonte che abbiamo usato per costruire l'archivio, invece un pezzo di terra possiamo dividerlo quanto ci pare e non sbagliamo mai. Il problema è che il numero di unità territoriali o è infinitamente grande, oppure le famiglie o le unità statistiche scelte risultano essere molto omogenee all'interno di una stessa partizione, quindi sarebbe meglio condurre l'indagine su un campione sparso, in cui le unità statistiche sono eterogenee tra loro.

La definizione di unità statistica può cambiare secondo la fase dell'indagine; infatti, possiamo avere unità campionarie, unità di analisi, unità di campionamento, unità d'informazione, unità di rilevazione, unità primaria, unità secondaria, etc.

Ad esempio per quanto riguarda i redditi l'unità statistica di riferimento è l'individuo, mentre per i consumi è la famiglia; in macroeconomia, dalla teoria di Keynes sul consumo  $C = a + b \cdot Y$ , sappiamo che per stimare il parametro  $b$  che rappresenta la

cosiddetta propensione marginale al consumo utilizziamo un modello di regressione lineare.

Ora, poiché per indagare i consumi utilizzo la famiglia come unità statistica, mentre per indagare i redditi utilizzo l'individuo, teoricamente non potrei effettuare un'analisi di regressione, poiché utilizzo due unità statistiche basate su unità di misura diverse; per risolvere questo problema nella nostra analisi dobbiamo per forza in una fase, considerare le due unità statistiche allo stesso modo, in pratica l'unità di analisi diventa la famiglia anche per il reddito altrimenti non potrei correlare le due variabili.



*Fig. 1.1 – Regressione consumo-reddito*

## **1.2 Glossario delle unità Statistiche**

### **Unità amministrativa**

E' l'unità statistica corrispondente a un ente delle pubbliche amministrazioni. Sono di questo tipo le unità amministrative che hanno un riferimento territoriale preciso, come le regioni, le province, i comuni, le USL, e quelle non territoriali, come gli ospedali, le scuole superiori e le università, etc.

### **Unità campionaria**

E' l'unità che compone un campione. Convieni distinguere questa denominazione da quella di "unità di campionamento", con la quale s'intende una delle unità che compongono un aggregato che deve essere sottoposto a campionamento. In un campionamento su più stadi, le unità campionarie si dicono "primarie" se compongono l'insieme di unità di primo stadio, "secondarie" se compongono il campione selezionato al secondo stadio, "finali" se compongono l'insieme delle unità dell'ultimo stadio.

### ***Unità complessa***

Sono le unità che derivano dalla combinazione di due o più unità elementari. Per esempio, il reddito nazionale annuo pro-capite, il numero di stanze delle abitazioni occupate per residente, sono variabili relative a unità complesse.

### ***Unità designata***

In un'indagine diretta, è l'unità campionaria sulla quale s'intende svolgere la rilevazione dei dati. Per esempio, in un'indagine sulle famiglie, l'unità designata a rispondere per conto della famiglia, può essere il capofamiglia o un adulto responsabile, quella designata a rispondere per i membri può essere il singolo membro o un familiare. Per i minori, è certamente un adulto responsabile. Talvolta l'unità designata non rispondente è sostituita da un'unità collaborativa. Ricorrendo al termine inglese, taluni denominano proxy questo rispondente.

### ***Unità di analisi***

L'unità appartenente all'insieme che è analizzato ai fini dell'analisi statistica. Se i dati sono forniti sotto forma di tabelle, si denomina unità di tabulazione. Può non coincidere con l'unità di rilevazione.

### ***Unità di classificazione***

L'unità di cui s'identifica la modalità di una data variabile. Se le variabili di classificazione sono due o più, dell'unità di classificazione si identificano le modalità di ogni singola variabile.

### ***Unità di campionamento***

E' l'unità che compone un aggregato da sottoporre a campionamento. Le unità di campionamento sono individuali e indivisibili nel momento in cui si effettua la selezione campionaria. Le unità da estrarre possono essere definite su basi naturali (persone, famiglie di fatto, animali etc.), giuridiche (famiglie anagrafiche, imprese etc.), amministrative (comuni, ospedali etc.), territoriali (aree in cui è ripartibile un territorio) o su qualsiasi altra base. Nel campionamento a più stadi, le unità di campionamento sono diverse ad ogni stadio: sono più grandi a livello superiore e gradualmente più piccole mano a mano che si scende di livello. L'unità di campionamento si distingue dall'"unità campionaria" con la quale si intende una delle unità che compongono un campione.

### ***Unità d'informazione***

E' quella che in un'indagine o in un esperimento statistico fornisce le informazioni. Può non coincidere con l'unità per la quale si rilevano le informazioni (unità di rilevazione) o con quella alla base dell'analisi dei dati (unità di analisi). Nel censimento dell'agricoltura l'azienda agricola è una delle unità di rilevazione; si possono analizzare i dati sia per azienda sia per appezzamento e l'unità di informazione è il conduttore dell'azienda.

### ***Unità di rilevazione***

Unità empirica su cui si basa la rilevazione. Non coincide necessariamente né con l'unità che fornisce le informazioni (unità d'informazione), né con l'unità statistica cui si è interessati. Nel censimento della popolazione, la famiglia di censimento è una delle unità di rilevazione; se la famiglia è composta di un solo membro, l'unità di rilevazione coincide con l'unità statistica di analisi, altrimenti si hanno tante unità di rilevazione quanti sono i membri della famiglia, in corrispondenza di una sola unità di informazione (il capofamiglia).

### ***Unità di tabulazione***

Unità di cui si presenta una classificazione secondo le modalità di carattere in una tabella. Il nucleo familiare in un censimento demografico non è l'unità statistica (il censito), né una di quelle di rilevazione, ma una derivazione da una di queste (la

famiglia di censimento) attraverso il carattere "relazione col capofamiglia", e costituisce oggetto di classificazione e pubblicazione in tavole specifiche.

### ***Unità elementare***

La più piccola unità che fornisce informazioni. Le unità elementari opportunamente aggregate permettono di stimare i parametri della popolazione oggetto di studio. Le unità elementari devono essere indicate con precisione come unità d'informazione e rispetto alle elaborazioni effettuabili (unità di analisi).

### ***Unità finale (di campionamento)***

Nel campionamento a più stadi, è un'unità selezionata all'ultimo stadio. Non necessariamente queste unità coincidono con le unità di rilevazione. Per esempio nella rilevazione trimestrale sulla forza lavoro sono unità finali di campionamento le famiglie, ma sono unità di rilevazione i membri delle famiglie stesse. Una volta individuata una famiglia campione, infatti, tutti i membri della stessa sono inclusi nel campione.

### ***Unità primaria***

Nel campionamento a più stadi è un'unità selezionata al primo stadio. Per esempio, in un campionamento areale per un'indagine sulle famiglie, sono primarie le unità territoriali selezionate al primo stadio (può trattarsi di comuni, di isolati cittadini, etc.). Il termine è talvolta usato per denotare unità statistiche di record di base che non derivano da alcun calcolo, per esempio, persone, tonnellate etc...

### ***Unità secondaria***

Nel campionamento a più stadi, è l'unità selezionata al secondo stadio. Se gli stadi di campionamento sono due, le unità secondarie si dicono anche finali, e possono o no coincidere con le unità di rilevazione.

### ***Unità sovrapposte***

Unità della popolazione dalla quale si traggono campioni che possono non essere mutuamente esclusivi. Se per qualsiasi motivo (mappe, unione di liste diverse etc.), le unità sono del tutto o in parte sovrapposte, il campionamento può essere comunque svolto, ma vanno applicate metodiche di stima particolari per ottenere stime corrette.

### *Unità sperimentale*

Unità sulla quale si effettua un esperimento statistico.

### *Unità statistica*

Unità elementare della popolazione statistica. Può trattarsi di una persona fisica (il censito, il dimesso da un istituto di cura), di una persona giuridica (l'impresa), di un'istituzione (un istituto scolastico), di un evento (un matrimonio, una nevicata) etc.

Si possono individuare vari tipi di unità statistiche: si hanno unità campionarie e unità della popolazione in relazione alla completezza della rilevazione. Mentre in relazione alla fase dell'indagine si hanno unità di campionamento, unità di rilevazione, unità d'informazione, unità di analisi, unità di tabulazione etc...

## **2. L'archivio**

La seconda fase consiste nella costruzione dell'archivio statistico. Come si costruisce un archivio d'impresa? Che fonte di dati utilizziamo? A chi chiediamo? Potremmo chiedere alla camera di commercio, però questa fonte d'archivio non è perfetta, in quanto alcune imprese potrebbero non essere registrate oppure potrebbero esserci altri tipi di problemi. Nella costruzione del nostro archivio dobbiamo sempre tener presente che esiste una popolazione vera ed il nostro archivio.

Il tema dell'integrazione degli archivi è diventato rilevante con la creazione del mercato unico dell'Unione Europea. La maggiore libertà e concorrenzialità derivanti dalla sua creazione hanno, infatti, posto l'esigenza di ottenere delle informazioni statistiche ed economiche più aggiornate e maggiormente confrontabili.

Nasce qui la necessità di disporre di sistemi informativi costruiti sulla base di registri ed archivi attendibili, da cui poter attingere le informazioni elementari. I dati raccolti da vari enti con specifiche finalità amministrative, possono essere utilizzati attraverso apposite procedure, per la costruzione di questi registri.

Un'ulteriore sollecitazione all'utilizzo dei dati amministrativi per scopi statistici è data dalle difficoltà che devono essere affrontate dalla statistica ufficiale nel costruire e mettere a disposizione degli utenti le informazioni di cui essi hanno bisogno.

Difficoltà legate alla naturale ritrosia dei cittadini a rispondere alle indagini, all'accesso sempre più difficile ad informazioni utili alla statistica ma sempre più protette dalle leggi sulla segretezza dei dati. Per non parlare della reticenza delle imprese di fronte a una massa crescente di questionari da riempire, o delle restrizioni di bilancio imposte dai governi. Nei registri amministrativi sono contenute informazioni utili allo statistico.

## **2.1 Un po' di storia**

Poter incrociare gli schedari amministrativi diventa perciò un'opportunità da sfruttare al meglio. Esistono ormai in Italia e all'estero numerosi esempi di archivi integrati cui fare riferimento. La Danimarca già nel 1980 rinunciò ad organizzare il censimento della popolazione preferendo utilizzare come fonte i ventuno archivi amministrativi contenenti informazioni sulle persone, che un codice identificatore individuale permetteva di integrare tra loro. L'obiettivo restava il censimento della popolazione, ma poi, grazie all'accuratezza delle informazioni ottenute, si riuscirono ad utilizzare i dati per rispondere ad ulteriori esigenze che esulavano dall'obiettivo primario.

In Italia, l'Istat ha realizzato il Censimento delle attività industriali e dei servizi privati destinabili alla vendita, denominato "intermedio" poiché, per la prima volta, non ha rispettato la cadenza decennale che, nel dopoguerra, ha visto la realizzazione di cinque censimenti dal 1951 al 1991. Le modalità di rilevazione dei dati sono state assolutamente innovative.

Il Censimento intermedio prevedeva, infatti, due fasi: la prima riguardava la costruzione dell'universo delle imprese; la seconda consisteva nell'approfondimento di particolari aspetti strutturali del sistema produttivo.

Per la realizzazione della prima fase del Censimento intermedio, l'Istat ha utilizzato la base informativa dell'Archivio Statistico delle Imprese Attive (*ASIA*), costruito mediante l'utilizzo dei dati forniti dalle imprese per fini amministrativi. Esso prevede la raccolta e l'aggiornamento dei dati identificativi e di stratificazione delle imprese (e delle unità locali ad esse collegate) che possono essere desunti da fonti fiscali, giuridiche o amministrative.

Un'altra importante esperienza nazionale è stata realizzata dall'Unioncamere che attraverso l'integrazione dei quattro archivi amministrativi per eccellenza (Registro Imprese, archivio Inps, archivio Inail e anagrafe Iva), ha costruito l'archivio Excelsior.

Gli obiettivi del progetto consistono nel misurare la domanda effettiva e potenziale dei diversi bacini territoriali del lavoro, in modo da fornire un supporto informativo a quanti – enti pubblici o privati – si trovano impegnati nel guidare l'offerta di lavoro verso le esigenze espresse dalla domanda. Un altro importante obiettivo è orientare le scelte degli enti istituzionali in materia di politiche della formazione scolastica e professionale, offrendo informazioni dettagliate sui bisogni di professionalità espressi dalle imprese per il breve e medio periodo.

Per analizzare le differenze esistenti tra l'archivio integrato e le fonti amministrative, è necessario premettere che l'archivio statistico integrato non coincide con nessuno degli archivi che ne costituiscono la fonte. Le singole informazioni contenute nell'archivio statistico non sono direttamente riconducibili alle fonti utilizzate, in quanto costruite con procedure statistiche. L'archivio statistico si differenzia da quelli giuridici e amministrativi perché le informazioni in esso contenute sono finalizzate all'analisi economica e non producono effetti giuridici. Inoltre hanno scopi statistici e non certificativi e il loro aggiornamento avviene periodicamente e non sempre.

Oggetto di indagine dell'archivio statistico sono le imprese e le relative unità locali economicamente attive. Non interessano invece le unità aventi un'esistenza meramente giuridica, come quelle create per motivi amministrativi o formali ("imprese di carta" o "società di comodo"), né le unità giuridicamente vive ma non attive, perché ancora in fase organizzativa in attesa di iniziare o perché hanno sospeso o cessato l'attività. Nei registri giuridici e nei registri amministrativi i due tipi di unità sono spesso indistinguibili. Ad esempio, sia la nuova, sia la vecchia unità giuridica dell'impresa che ha modificato l'assetto proprietario, la forma giuridica o la ragione sociale, ma non l'attività né l'organizzazione delle risorse produttive, possono risultare contemporaneamente iscritte nello stesso registro per tutto il tempo necessario all'espletamento di alcuni obblighi.

Mentre l'archivio statistico deve assicurare che i caratteri in esso registrati siano attendibili e corrispondenti alla realtà economico-produttiva cui si riferiscono, i registri giuridici, dato il loro scopo di pubblicità legale, si preoccupano più della correttezza formale che dell'attendibilità sostanziale delle dichiarazioni.

I registri amministrativi, dal canto loro, organizzati per il controllo dei pagamenti periodici, tendono a concentrare la loro attenzione sui caratteri relativi all'entità e modalità dei pagamenti stessi e a trascurare la sorveglianza su altri caratteri che possono essere irrilevanti per gli obblighi di pagamento, tuttavia importanti per l'analisi economica.

Inoltre, le informazioni individuali contenute nel registro costruito a scopi statistici possono essere statisticamente accettabili anche se incomplete, purché tale incompletezza e l'incertezza che ne deriva, siano opportunamente misurate e contenute entro limiti prestabiliti.

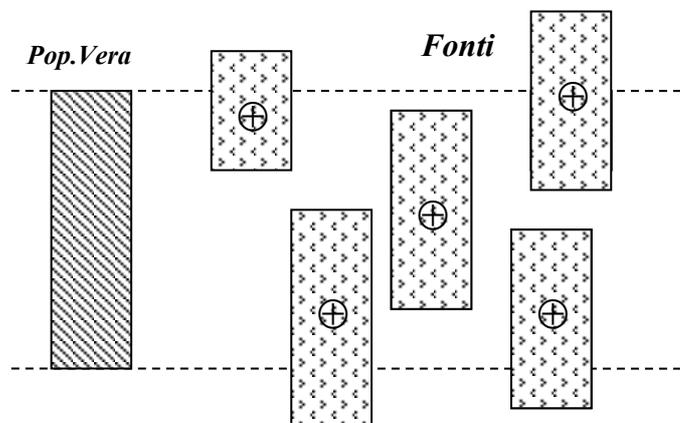
Al contrario, i dati dei registri giuridici e amministrativi che certificano posizioni giuridiche, devono essere completi e certi. Ne consegue che mentre i dati del registro statistico possono essere controllati ed eventualmente modificati con rapide procedure statistiche, quelli degli altri registri devono essere modificati con ben definite procedure amministrative, inevitabilmente lente perché stabilite nel giusto rispetto dei diritti e degli interessi dei soggetti coinvolti.

L'aggiornamento del registro statistico infine può essere effettuato con cadenza periodica (annuale, trimestrale, mensile) mentre quello dei registri amministrativi e giuridici deve essere continuativo per tenere conto tempestivamente degli eventi modificativi quali nascite, morti, cambiamenti di indirizzo, attività e altri che provocano effetti sulla posizione giuridica dei soggetti.<sup>2</sup>

## **2.2 La costruzione di un archivio ed i problemi relativi**

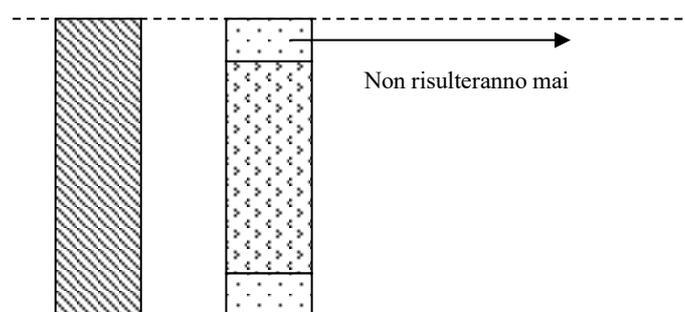
In genere l'archivio è soggetto a due tipi di errori: *l'errore di sottocopertura* e *l'errore di sovracopertura*. L'errore di sottocopertura è un errore irrisolvibile, ossia ci sono ad esempio delle aziende esistenti nella popolazione vera ma che non risultano nel nostro archivio. Mentre per quanto riguarda l'errore di sovracopertura questo indica che nel nostro archivio sono presenti più unità rispetto alla popolazione vera; in ogni caso questo è un errore risolvibile in quanto ci possiamo permettere una stima della sovracopertura mentre l'errore di sottocopertura non potrà mai risultare. Questo significa che per limitare l'errore di sottocopertura dobbiamo incrociare più fonti, nel caso delle imprese, fonti amministrative come l'Inps, la camera di commercio etc. Possiamo

quindi costruire l'archivio anche esagerando con l'inserimento dati pur rischiando l'errore di sovra copertura, che eventualmente potremmo correggere successivamente.



*Fig. 2.1 - Sovracopertura*

<sup>2</sup> Caratteristiche di un archivio statistico integrato, *Tratto da ASTRI "Archivio Statistico Regionale delle Imprese Attive" della regione Sardegna.*



*Fig. 2.2 - Sottocopertura*

### 2.3 L'archivio d'impresa

Come agganciare le varie fonti dell'archivio? Per le imprese dobbiamo avere naturalmente un codice identificativo univoco dell'unità statistica, nella fattispecie la *partita IVA*, non può che essere quest'ultima, *la ragione sociale, la forma giuridica,*

*l'indirizzo* comprensivo del codice di avviamento postale, *numero di telefono*, *e-mail*, e successivamente è fondamentale il *codice identificativo territoriale*, che alcune volte è utilizzato addirittura per realizzare delle stime, quindi come una variabile.

Questo codice è molto importante tanto da essere armonizzato a livello internazionale con la sigla *NUTS* (*NUTS1*, *NUTS2*, *ETC.*), in Italia è composto di due codici che rappresentano la regione, tre codici che rappresentano la provincia, e tre codici che rappresentano il comune. In un'indagine su scala internazionale vanno aggiunti altri tre codici per indicare la nazione.

3	2	3	3
<i>NAZ.</i>	<i>REG.</i>	<i>PR.</i>	<i>COM</i>

Ovviamente quando parliamo di indirizzo, telefono, etc., intendiamo la sede amministrativa dell'impresa. Va inserito in seguito il *settore di attività economica*, per il quale esiste una classificazione internazionale, denominata con la sigla *NACE*, quattro codici uguali in tutta l'Europa ed il quinto è concesso all'Italia per far coincidere meglio alcune attività economiche; quando il codice identificativo è formato quindi da cinque cifre rientra in un'altra classificazione denominata con la sigla *ATECO*.

Si può dire che senza questo tipo di classificazioni ogni analisi o indagine statistica diventa priva di valore. Potremmo ottenere in un'indagine migliaia e migliaia di attività diverse che sarebbe difficile classificare; comunque a livello d'impresa l'attribuzione di attività economica si riferisce a quella prevalente perché le imprese potrebbero, in molti casi, svolgere più di un'attività. Successivamente inseriamo nel nostro archivio *il numero di addetti*, variabile che è molto importante in quanto ci fornisce la dimensione dell'impresa, ad esempio non è la stessa cosa campionare un tabaccaio o analizzare una multinazionale.

Il *fatturato* è molto meno presente nella camera di commercio perché ha una correlazione pari ad uno per settore di attività economica, quindi l'aspetto dimensionale ci conviene misurarlo tramite numero di addetti.

Per la *sede* non è tanto importante la sede dell'impresa ma a quante unità locali l'impresa fa capo, soprattutto allo scopo di richiedere le stesse informazioni a tutte le unità locali dell'impresa. Inoltre attribuire un codice di attività economica ad un'unità locale è molto più logico perché la produzione, l'attività economica, è molto più

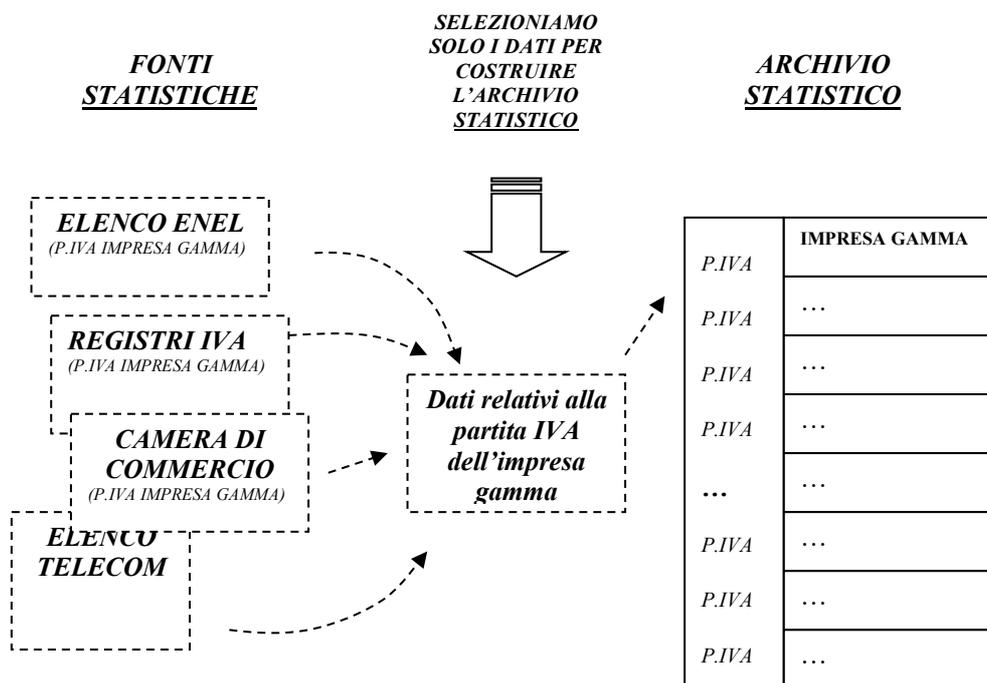
omogenea in quest'ultima rispetto all'impresa, in quanto l'impresa può far capo ad un insieme di unità locali.

Per quanto riguarda il codice territoriale, ciascuna unità locale avrà un proprio codice territoriale, dall'impresa passiamo ad un ordine gerarchico inferiore.

Queste informazioni possono essere ottenute a pagamento tramite la camera di commercio, che delega la società *CERVED* nella realizzazione di questi archivi di informazioni. Inoltre l'Istat ci mette a disposizione un archivio già riagganciato e ripulito detto *ASIA*<sup>3</sup> (Archivio statistico delle imprese attive) cui l'*ISTAT* aggiunge il censimento dell'industria che non è disponibile essendo utilizzato solo a scopi statistici, in ogni caso l'archivio della camera di commercio presenta un 93%, 94% dell'archivio Asia.

---

<sup>3</sup> (E' l'archivio delle Unità Statistiche di osservazione delle Indagini economiche dell'Istituto, creato in ottemperanza al Regolamento (CEE) n.2186/93 del Consiglio del 22 luglio 1993, relativo al "coordinamento comunitario dello sviluppo dei registri di imprese utilizzati a fini statistici" e in attuazione del decreto del Presidente del Consiglio dei Ministri del



*Fig. 2.3 – Il passaggio dalle fonti all'archivio statistico*

## 2.4 L'archivio delle famiglie e degli individui

La fonte principale cui dobbiamo attingere è l'anagrafe ovvero il suo aggiornamento tramite il censimento. L'anagrafe ha la stessa funzione a livello tecnico dell'archivio delle imprese, a livello sostanziale però è molto difficile riaggregare a livello nazionale le anagrafi comunali quindi è molto probabile che questa fonte sia piena di errori.

---

26 febbraio 1994. Raccoglie le informazioni identificative (denominazione, localizzazione), strutturali (addetti, attività economica prevalente e secondaria, natura giuridica, volume degli affari) e demografiche (data inizio attività, data di cessazione, stato di attività, presenza di procedure concorsuali) di tutte le imprese (e relative unità locali) attive in tutti i settori di attività economica (ad eccezione delle sezioni A, B, L, P e Q e dei soggetti privati non-profit) della classificazione Ateco '91. E' stato creato, ed è aggiornato annualmente, sulla base del trattamento statistico e dell'integrazione delle informazioni residenti in differenti archivi giuridici, amministrativi e di esazione. In particolare, le principali fonti utilizzate sono: il Repertorio Economico Amministrativo (REA) gestito dalle Camere di Commercio, l'archivio anagrafico dell'Inps integrato con le dichiarazioni annuali desunte dai modelli presentati dalle imprese (DM/10), l'archivio anagrafico dell'Inail, l'Anagrafe Tributaria integrata con dichiarazioni annuali IVA, gli archivi delle "utenze affari" dell'Enel e della Telecom, l'archivio anagrafico delle banche della Banca d'Italia. Per l'aggiornamento delle informazioni contenute in ASIA vengono inoltre utilizzate tutte le informazioni desumibili da tutte le indagini statistiche sulle imprese).

Tra i problemi che possiamo incontrare vi sono la mortalità e la residenza aggiornate in modo imperfetto, ed anche supponendo che la mortalità sia perfetta la residenza non potrà mai esserlo, è impossibile aggiornare giorno per giorno la registrazione in quanto questa cambia più volte, cambia molto spesso. Ma vediamo i contenuti di questo archivio, come per le imprese procediamo in modo gerarchico e poi andiamo sui singoli componenti. Le più importanti sono le informazioni riguardanti il *capofamiglia*, sono le prime informazioni che richiediamo (alla fonte qualsiasi essa sia), quindi *nome cognome*, *indirizzo* della famiglia (residenza) ma attenzione, non la residenza del capofamiglia in quanto questo potrebbe risiedere in altri luoghi; *telefono*, etc. Di solito, in ogni caso, il capofamiglia è il maggiore percettore di reddito ed anche l'unico.

Successivamente passiamo alle caratteristiche lavorative del capofamiglia ossia la *condizione professionale del capofamiglia* (dipendente, professionista, etc.). Non è banale chiedere questa informazione poiché molto spesso è una domanda filtro; *la condizione lavorativa del capofamiglia* (muratore, ingegnere, imbianchino, parrucchiere, etc.), per quest'ultima esiste una possibile classificazione standardizzata che comprende circa 7000-8000 professioni; *stato civile* (vedovo, divorziato, separato, etc.);  *sesso e età*; il *livello di avanzamento nella professione* poiché gli aspetti lavorativi che avevamo chiesto, la condizione professionale e la professione non ci dicono il grado di avanzamento; il *numero di componenti della famiglia* è l'ultima cosa che chiediamo e riproponiamo tutto quello che abbiamo chiesto al capofamiglia e soprattutto la residenza di ogni componente. Tra tutte le variabili cui abbiamo accennato quella che è più correlata con il reddito è il titolo di studio, ha il coefficiente di correlazione più alto rispetto al reddito che non altre variabili. Si può notare dalla struttura dell'archivio che non c'è alcuna domanda riguardante il reddito poiché questa richiesta viene difficilmente esaudita o comunque dà origine a mancate risposte.

## **2.5 Alcune osservazioni**

Alla luce delle nuove conoscenze, possiamo capire ora perché utilizziamo una definizione puramente territoriale di unità statistica (terra, poligoni, punti), perché partizioniamo il territorio? Fondamentalmente per tre ragioni: a) siamo obbligati dal fenomeno di analisi; b) non possiamo disporre di un archivio di unità statistiche, nessun istituto ci concede l'opportunità di visionare il proprio archivio, dobbiamo costruircelo

da soli; c) infine perché per definizione questo è l'unico archivio che non risente degli errori di sovracopertura e di sottocopertura, per definizione abbiamo messo dentro di tutto senza duplicare nulla. Ma allora verrebbe da chiedersi perché non utilizziamo sempre le partizioni del territorio. La risposta risiede nel fatto che dovremmo intervistare troppe unità, a parità di precisione bastano meno unità campionare, inoltre questo è sicuramente il metodo che ha il costo maggiore, è il metodo che ci obbliga ad intervistare più unità rispetto ad un archivio già pronto, inoltre la stragrande maggioranza delle variabili economiche presenta una grossa omogeneità territoriale. Ribadiamo che questo metodo di selezione comporta la divisione del territorio di analisi in pezzi e la successiva selezione di un campione di pezzi.

### **3. Classificazioni**

#### **3.1 Classificazione per settore di attività economica**

Parlando della costruzione degli archivi abbiamo accennato in parte alla classificazione delle unità statistiche; ci soffermiamo ora su quest'ultimo aspetto poiché la classificazione delle unità statistiche è particolarmente interessante ed utile per chi si occupa di dati economici.

Ma a cosa serve una classificazione di attività economiche? Appartenere ad un codice di attività economica piuttosto che ad un altro potrebbe significare accedere a particolari benefici, finanziamenti collegati a determinate attività, etc. Può essere più o meno vantaggioso appartenere ad un determinato codice; non è infrequente, infatti, che le imprese denuncino gli istituti che si occupano di attribuire questi codici poiché viene loro preclusa la possibilità di accedere a determinati tipi di vantaggi. Ad esempio accade spesso che il governo o altre istituzioni decidano di stanziare fondi per favorire lo sviluppo di una determinata attività economica oppure di un determinato settore che hanno un proprio codice specifico; se fossero stanziati dei fondi per il settore di attività economica che si occupa, ad esempio, della macellazione di polli, chi macella i polli ne usufruirebbe mentre ne rimarrebbe esclusa l'attività che si occupa dell'imballaggio poiché appartiene ad un diverso codice; Tra le classificazioni demo-sociali di carattere economico che potrebbero interessarci ce n'è una ma non la trattiamo, trattiamo solo classificazioni che possono avere interesse di carattere economico.

#### **3.2 La classificazione dei prodotti all'origine**

Come sappiamo la classificazione delle imprese e delle unità locali avviene per origine di attività produttiva; a questo proposito una classificazione simile è utilizzata anche per i prodotti. A questo proposito per il commercio estero il prodotto che entra o esce dai confini nazionali è importante prima di tutto per scopi fiscali e chiaramente anche per l'analisi economica.

La classificazione dei prodotti del commercio estero avviene tramite un sistema armonizzato internazionale e deve essere condivisa da tutti i paesi del mondo per cui il codice è uguale per tutti a livello internazionale. Il sistema armonizzato si traduce in una

classificazione uguale in tutto il mondo mentre in Europa assume la denominazione *NC (nomenclatura combinata)* che vale solo all'interno della comunità europea.

Una classificazione di prodotti di questo tipo è importante sia da un punto di vista fiscale e sia perché alcuni prodotti (codici) non possono passare la dogana, ossia non si possono importare.

Qualsiasi legge fa riferimento a questo tipo di nomenclatura. L'Istat mantiene ed aggiorna questa classificazione in concomitanza con il ministero, nei convegni dedicati a questo tema vi è sempre un funzionario dell'Istat ed uno del ministero. La classificazione a livello europeo viene definita come *regolamento Prodc<sub>2</sub>*, che rappresenta lo strumento normativo adottato dall'Unione europea per armonizzare l'osservazione statistica della produzione industriale negli Stati membri. L'armonizzazione della rilevazione della produzione industriale nei diversi paesi dell'Unione europea ha richiesto la definizione e l'adozione di un'unica classificazione merceologica da parte di tutti gli Stati membri.

La predisposizione della classificazione (*elenco Prodc<sub>2</sub>*) è stata curata dall'Eurostat in collaborazione con i paesi e con le rappresentanze delle associazioni industriali. La lista si riferisce principalmente ai beni materiali, ma sono compresi anche alcuni servizi industriali (riparazione, manutenzione, installazione e montaggio).

---

<sup>4</sup> L'idea di un progetto armonizzato per la rilevazione della produzione industriale nei diversi paesi dell'Unione europea nasce verso la metà degli anni ottanta. Fino ad allora, infatti, sebbene nella maggior parte dei paesi europei esistessero rilevazioni statistiche sul fenomeno, esse facevano ricorso a concetti, classificazioni e metodi spesso eterogenei fra paesi. Attualmente, a oltre 12 anni dall'entrata in vigore del regolamento, il sistema Prodc<sub>2</sub> rileva statistiche di produzione su 5.777 voci che comprendono principalmente prodotti ma anche molte lavorazioni e servizi industriali. Quindici stati membri dell'unione (più Norvegia e Islanda) producono correntemente dati di produzione in quantità e in valore. Altri 13 paesi europei (cosiddetti Phare) hanno avviato la procedura di ingresso nel sistema Prodc<sub>2</sub>. L'obiettivo fondamentale del progetto Prodc<sub>2</sub> è, pertanto, quello di creare una fonte statistica in grado di delineare il quadro strutturale ed evolutivo dei settori produttivi e dei mercati dell'industria europea, assicurando la comparabilità dei dati di produzione rilevati nei diversi paesi e la possibilità di integrare queste informazioni con quelle che scaturiscono correntemente dalle statistiche sull'interscambio con l'estero.

In generale, il regolamento prevede che l'osservazione della produzione industriale venga condotta a cadenza annuale; tuttavia, per alcune voci merceologiche, tra le quali rientra la quasi totalità dei prodotti dell'industria tessile e dell'abbigliamento e di quelli dell'industria chimica e delle fibre sintetiche e artificiali, la normativa comunitaria stabilisce che la rilevazione venga effettuata per trimestri, allo scopo di monitorare con maggiore frequenza l'andamento della produzione. L'applicazione del regolamento Prodc<sub>2</sub> in Italia ha comportato la definizione di un complesso disegno d'indagine. Oltre all'introduzione delle rilevazioni a periodicità trimestrale, i vincoli di rappresentatività previsti dal regolamento hanno reso necessario estendere l'osservazione statistica anche alle unità produttive appartenenti a classi di dimensione molto piccola. Infatti, il regolamento prevede che la rilevazione venga condotta presso imprese che rappresentino almeno il 90 per cento della produzione nazionale a livello di singola classe Nace Rev. 1.

La descrizione dei singoli prodotti è accompagnata da un codice identificativo a otto cifre concordato a livello comunitario e adottato da tutti i paesi dell'Unione europea (codice *Prodc*om). In linea generale, le prime sei cifre corrispondono alla classificazione *Cpa* (*Nomenclatura comunitaria dei prodotti per attività*), mentre le ultime due cifre stabiliscono un riferimento alla Nomenclatura combinata, utilizzata nell'ambito delle statistiche sull'interscambio di merci con l'estero.

Per effettuare la rilevazione l'Istat invia alle imprese, congiuntamente al questionario di rilevazione, un repertorio merceologico di settore. Il repertorio riporta l'elenco dei prodotti previsti dalla lista ufficiale *Prodc*om per ciascun settore industriale.

In generale, un repertorio merceologico contiene la gamma dei prodotti associati a una singola divisione di attività economica; per alcuni settori, dove risulta più frequente la fabbricazione congiunta di prodotti classificati in divisioni di attività economica diverse, il repertorio riporta i prodotti associati a tutte le divisioni interessate.

I repertori nazionali si caratterizzano, rispetto alla lista *Prodc*om, per la presenza di codici a nove cifre anziché a otto cifre. Le prime cinque cifre di ciascun codice del repertorio nazionale stabiliscono una corrispondenza con la classificazione delle attività economiche italiane (*Ateco91*), che presenta un livello di dettaglio aggiuntivo rispetto alla Ance Rev. 1 (caratterizzata da un dettaglio massimo di quattro cifre). Durante il processo di aggregazione le informazioni rilevate a nove cifre di dettaglio vengono ricondotte al livello di otto cifre *Prodc*om. I repertori merceologici nazionali sono disponibili sul sito Internet dell'Istat, all'indirizzo <http://www.istat.it/Imprese-e-/prodc>.

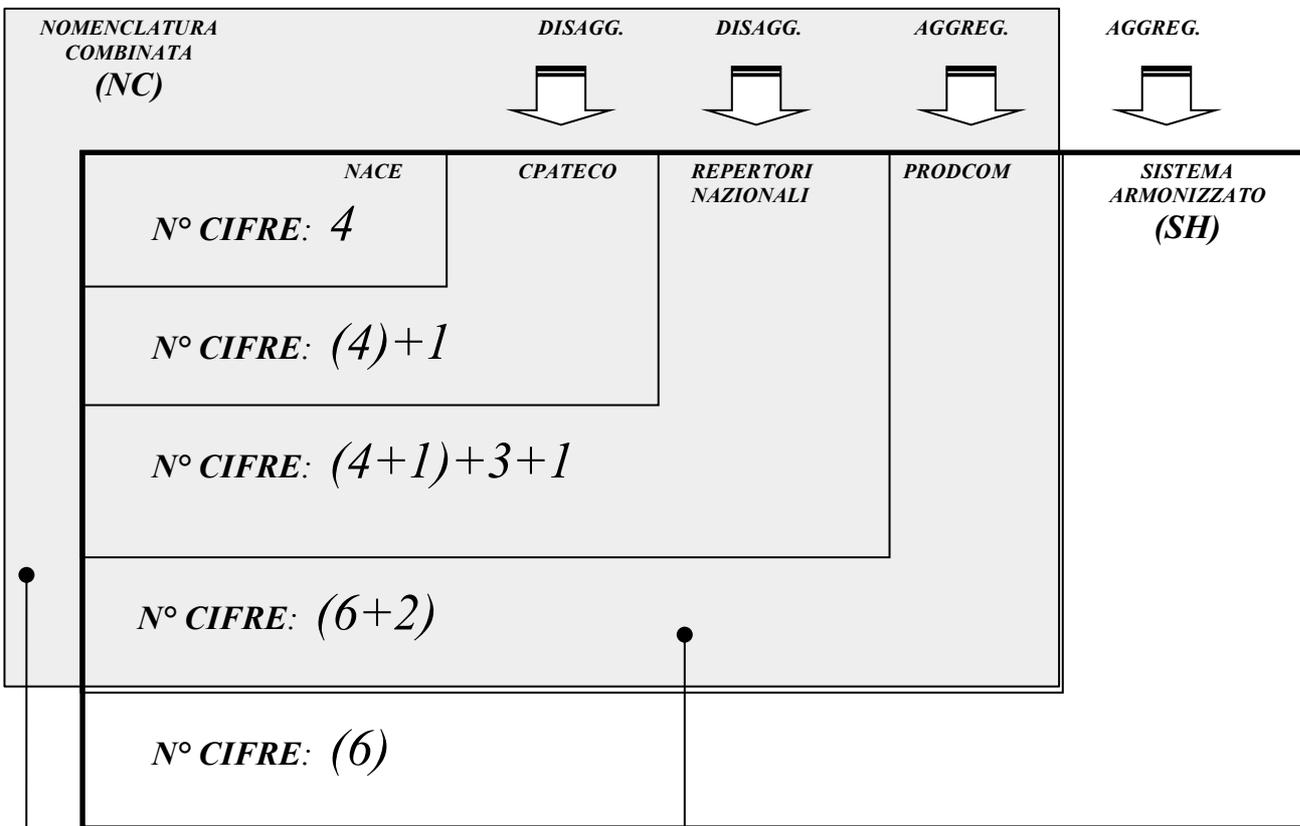
In Italia la classificazione dei prodotti per settore di attività economica è la *CPATECO*, che deriva sempre dalla classificazione *ATECO* ma ne costituisce un'ulteriore disaggregazione. La *CPATECO* è definita a partire dalle voci della Nomenclatura combinata, raggruppate sulla base della classificazione dei prodotti associata a quella delle attività economiche. La Classificazione delle attività economiche ufficialmente adottata dall'ISTAT è l'*ATECO2002* assimilabile, fino alla quarta cifra di dettaglio, alla classificazione ufficiale dell'Unione Europea *NACE rev. 1*.

Per *sistema armonizzato (SH6)* intendiamo la classificazione economica delle merci stabilita a livello mondiale dal Comitato di Cooperazione Doganale. E' costituita

da raggruppamenti di merci in oltre 5.000 posizioni a sei cifre (vedi Nomenclatura combinata).

Per *nomenclatura combinata (NC8)* invece intendiamo la classificazione economica delle merci adottata nelle rilevazioni del commercio estero dai paesi dell'UE. E' costituita da raggruppamenti di merci in 10.000 posizioni ad otto cifre e costituisce un'ulteriore disaggregazione del Sistema Armonizzato. Con la sigla *SH-NC* intendiamo quindi "nomenclatura combinata-sistema armonizzato", in pratica questi metodi di classificazione vengono usati in modo congiunto a seconda dell'ambito di applicazione.

Fig. 3.1 – Disaggregazione ed aggregazione dei codici a livello nazionale, comunitario ed internazionale



Nomenclatura combinata a livello comunitario che riprende quella del sistema armonizzato

DISAGGREGAZIONE PER IL COMMERCIO ESTERO, PER IMPORRE DAZI E TARIFFE, la settima e l'ottava cifra identificano le sottovoci NC. Quando le voci o sottovoci del sistema armonizzato non sono ulteriormente suddivise per esigenze comunitarie, la settima e l'ottava cifra sono "00".

### 3.3 La classificazione dei prodotti alla destinazione

Come s'intuisce facilmente la classificazione precedente fornisce un elenco di prodotti definiti sulla base della produzione industriale o meglio sui settori di attività economica; possiamo utilizzarlo per condurre un'indagine sulle famiglie? Assolutamente no, poiché se i prodotti li definiamo su base produttiva, cioè all'origine e non alla destinazione, a livello logico commettiamo un grave errore; ad esempio a livello di produzione industriale, se vogliamo classificare una lattina in quanto contenitore di qualcosa, quest'ultima dal punto di vista produttivo avrà un suo codice ma non sarà specificato il suo contenuto, e chi produce le lattine sarà un produttore di lattine, ma solo ed esclusivamente di lattine, non di lattine con un qualche tipo di liquido contenuto all'interno; ora se assumiamo che la nostra indagine verta sul consumo di coca-cola, di lattine di coca-cola, certamente non potremo effettuare l'indagine basandoci sulla codifica per origine di attività produttiva poiché ovviamente bere una lattina di coca-cola o bere una lattina di nafta dal punto di vista del consumo cambia in modo considerevole. Quindi classificare all'origine del prodotto o alla destinazione, cioè al consumo è profondamente diverso.

La classificazione alla destinazione, che utilizziamo per i prodotti è denominata con la sigla *COICOP* ed anche questa è una classificazione standardizzata a livello comunitario.

Abbiamo visto quindi due tipologie di classificazione dei prodotti; ma le classificazioni che abbiamo utilizzato per il commercio estero sono classificazioni costruite per omogeneità di tipologia produttiva o per omogeneità di prodotto e quindi classificazioni alla destinazione? Naturalmente non possono essere costruite che per omogeneità produttiva; di solito nel commercio estero l'ultimo passo, l'ultima fase della realizzazione di un prodotto viene effettuata nel paese in cui il prodotto stesso viene commercializzato; riprendendo l'esempio delle lattine, la distinzione tra lattine di nafta o di coca-cola viene fatta alla destinazione, ovvero è nel paese di destinazione che le lattine, che hanno passato la frontiera con un determinato codice per tipologia di produzione, per omogeneità di attività produttiva, vengono riempite con il liquido loro destinato; quindi tra il *sistema armonizzato-[nomenclatura combinata]* e il sistema *CPATECO* esiste una comparabilità quantomeno logica. Tra l'altro esistono delle tabelle

di confronto che indicano quali prodotti della *CPATECO* corrispondono al *sistema armonizzato-nomenclatura combinata* e viceversa.



Quindi queste due classificazioni sono confrontabili quantomeno a livello logico, non con proporzionalità 1:1 ma ci avviciniamo molto. Invece non potrà mai esserci comparabilità tra una classificazione all'origine ed una classificazione alla destinazione dei prodotti.



Tuttavia esistono alcune eccezioni; effettivamente alcuni prodotti vengono già imballati ed impacchettati all'origine, quindi in alcuni casi particolari siamo costretti a classificare per destinazione e non per omogeneità di attività produttiva. Questi casi comunque non superano la percentuale del 5%. La classificazione *COICOP* è estremamente importante per stimare l'indice dei prezzi al consumo; famosa è la polemica riguardante il paniere adottato per la stima di questo indice su cui, ovviamente, la *COICOP* non influisce dal punto di vista metodologico in quanto costituisce solo, tramite i codici, un elenco infinito di prodotti e nient'altro, e nel paniere sono inclusi solo alcuni prodotti di questo elenco, quelli che probabilmente sono i più consumati all'interno del paese, 600-700 prodotti. La polemica che si è creata è basata in pratica sulla velocità con cui l'Istat aggiorna l'elenco dei prodotti, cosa che l'Istat non può fare arbitrariamente se non tramite un accordo con altri organi, ministero del lavoro, etc.

Inoltre oltre ai prodotti vanno aggiunti tutti i servizi; l'80% del consumo è costituito da servizi. La classificazione dei servizi non è presente all'interno del *COICOP*, non è

armonizzata, ha esclusivamente carattere nazionale, prettamente italiano se stiamo considerando i prodotti alla destinazione; invece nel caso in cui dobbiamo classificare un'attività produttiva di servizi allora è importante attribuire il codice giusto per omogeneità di tipologia produttiva.

### **3.4 Approfondimenti**

Come pensiamo che funzioni l'attribuzione di un codice *ATECO* ad un'impresa? C'è un sistema informatizzato che ad ogni impresa o ad ogni unità locale attribuisce un codice, basato sulla descrizione e definizione italiana dell'unità statistica, naturalmente in modo sintetico; ma spiegare ad un computer l'italiano è una cosa molto complicata, quindi potrebbero sorgere degli errori. Questa attività di codifica automatica ha una probabilità di errore di attribuzione codice pari più o meno al 30%. Ma questo errore è lo stesso che otteniamo se chiediamo a due aziendalisti di attribuire un codice ad un'attività, quindi sbagliare tre volte su 10 è un problema fisiologico del sistema.

Il più importante dato tra quelli di cui sentiamo parlare spesso nei telegiornali è il Pil, che non è un'indagine diretta ma è il risultato di un calcolo basato sulla contabilità nazionale, è sempre l'Istat che esegue l'operazione, svolgendo prima indagini su altre variabili (consumi, investimenti, etc.), e successivamente tramite semplici calcoli di contabilità nazionale definisce il valore del Pil. Questo calcolo è basato su specifiche disposizioni della comunità europea riguardo alle operazioni di aggregazione dei risultati provenienti dalle altre indagini. Sappiamo che supponendo che nel sistema economico di un paese dove non esiste il commercio estero l'equazione di contabilità nazionale è  $Y = C + I + G$ , è da questa equazione, qui estremamente semplificata, che si ricava il Pil.

Ma perché il prodotto interno lordo è così importante? Come avremo avuto modo di sapere, da qualche anno la permanenza all'interno dell'area euro è condizionata al rispetto di alcuni parametri tra i quali il Pil (soglia del 3%) nonché il rispetto di un altro parametro calcolato dall'Istat sull'aumento dei prezzi (inflazione nazionale non superiore all'1,5% della media dei best performing all'interno dell'area-euro).

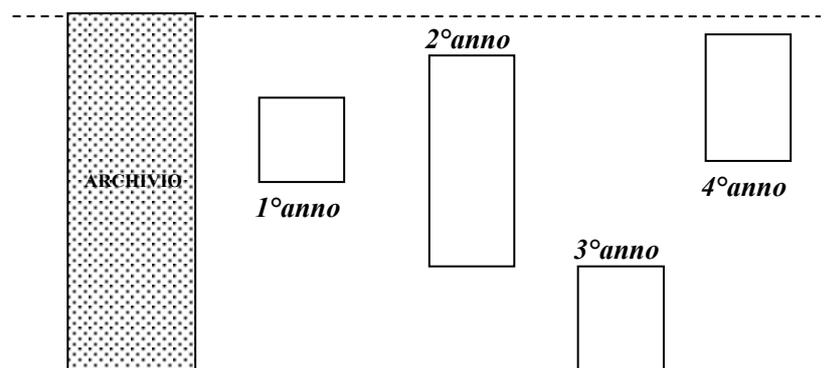
### **3.5 Indagine sull'andamento dei prezzi al consumo**

Vediamo finalmente come si conduce un'indagine sull'aumento dei prezzi al consumo; prima di tutto dobbiamo definire l'unità statistica, chiediamo alle imprese/unità locali del codice di attività economica *ATECO* commercio oppure al consumatore?

Ad ambedue i soggetti. Certamente non potremo chiedere il prezzo alla famiglia poiché dovremmo chiedere quantità e valore per poi moltiplicarli, cosa che per quanto banale è impossibile, il prezzo unitario è rilevabile solo al venditore. Le indagini sui consumatori e quindi intervistando le famiglie per valore e quantità vengono utilizzate di solito esclusivamente come indagini di mercato perché praticamente bisogna rilevare tutti prodotti che entrano in una casa per quantità e valore o in altri termini per codici a barre. L'indagine sui prezzi quindi viene svolta tramite due rilevazioni, in una rilevazione chiediamo il prezzo unitario al rivenditore, mentre nell'altra rilevazione intervistiamo la famiglia riguardo ai consumi, quest'ultima utilizzata per sommare i prezzi unitari, per mediarli, per fare i cosiddetti pesi del paniere, poiché naturalmente non tutti i codici pesano allo stesso modo, ogni singolo codice pesa per il valore consumato.

### 3.6 Indagini trasversali e longitudinali

Ciò che ci interessa non è il livello dei prezzi ma la loro variazione. Per questo è fondamentale introdurre il concetto di *INDAGINI TRASVERSALI E LONGITUDINALI*. Dicesi *TRASVERSALE* un'indagine in cui da un'occasione all'altra il meccanismo di selezione delle unità statistiche è indipendente, ovvero se abbiamo una popolazione ed estraiamo un campione in qualche modo, in ogni periodo prendiamo un campione che non tiene conto di quello del periodo precedente, quindi estraiamo il campione in modo indipendente;



*Fig. 3.2 – Indagine trasversale*

Dicesi *LONGITUDINALE* invece un'indagine in cui con il nostro sistema di selezione tende ogni anno a selezionare le unità in modo ordinato, secondo il metodo che è stato utilizzato nelle occasioni precedenti.

Quindi per conoscere la variazione sui prezzi lo stesso rivenditore deve essere intervistato tutti gli anni se l'indagine è su base annuale. Dunque per ottenere delle variazioni bisogna intervistare le stesse unità in più occasioni temporali, è l'unico modo per non incorrere nell'*EFFETTO MEMORIA* che distorce tutta la nostra indagine.

Quindi l'indagine sul commerciante al dettaglio deve essere fatta obbligatoriamente in modo longitudinale, ovvero nel tempo bisogna tenere conto del meccanismo di selezione delle unità campionarie. Ogni quante volte rileviamo? La rilevazione può essere anche mensile. Ritornando alla polemica di cui abbiamo parlato nelle pagine precedenti, il ventre molle dell'indagine non si ravvisa nella costruzione del paniere e nemmeno nel ponderare i consumi, il vero problema va individuato nel rilevatore, solitamente è un ufficiale comunale che rileva fisicamente i prezzi al rivenditore e certamente non è un rilevatore qualificato (proprio perché la rilevazione è fatta prendendo un campione di comuni e poi un campione di rivenditori), per legge però deve essere una guardia comunale a svolgere la rilevazione, cui non interessa avere dettaglio ed attenzione e non è cosciente del valore del compito che sta svolgendo. A livello sostanziale il meccanismo di rilevazione dei dati non è affidabile, poi tutte le altre considerazioni riguardo alla polemica sono poco importanti, sarebbe più opportuno cambiare la legge che obbliga gli ufficiali comunali ad effettuare le rilevazioni.

Naturalmente per quanto riguarda le varie interpretazioni economiche del dato riguardante la variazione dell'indice dei prezzi al consumo, i vari allarmismi derivanti da interviste fatte in modo sconsiderato ai consumatori nei mercati regionali, che solo su impressioni personali pensano che il prezzo di un prodotto sia aumentato, sono spesso infondati. Tutti i dubbi potrebbero essere risolti studiando la documentazione disponibile presso l'Istat che riporta i vari pesi per ogni codice del paniere, in modo da evitare allarmismi ingiustificati. È naturale poi che la media dei vari pesi possa nascondere valori elevati dell'indice, ma non è l'Istat responsabile di un comportamento scorretto o incompetente.

### 3.7 Approfondimenti

In un anno vengono rilevati, mese per mese circa 300.000 rivenditori, e la dimensione è così grande, proprio perché dobbiamo rilevare un numero di codici molto elevato; dobbiamo evitare che un prodotto entri nel campione una sola volta, o comunque troppo poco, proprio perché 800 dati, i codici COICOP sono molti.

L'indagine sui bilanci delle famiglie, dalla quale rileviamo i consumi su cui possiamo pesare i prezzi unitari, prende un campione di 20.000 famiglie, teoriche 30.000, poiché una famiglia consuma svariati prodotti, e 20.000 famiglie sono più che sufficienti per calcolare i prodotti in valore.

Continuando a parlare dell'indagine sui bilanci famigliari per quali altri motivi è importante un'indagine sui consumi famigliari? Sono le più importanti, se non le uniche, che solitamente interessano le società private, soprattutto a scopo di marketing, sono indagini riguardanti le famiglie intese come insieme di consumatori, le potenzialità della famiglia come insieme di consumatori.

Dopo aver costruito il nostro archivio statistico, cosa chiediamo alla famiglia che rappresenta la nostra unità di rilevazione? Prima di tutto qualsiasi indagine siamo chiamati a svolgere dobbiamo costruire un quadro di riferimento altrimenti poi non siamo in grado di riportare le stime all'universo, ovvero dobbiamo sapere a cosa si riferiscono i consumi che abbiamo rilevato, quali sono i consumi tipici di determinate categorie di famiglie, ovvero dobbiamo categorizzare la famiglia in un modo che può essere utile; inoltre dobbiamo sempre ricordare di chiedere i dati che sono contenuti nell'archivio poiché quelli della fonte dell'archivio di solito non sono dati statistici ma dati amministrativi; ad esempio il numero di addetti che sono iscritti sull'archivio delle imprese è un dato amministrativo, e quindi anche se hanno la stessa definizione i dati d'archivio e quelli del questionario sono sempre diversi, anche se fanno affidamento quelli dell'archivio; un'altra cosa molto importante, nella definizione del disegno dell'indagine è la periodicità. Quale prendiamo? Chiediamo i consumi del giorno, della settimana, del mese o dell'anno? E per quanto riguarda la periodicità del periodo di

rilevazione, quale prendiamo? Teniamo il questionario presso le unità di rilevazione una settimana, un mese, un anno?

La scelta dell'Istat nell'indagine sui consumi è settimanale, la famiglia campionata viene intervistata per una settimana, viene lasciato il questionario presso la famiglia. E le famiglie ruotano in modo da stare in un dodicesimo, per ogni mese, e ruotate nelle settimane all'interno del mese. Perché questa distribuzione temporale? Perché se svolgiamo l'indagine su un solo mese otteniamo delle stime non utili alla nostra indagine, in quanto esistono molti prodotti che risentono di una forte stagionalità. Ad esempio se facciamo un'indagine durante il periodo pasquale il consumo di agnello sarà sicuramente limitato, ecco perché abbiamo bisogno di distribuire il campione durante l'anno. Quindi dobbiamo considerare tre tipi di periodicità, ovvero il periodo di riferimento della variabile, quanto deve durare l'indagine e quanto tempo bisogna lasciare il questionario presso le famiglie.

Quali altre indagini sono importanti? Reddito, risparmi e consumi sono aspetti della stessa indagine, quindi un'indagine solo sul reddito non ha senso. Le indagini sui bilanci delle famiglie vengono effettuate dall'Istat, dalla Banca d'Italia (governativa) e da una società privata, multinazionale, chiamata *AC-NIELSEN*. L'indagine dell'Istat è particolarmente dettagliata sui consumi in quanto questi ultimi servono a pesare i prezzi unitari mentre i redditi ed i risparmi si chiedono come totali in qualche altro modo; l'indagine della Banca d'Italia è la consorella, il campione è più o meno simile, stesse numerosità più o meno ma molto dettagliata sui risparmi è poco sui consumi, mentre l'indagine dell'*AC-NIELSEN* è molto dettagliata sui consumi.

Molto importanti sono le indagini congiunturali sulle imprese (tra cui c'è anche quella sulla produzione industriale) e le indagini sulle forze lavoro (stima del tasso di disoccupazione). Per indagine congiunturale s'intende un'indagine fatta su un piccolo campione d'imprese, di solito le più grandi, ovvero la produzione industriale viene rilevata solo sulle grandi imprese e quindi campioni a bassa significatività statistica ma con un'altissima frequenza temporale, e non viene fatta solo sulla produzione industriale ma anche sul fatturato, etc. Quindi campioni non molto grandi, piccoli, e solitamente le indagini non sono indirizzate a fare stime di livello ma di variazione, ovvero dall'indagine sulla produzione industriale ad esempio si può dire, se quest'ultima è

aumentata o è diminuita, ma non il livello in cui si trova. Queste tre indagini che abbiamo definito sono le più importanti a livello politico ed economico.

## 4. Tipologia dell'indagine

### 4.1 Indagini trasversali e longitudinali

Nelle pagine precedenti abbiamo accennato, riguardo all'indagine sui consumi, ai due tipi di meccanismi di selezione delle unità statistiche. Quando c'è una ripetizione della rilevazione sulle unità avremo una rilevazione *PANEL*. Ma perché dovremmo fare una cosa del genere? Perché ripetiamo la rilevazione sulle stesse unità statistiche? Perché stabiliamo una dipendenza tra le unità statistiche? Oppure potremmo costruire un cosiddetto *PANEL RUOTATO* e sottoporre nuovamente ad esempio il questionario all'80% delle unità campionarie precedentemente intervistate; perché lo facciamo? Perché stabiliamo che ci debba essere questa sorta di coordinamento tra le unità statistiche?

L'indagine *PANEL* viene eseguita principalmente per quattro motivi: a) in quest'indagine un tasso di variazione è più facilmente calcolabile; b) permette di evitare problemi di *EFFETTO MEMORIA*; c) nell'indagine con intervista di solito buona, parte del costo è giustificato dal contatto dell'unità statistica, quindi nelle indagini *PANEL* diminuisce il costo "di rompere il ghiaccio" tra rilevato e rilevatore; d) di solito la qualità della compilazione del questionario da parte del rilevato è maggiore. D'altra parte presenta dei difetti tra cui i principali sembrano essere tre: a) potrebbe verificarsi quella che nella letteratura viene definita *MOLESTIA STATISTICA*, oppure con un termine anglosassone "*PANEL ATTRITION*"; b) ha un grosso difetto metodologico in quanto se sbagliamo il campione ci portiamo dietro tutti gli errori a quest'ultimo attribuibili; c) supponendo anche di non fare alcun errore nell'estrazione del campione, la popolazione cambia nel tempo, quindi il campione che avevamo progettato su un determinato tipo di popolazione e secondo determinati criteri si allontana sempre di più dalla popolazione reale e quindi diminuisce la sua rappresentatività, per questo motivo si usa fare una rotazione del panel, aggiornare il campione almeno di un 10%-15% serve proprio ad evitare i due errori precedenti, il 10%-15% dell'archivio viene estratto ex

novo casualmente dalla popolazione proprio per seguire i cambiamenti e sostituire eventuali famiglie soggette a *PANEL ATTRITION*;

Alla luce delle nuove conoscenze un'indagine sui consumi delle famiglie potrebbe essere costruita in modo PANEL? Delle tre indagini di cui abbiamo parlato nel capitolo precedente l'*AC-NIELSEN* effettua il *PANEL PURO*, poiché essendo una società privata difficilmente trova famiglie disponibili a rispondere a questionari e domande, a collaborare, quindi appena individua una famiglia disponibile, si guarda bene dal farsela scappare coccolandola, in quanto società privata, con gadget, regali, etc.

Anche l'indagine della Banca d'Italia viene fatta con *PANEL PURO* perché a quest'ultima interessa, poiché il risparmio in un anno non si basa né sul consumo né sul reddito, seguire l'andamento di un progetto globale di vita sul risparmio. Il risparmio è poco funzionale al consumo quotidiano di una famiglia, magari una dipendenza più stretta si può verificare con i beni durevoli in quanto il risparmio si presume sia funzionale a questi ultimi, dunque è giustificata l'indagine della Banca d'Italia nel seguire la variabile su un arco di tempo più esteso; questa è un'indagine che esiste dagli anni '70, addirittura esistono quindici famiglie facenti parte della prima ondata campionata; avranno sposato i rilevatori?

L'indagine dell'Istat non è *PANEL* ma fatta tramite campioni casuali indipendenti; Quindi come vediamo stabilire il modo in cui deve essere estratto il campione nei diversi periodi ha una sua importanza che dipende dal tipo di variabili e dal modo in cui devono essere studiate.

## **5. Il Questionario**

### **5.1 Tecniche di somministrazione**

Le più usate tecniche di somministrazione del questionario sono quattro: l'indagine telefonica, postale, faccia a faccia o rilevazione diretta e tramite e-mail. *L'indagine telefonica* viene svolta principalmente per tre motivi: è molto veloce, chi sta al telefono digita contemporaneamente sul computer le risposte degli intervistati, quindi tempo di rilevazione e registrazione vengono compressi in modo sensibile; i costi sono contenuti; l'interazione potrebbe essere un pregio in quando l'intervistatore può spiegare le domande, ma non sempre. Tra i difetti invece abbiamo che c'è un alto tasso di mancata

risposta, chi viene intervistato può abbassare la cornetta immediatamente, e quindi alto tasso di mancato reperimento; comunque in tutti i contratti fatti con le società delegate a fare queste interviste c'è sempre un numero minimo di contatti da garantire; non abbiamo il controllo su chi risponde all'intervista, se risponde una donna o un bambino, controllo sul rispondente e quindi l'unità informativa, la stessa cosa avviene ad esempio quando vengono fatte delle interviste alle imprese, non sempre chi risponde è l'ufficio giusto, non c'è controllo; non è applicabile per questionari che richiedono tempo per la loro compilazione, certamente non potrebbe essere applicata per un'indagine sui consumi o quantomeno non dovrebbe essere svolta in forma pura ma occorrerebbe un'indagine mista; in generale il questionario telefonico si può fare solo per domande brevi e per poche domande, e soprattutto molto specifiche senza molte alternative. Insomma l'intervista telefonica è quella che pone più vincoli sia sulla dimensione sia sulla tipologia di questionario sia si può utilizzare.

Cinque sono invece i motivi principali che spingono all'utilizzo della somministrazione tramite *indagine postale*: tempo sufficiente per rispondere al questionario; è meno costosa dell'intervista telefonica; si può compilare il questionario in modo molto più dettagliato; il questionario può avere qualsiasi lunghezza; c'è un tasso di mancata risposta inferiore all'intervista telefonica. Anche in questo caso però possiamo individuare dei difetti: è obbligatorio correlare il questionario con ampie, precise e dettagliate istruzioni di compilazione poiché viene meno il contatto con il rilevatore, e soprattutto mai fare un'intervista postale senza inserire un numero verde per eventuali ulteriori spiegazioni; di solito è su base annuale; è molto lenta, proprio per il fatto che chi deve rispondere rimanda continuamente la compilazione, addirittura spesso si invia il questionario tramite posta e poi si chiedono le risposte tramite telefono; è l'indagine in cui risponde il minor numero di persone, ha il più alto tasso di mancata risposta, poi viene la telefonica e poi le altre, di solito ad un'intervista postale si appoggia sempre un'intervista telefonica per sollecitare la risposta, di solito due o tre solleciti. Questo tipo di indagine è indirizzata alle imprese, perché le indagini sulle imprese risentono poco dei difetti ma molto dei pregi, si chiedono poche cose ma molto complicate, possono essere necessarie settimane per la compilazione, inoltre l'impresa risponde a prescindere dal tipo di indagine ed abbiamo più probabilità che chi compila il questionario sia la persona giusta. Il tasso di mancata risposta è vicino al 30%.

Nell'indagine faccia a faccia tra gli aspetti positivi abbiamo che: si può scegliere chi intervistare; la qualità dei dati è maggiore quando è obbligatoria e necessaria la presenza di un rilevatore rispetto ad altre indagini; mentre tra gli aspetti negativi abbiamo un effetto rilevatore che potrebbe influenzare l'intervistato e chiaramente il costo è maggiore di qualsiasi altro tipo di indagine. Questa tecnica ci permette di riconoscere quando le unità statistiche hanno interesse a mistificare i dati, quindi il rilevatore può osservare se i risultati sono coerenti con la realtà che osserva;

Infine per quanto riguarda *l'indagine tramite e-mail*, innanzitutto possiamo dire che quest'ultima rientra nell'ambito dell'indagine postale in quanto si potrebbe, quando possibile, compilare il questionario ed inviarlo tramite la rete. In questo tipo di indagine il grande difetto è che il Web è ancora poco diffuso e di solito viene utilizzata per indagini dirette alle imprese ed il motivo è chiaro, in questo caso almeno il 90% delle imprese risponde, tolti i casi di mancata risposta non dovuti a motivi tecnologici; solitamente le indagini tramite e-mail vanno a sostituire la parte della trasmissione dati, ovvero viene mandato il questionario per carta, dicendo che lo stesso questionario è disponibile su un sito in via telematica, dove è possibile compilarlo, e quindi il trasferimento dati è molto semplificato. Questo tipo d'indagine viene chiamata EDI (Electronic Data Interchange).

## **5.2 Alcuni casi concreti sull'intervista faccia a faccia**

Per quanto riguarda l'indagine sui consumi delle famiglie è obbligatorio fare l'intervista faccia a faccia? In realtà sì, perché l'auto compilazione del questionario è molto complicata, addirittura l'indagine dell'*AC-NIELSEN* è tutta basata su codici a barre.

L'indagine della banca d'Italia è realizzata faccia a faccia, è un funzionario della banca che esegue l'intervista. L'indagine sulle forze lavoro viene fatta in modo simile al censimento, nonché faccia a faccia tramite un rilevatore selezionato dal comune, purtroppo poco competente;

Il censimento è il caso classico dell'esistenza del rilevatore faccia a faccia, è fatto in modo misto, di solito il rilevatore consegna il questionario fisicamente limitandosi a questo, allora il censimento è più simile ad un'indagine postale. Ma perché allora si reca personalmente nelle case? Perché il censimento è una forma mista di rilevazione, di

solito il rilevatore si reca di persona presso le famiglie per constatare se è necessario il suo aiuto per problemi culturali, per problemi di lingua, etc. Nel caso si riscontrino problemi di questo tipo il rilevatore diventa tale a tutti gli effetti. Per le indagini sui dati di un'azienda non si usa assolutamente l'intervista faccia a faccia, mai nessuna azienda permetterà ad uno sconosciuto di invadere la segretezza dei dati, per questo si usa un'indagine postale, e l'azienda può rispondere quando vuole. L'intervista faccia a faccia invece è obbligatoria nelle indagini riguardanti l'agricoltura, poiché il diverso livello culturale potrebbe inficiare i dati, inoltre esistono problemi relativi alle unità di misura, ad esempio in Italia esistono un'infinità di unità di misura per determinare l'estensione di un terreno; quindi bisogna mandare un rilevatore che faccia un'adeguata conversione. Da un'esperienza personale del prof. Benedetti, ad esempio, si è riusciti a contare 39 modi di pesare i fiori .

### **5.3 Esercitazione sui voli aerei**

Vogliamo sapere se il pubblico è soddisfatto di un servizio che noi offriamo, ad esempio sui voli aerei che offre la nostra compagnia aerea; dobbiamo definire la forma di somministrazione del questionario (telefonica, faccia a faccia, postale, etc.), l'unità statistica, la costruzione dell'archivio, la periodicità, (attenzione, i voli aerei hanno una stagionalità), etc., insomma dobbiamo applicare tutte le conoscenze acquisite finora.

Prima di tutto dobbiamo studiare il fenomeno: *grado di soddisfazione dei nostri voli*. Riguardo all'unità statistica se prendiamo come unità di rilevazione l'individuo, sbagliamo in quanto non possiamo costruire, dopo aver definito l'unità statistica, l'archivio; nel nostro caso prendiamo come definizione di unità statistica il vettore ovvero *l'aereo*, e volendo possiamo stratificare per *durata del volo*; quindi l'unità di rilevazione di primo stadio è *l'aereo*, l'unità di rilevazione finale è *il viaggiatore* e corrisponde con l'unità di analisi. Quindi per costruire l'archivio dobbiamo scegliere la giusta unità di rilevazione. Distribuiamo i *questionari sull'aereo* e poi questi vengono raccolti alla fine del volo dal personale di bordo, poiché i passeggeri sfruttano il tempo a loro disposizione per riempirli. La periodicità dell'indagine sarà con cadenza almeno *trimestrale* in quanto il nostro fenomeno ha una forte stagionalità; inoltre c'è una forte stagionalità lungo *la settimana*, quindi la rilevazione viene fatta su base trimestrale ed il

singolo campione viene distribuito lungo i giorni delle settimane componenti i trimestri; è molto importante distribuire il campione lungo la settimana, tra l'altro il trimestre invernale è molto particolare, quindi dobbiamo dividere l'anno in blocchi omogenei tramite delle informazioni ottenuti a priori. L'indagine non può essere panel logicamente, perché è impossibile intervistare più di una volta lo stesso passeggero, quindi sarà *trasversale*, inoltre un'indagine longitudinale non avrebbe significato in quanto in un'indagine sulla *customer satisfaction* non occorre rilevare alcuna variazione;

#### **5.4 Esercitazione sui treni**

Dobbiamo prima di tutto tener presente che cambiando il vettore può cambiare la metodologia d'indagine. La prima grande differenza è che il treno è molto più capiente, trasporta molta più gente ed il viaggio non necessariamente dura delle ore, ad esempio esistono i treni locali, quindi l'unica soluzione possibile sta nell'estrarre il vagone e far ritirare i questionari dal controllore durante la validazione dei biglietti. In questo tipo di indagini non possiamo mai chiedere il numero di telefono.

Un grosso vantaggio deriva dal fatto che esiste solo Trenitalia, quindi in teoria potremmo usare anche il censimento, ma sarebbe costoso; quindi si potrebbe utilizzare, come definizione di unità statistica il passeggero, senza passare per il treno.

#### **5.5 Impostazione del questionario**

L'esigenza informativa del committente deve essere tradotta dallo statistico nel modello concettuale statistico. Supponiamo si voglia rappresentare un campo di girasoli. A fronte di questa esigenza conoscitiva vi sono diversi approcci. Un giornalista può descrivere questo fenomeno raccontando con il linguaggio naturale quanto è esteso, quanto maturi sono i frutti, etc. Un poeta può dare un'interpretazione che evoca sentimenti, fino a Van Gogh dove il campo di girasoli evoca un turbine di piccoli soli.

Lo statistico a fronte del campo di girasoli mira a una rappresentazione quantitativa e oggettiva, pur all'interno della scelta soggettiva del modello concettuale statistico. Il

modello concettuale statistico è lo strumento che lo statistico sceglie per rappresentare un fenomeno collettivo. Il modello concettuale statistico è definito da:

- *La popolazione obiettivo* e le unità che vi appartengono;
  - *I caratteri* (variabili) da rilevare su ciascuna unità (le domande e loro tipologia);
  - *Le modalità* dei caratteri (le risposte previste per ciascuna domanda);

La popolazione obiettivo è un insieme definito in modo che data una qualunque unità o evento reale si possa dire se appartiene o non appartiene all'insieme. Le unità della popolazione obiettivo prendono il nome di unità statistiche. Occorre che la definizione sia precisa e preveda scelte univoche per situazioni talvolta incerte. Si pensi alla definizione di famiglia, di impresa, abitazione, struttura ricettiva, ecc. Spesso le norme giuridiche forniscono un buon punto di riferimento ma casi dubbi si trovano anche inaspettatamente.

Una delimitazione della popolazione obiettivo è quella rappresentata dal tempo (istante o periodo) e dallo spazio. Talvolta l'indagine riguarda più popolazioni obiettivo gerarchicamente collegate (es. famiglie e individui, imprese e unità locali), in tal caso si devono definire esattamente le diverse popolazioni e la relazione che le collegano. In presenza di più popolazioni obiettivo è molto utile costruire il modello concettuale *E/R* (Entità/Relazione) col quale si rappresentano le popolazioni (Entità) e le relazioni che le collegano (Relazioni). Le informazioni statistiche sono valori numerici associati a queste popolazioni e le caratterizzano quantitativamente.

I caratteri sono le misure (le domande) che vogliamo effettuare su ciascuna unità statistica appartenente alla popolazione obiettivo. Queste vanno definiti tenendo conto dell'esigenza informativa. Ad esempio per rappresentare se e quanto i residenti in Toscana usano internet potremmo considerare i seguenti caratteri:

Possesso del computer	<i>(Possiede un computer?)</i>
Collegamento a internet	<i>(E' collegato a internet?)</i>
Frequenza settimanale di uso	<i>(Quante volte si è collegato nell'ultima settimana?)</i>

I caratteri si presentano in ciascuna unità statistica con un insieme di modalità che il ricercatore definisce sempre avendo presente gli obiettivi informativi. Per un dato carattere si possono considerare più modalità (modalità di risposta).

Esempio: per la domanda "Quante volte si è collegato a internet nell'ultima settimana?" si possono prevedere diverse tipologie di modalità di risposta:

*0, 1, 2, 3, 4, 5, 6, 7*

Mai, qualche volta, tutti i giorni.

A una domanda sulla soddisfazione rispetto a un servizio si può prevedere in risposta:

Una scala numerica, *1.....10*

Una scala di giudizio: "per niente, poco, abbastanza, molto".

Una rilevanza particolare assume la distinzione dei caratteri in qualitativi e quantitativi. Questa distinzione ha una particolare rilevanza nella statistica e nel campionamento in particolare: ad esempio per i primi è più facile fare ipotesi sulla varianza del carattere (variabile) nella popolazione.

Modelli concettuali statistici diversi adottati per uno stesso obiettivo informativo conducono a risultati diversi. E' questo spesso motivo di scontro fra interlocutori che, pur appoggiandosi a dati statistici, sostengono tesi diverse, perché i dati a sostegno delle rispettive tesi si riferiscono a modelli statistici diversi.

Nell'ambito della statistica pubblica ci sono organismi che mettono a punto metodologie e classificazioni standard per limitare al minimo le differenze fra i modelli statistici e garantire quindi la comparabilità dei risultati.

Un ulteriore fattore che influenza i risultati è legato alle modalità di rilevazione<sup>5</sup>.

Il questionario deve essere sempre anonimo a meno che non sia strettamente necessario; deve essere sempre diviso in sezioni (0,1, 2,etc.) ed ogni sezione deve riguardare un argomento omogeneo al suo interno e diverso rispetto alle altre sezioni, magari mettere anche il titolo della sezione. Principalmente il questionario può essere diviso in quattro sezioni:

- I. *Sezione iniziale:* logo della società che effettua l'indagine, le varie leggi sulla privacy, etc., dobbiamo far capire chi siamo;
- II. *Prima Sezione:* questa sezione deve sempre riguardare i dati personali dell'intervistato poiché il nostro obiettivo è riuscire a discriminare in qualche modo le unità statistiche, altrimenti l'indagine non serve a nulla (caratteristiche strutturali dell'unità statistica, età, sesso, professione, quante volte viaggia, da solo o in compagnia, quanto si trattiene, quante volte viaggia etc.);
- III. *Seconda sezione:* dopo le caratteristiche generali bisogna poter effettuare la distinzione tra i passeggeri occasionali e non, un po' di passato, presente e futuro del passeggero che vola come nell'esercitazione precedente, ad esempio quante tratte si fanno, se in futuro si volerà di più o di meno, insomma dobbiamo conoscere con che tipo di passeggero abbiamo a che fare, prima abbiamo definito le caratteristiche strutturali mentre ora quelle del passeggero;
- IV. *Terza sezione* - chiediamo il grado soddisfazione ovvero la nostra variabile;

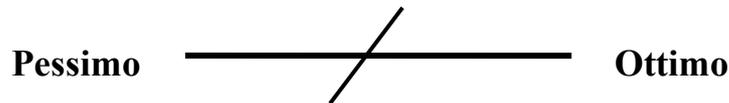
---

<sup>5</sup> Le indagini telefoniche CATI, *esperienze di disegno e implementazione presso la Regione Toscana*,

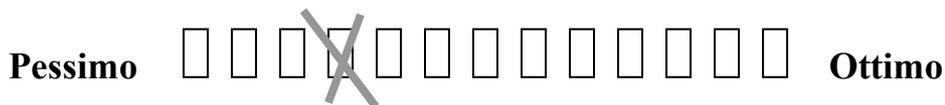
### **5.5.1 Come chiediamo il grado di soddisfazione?**

Vista la soggettività del giudizio su possibili aggettivi come ad esempio, buono, ottimo, sufficiente etc., dobbiamo introdurre un diverso criterio, non possiamo stabilire la distanza che c'è tra un giudizio ed un altro, alcuni addirittura associano in fase di codifica i giudizi ad i numeri, ma la distanza che c'è tra 1 e 2 ad esempio non è la stessa che c'è tra un buono ed un sufficiente, qualcosa di concettualmente astratto ed indefinibile, se proprio dobbiamo introdurre degli aggettivi e sempre meglio metterne un numero in modo da formare una simmetria, quindi quantomeno bilanciare i giudizi, sia negativi sia positivi, e magari in numero pari se si vuole evitare un'eccessiva concentrazione delle risposte nella domanda centrale, ad esempio, per niente soddisfatto, insoddisfatto, soddisfatto, molto soddisfatto, oppure al centro una risposta neutra, indifferente, quindi buona norma è che da un lato e dall'altro dell'indifferenza mettiamo stessi termini ma opposti, allontanandoci dal punto centrale.

Se proprio vogliamo introdurre un metodo per trasformare un giudizio qualitativo in un quantitativo possiamo introdurre il cosiddetto *DIFFERENZIALE SEMANTICO*, da un lato di un segmento mettiamo il giudizio peggiore ed dall'altro mettiamo il giudizio migliore e chiediamo all'intervistato di mettere un segno sul segmento:



Ma questa è solo un concetto teorico poiché a livello pratico non è accettabile, il segno potrebbe andare a finire da tutte le parti tranne che sulla riga; inoltre come misuriamo, con il righello? Questo problema di solito viene risolto con una linea composta da quadratini, sicuramente in modo più efficace ed efficiente;



Questa modalità ha diversi vantaggi: possiamo tirar fuori non una variabile qualitativa ma una vera e propria misura quantitativa su cui effettuare le nostre medie, varianze, etc.; non ha tutti gli svantaggi che abbiamo visto sopra, non raccoglie ad esempio le risposte nella parte centrale; altro grosso vantaggio è che permette di ridurre la lunghezza del questionario e riduce i tempi, un questionario fatto in questo modo è molto più scorrevole, è la forma migliore per chiedere giudizi. Forse l'unico svantaggio tipicamente italiano è che di solito si hanno meno risposte sulle code.

## 5.6 Come si formulano le domande di un questionario?

Redigere un questionario è una cosa complicatissima e richiede una certa multidisciplinarietà; molti dilettanti considerano la realizzazione di un questionario una fase molto banale quando è invece la più complicata. C'è bisogno che un esperto coadiuvi questa fase, chi risponde può considerare in modo diverso le domande, ad esempio per le imprese non potremmo chiedere mai il valore aggiunto, in quanto sul bilancio non esiste quella voce e chi risponde potrebbe non considerare accettabili le domande o troppo onerose sotto diversi punti di vista, riguardo al valore aggiunto si

richiede il lavoro del commercialista, quindi magari chiederemo altri dati, abbiamo bisogno di un esperto sull'argomento. Questo vale anche per le indagini sulle famiglie, soprattutto se chiediamo cose psicologicamente complesse avremo bisogno di uno psicologo o di un sociologo.

Inoltre un questionario va redatto in modo da formulare domande in ordine crescente di difficoltà e con l'aiuto di un esperto, principalmente per individui e famiglie. Quali sono le categorie di domande formulabili? Quali sono le cose più facili da chiedere? Elenchiamo le categorie in ordine crescente di difficoltà.

1. *Chiedere fatti* (lo sei, non lo sei, etc.);
2. *Conoscenze* ( lo sai che...);
3. *Opinioni* (come la pensi?);
4. *Atteggiamenti* (cosa hai intenzione di fare?);
5. *Motivazione* (perché mangi la nutella?);

Questa è una classificazione tipica nelle indagini di mercato, di *customer satisfaction*, queste cinque informazioni su un prodotto permettono di orientare le politiche aziendali. Ora volendo enunciare una serie di regole generali per formulare le domande, dobbiamo fare in modo di:

1. *Evitare formulazioni generiche* (ma cosa vuole sapere...);
2. *Non usare terminologie oscure*, i termini che conosciamo noi possono non essere gli stessi del rispondente;
3. *Assicurarci che la domanda rientri nella sfera culturale dell'intervistato*, ossia l'argomento deve essere conosciuto, chi risponde deve sapere di cosa stiamo parlando;

4. *Non indispettire l'intervistato;*
5. *La domanda deve essere definita univocamente per tutti gli intervistati, e non cambiare quindi definizione a seconda dell'intervistato, per tutti deve essere chiara;*
6. *Le domande devono richiedere piccolo sforzo mnemonico;*
7. *Evitare domande che possono creare stati di tensione nell'intervistato (opinioni politiche, etc., di qualsiasi genere);*
8. *Evitare di formulare domande orientate, regola che è impossibile da rispettare per chi redige il questionario, l'unico modo per rispettarla è che ci sia una controparte che ricontrolli le domande;*
9. *In generale i questionari devono essere brevi;*

### **5.6.1 Esercitazione sulla disponibilità economica degli studenti**

Quanti errori ci sono nella domanda: *che lavoro fanno i tuoi genitori?*

1. Se lavorano tutti e due a chi ci riferiamo?
2. Ipotizziamo implicitamente che l'intervistato abbia i genitori, potremmo essere anche offensivi;
3. Potrebbe essere che intervistato abbia genitori ma che non siano loro i percettori di reddito;
4. Il concetto di lavoro si divide tra condizione professionale ed il tipo di lavoro, ad esempio disoccupato ed imbianchino, tra l'altro uno stesso tipo di lavoro può variare nel suo significato a seconda di dove viene svolto, insomma all'unidimensionalità dell'intervistatore corrispondono più dimensioni dell'intervistato, chi risponde può rispondere ciò che vuole;
5. Le risposte aperte sono la trappola degli incapaci, mai e poi mai utilizzare risposte aperte a meno che non ne possiamo fare a meno, ovvero non sappiamo niente del fenomeno che indaghiamo magari qualcosa che non è mai stato indagato, comunque

anche in questo caso la consuetudine vuole che prima facciamo delle domande a risposta aperta come test su un campione molto basso e sulla base delle risposte impostiamo un buon questionario;

6. C'è il problema della codifica, è la stessa cosa che avviene quando si va alla camera di commercio e bisogna attribuire un codice per il settore di attività economica, quando l'imprenditore inserisce i dati all'interno dell'elaboratore elettronico e quest'ultimo deve attribuire un codice; in questo caso però il fenomeno, le attività economiche sono conosciute, figuriamoci nel caso in cui indagiamo un fenomeno che non è mai stato indagato, che non conosciamo, nel caso del settore di attività economica il problema principale era l'interpretazione della lingua Italiana. Nemmeno la domanda semiaperta è consigliabile, ad esempio con opzioni *A, B, C*, altro specificare..., quindi con delle opzioni a risposta chiusa ed una aperta, è proprio questo "altro specificare..." che crea problemi, e qui che si prova la propria ignoranza, in quanto inserire questa opzione vuol dire invitare il rispondente a non rispondere, tantissima gente risponde a quest'ultima opzione non perché ha altro da dire ma perché non si vuole esporre, e non si specifica assolutamente nulla, e tra l'altro cosa ci facciamo con questo "altro specificare...", serve più che altro a buttare i dati, lo mettiamo effettivamente solo quando sappiamo che c'è qualcos'altro da specificare ma non sappiamo cos'è e solo quando c'è un motivo, metterlo come atto democratico non serve a nulla. Non sono in ogni caso da considerare risposte aperte le richieste di dati quantitativi, anche se graficamente hanno la stessa forma di una domanda aperta poiché la risposta è precisa.

7. Crea stati di tensione nell'intervistato (imbarazzo a rivelare il lavoro dei propri genitori, può darsi che il padre faccia il becchino oppure genitori non siano in vita).

### **5.6.2 Come vanno formulate le domande sensibili**

Riprendendo l'esercitazione precedente, come va effettivamente chiesta la classe sociale dell'intervistato, la categoria della persona? Non c'è comunque un solo modo di redigere il questionario. Prima di tutto non si fa con una sola domanda ma con una batteria di domande; si predispone una domanda filtro - *i percettori di reddito sono all'interno del nucleo familiare?* Ma anche aver ricevuto lasciti ad esempio da un parente morto in America; solo se ha risposto positivamente alla domanda filtro si passa

alla domanda successiva - *quanti sono i percettori di reddito all'interno del nucleo familiare?* Per ciascun percettore di reddito poi facciamo una batteria di domande: *qual è la condizione professionale? Il tipo di lavoro? Il settore di attività economica?* Volendo possiamo anche chiedere la relazione di parentela se c'interessa (padre, figlio, madre, etc.) perché in effetti sapere chi fa l'impiegato o l'operaio non è tanto importante; la risposta aperta potrebbe essere utile solamente se vogliamo conoscere tutti i particolari della professione, ma in questo caso c'interessa solo la posizione professionale.

Un'altra cosa che possiamo fare è andare direttamente al dunque, chiedere il reddito, ma come si chiede il reddito? Un metodo utilizzato dall'Istat è quello d'inserire degli intervalli all'interno delle domande, ad esempio (0-1000, 1000-2000, etc.); l'Istat lo fa all'interno dell'indagine sui consumi ed anche se non è molto intelligente questo procedimento all'Istat non importa, in quanto interessano più che altro i consumi, la richiesta la inserisce per obbligo comunitario. Al massimo è un modo per vedere quanto i consumi si distanziano dal reddito, non si può consumare da 3000 ed avere un reddito di 1000.

Un altro metodo abbastanza paradossale è quello utilizzato dalla banca d'Italia, con il 740 alla mano il funzionario esegue tutti i calcoli ed estrapola una cifra, ma anche qui vi è un errore madornale, in quanto il 740 non considera il fatto che il reddito disponibile per la famiglia non è lo stesso di quello dichiarato poiché potremmo avere altre entrate non dichiarate. Tra l'altro bisogna considerare le deduzioni delle detrazioni dovute alla condizione professionale, ad esempio i lavoratori autonomi hanno la possibilità di scaricare numerose spese; insomma il 740 da un reddito lordo che non considera altre fonti di reddito.

Quindi in definitiva come si chiede il reddito? Prima di tutto introduciamo il concetto di *domande dirette* e *domande indirette*, queste ultime servono in qualche modo per deresponsabilizzare il giudizio, tipiche dei questionari psicologici. Quanto è alto il tuo reddito? Questa è una domanda diretta, invece una domanda indiretta chiede qualcos'altro per ottenere quello che c'interessa, il reddito.

La prima possibilità che ci si prospetta per chiedere il reddito sembra banale ma è quella che funziona di più - signora se lei disponesse di una quantità pari al suo reddito mensile, cosa acquisterebbe? Ci sarà un elenco dei prodotti di cui più o meno

conosciamo il valore, sommiamo i valori ed otteniamo il reddito, di solito il margine di errore è molto basso, intorno al 5-10% al massimo; 5-10% che sulla somma di tutti i redditi, ad esempio i redditi di Pescara, diventa approssimativamente lo 0,1% di errore. Potremmo fornire un elenco di prodotti tra cui scegliere, da inserire nella domanda ma senza il valore che potrebbe mettere in allarme il rispondente, oppure un elenco di prodotti a scelta del rispondente.

La seconda possibilità si basa su due domande che vanno messe ad una grande distanza l'una dall'altra, il rispondente deve dimenticare la prima una volta arrivati alla seconda - secondo lei, al giorno d'oggi, quanto servirebbe per poter vivere dignitosamente? E poi con grande distacco - date le sue disponibilità economiche, di quanto avrebbe bisogno lei per poter vivere dignitosamente? Per semplice differenza scopriamo il reddito, ma dov'è che possono avere delle lacune? Sui rispondenti estremamente poveri ed estremamente ricchi ovvero sulle code della distribuzione dei redditi, dato che questa batteria di domande ipotizza che vi sia una differenza negativa, un differenziale da applicare tra la prima e la seconda domanda, vale quindi per individui che si trovano sulla parte centrale della distribuzione dei redditi;

### **5.6.3 Le domande randomizzate o casualizzate**

Supponiamo che la domanda sia a risposta chiusissima e con due sole modalità, una domanda vera che c'interessa – ti droghi? Ed una domanda totalmente astrusa ed innocua con lo stesso elenco di risposte si/no - sei femmina? Sei maschio? E quindi di cui conosciamo la frequenza di risposta dei sì e dei no, poi andiamo dall'intervistato, gli consegniamo un qualsiasi meccanismo di estrazione di numeri casuali, ad esempio un dado o una moneta, e gli diciamo che se esce pari deve rispondere alla domanda vera se esce dispari deve rispondere a quella falsa, e saprà solo lui se ha risposto a quella vera o a quella falsa. Questa metodologia garantisce contemporaneamente la privacy dell'intervistato e la qualità statistica degli stimatori, a noi consegnerà solo il si/no, non sapremo mai a quale domanda ha risposto, poiché a noi non interessa se lui si droga ma quanti si drogano, quindi a livello di totale possiamo ritornare indietro.

$$NR\% = \pi \cdot NRC\% + (1 - \pi)NRS\%$$

Dove  $NR$  è la percentuale di risposte positive,  $\pi$  è la probabilità del meccanismo di estrazione, sarà 0.5,  $NRC$  è il numero percentuale di risposte positive alla domanda fasulla,  $(1-\pi)$  è la probabilità che si risponda alla domanda vera, per il numero percentuale di risposte positive alla domanda vera  $NRS$  ;

$$NRS\% = \frac{NR\% - \pi \cdot NRC\%}{1 - \pi}$$

Cioè 0.5 risponderanno alla domanda falsa e 0.5 volte risponderanno all'altra, e quindi a livello di totale il numero delle nostre risposte sarà una media ponderata delle probabilità di risposta all'una o all'altra, poiché  $NR$  la conosciamo ed  $NRC$  la costruiamo proprio in modo da conoscere le frequenze a livello di totale,  $\pi$  la conosciamo, e quindi esplicitando rispetto a  $NRS$  troviamo il numero totale di persone che fanno uso di droga. Questo lo possiamo fare per la legge dei grandi numeri, l'unica fonte di errore quindi può venire solo dal fatto che le risposte possono discostarsi dalla probabilità di 0.5 tra la risposta alla domanda vera e la risposta a quella falsa. In ogni caso questo procedimento viene usato di solito per le indagini su individui e famiglie.

#### 5.6.4 Altre tipologie di domande

Esistono domande a *risposta singola* ed a *risposta multipla* ovvero in cui si deve mettere una sola risposta o dove metterne più di una, concetto che deve essere ben specificato in quanto di solito il rispondente non conosce il numero di risposte che è possibile inserire; questo problema si risolve di solito inserendo un esempio all'interno del questionario - quali prodotti usi? Bisogna specificare che si può rispondere a più di un prodotto.

Esistono *domande primarie* da cui dipende la possibilità di poter fare altre domande; esistono *domande filtro*, domande primarie molto particolari poiché a seconda della risposta, sì o no, si annullano intere sezioni di questionario, è una domanda primaria molto importante poiché il questionario cambia totalmente.

Bisogna tenere sempre bene in considerazione che tutto quello che abbiamo detto finora vale solo se non abbiamo a che fare con fenomeni o argomenti molto delicati; nel caso invece abbiamo a che fare con questo tipo di fenomeni dobbiamo fare in modo di far sentire l'intervistato sempre più a suo agio e che la sua risposta sarà sempre anonima.

### **5.6.5 Domande per le imprese**

Per le indagini sulle imprese tutto ciò che abbiamo detto si semplifica ampiamente, poiché un'indagine sulle imprese è un elenco di variabili, di dati quantitativi, e quindi i questionari sono molto semplificati, stringati, l'unico motivo di difficoltà per quanto riguarda i questionari alle imprese è che per calcolare i pochi dati richiesti s'impiega molto tempo, anche se chiediamo poche cose è comunque molto oneroso calcolarle, non è un problema di questionario, più che altro c'è bisogno di un'adeguata istruzione alla compilazione; una determinata variabile quantitativa deve essere ben definita, poiché magari ciascuna azienda ha una concezione diversa della stessa variabile, ad esempio quando si chiede il fatturato, va specificato tutto quello che è contenuto all'interno di questa variabile. Naturalmente solo se l'indagine non è diretta all'imprenditore, in questo caso rientriamo nell'analisi fatta precedentemente.

### **5.7 Il piano di compatibilità**

Fare un questionario comporta una difficoltà aggiuntiva rispetto a tutte quelle menzionate finora, per fare un questionario non basta trascrivere le risposte e stamparle ma, lavoro grande, pesante e tedioso, al nostro questionario dobbiamo associare quello che viene chiamato *piano di compatibilità* ovvero un elenco il più completo ed esaustivo possibile di tutte le regole che devono essere soddisfatte dalle risposte all'interno del questionario, ad esempio – età: 5 anni, professione: pensionato - chiaramente non c'è compatibilità, quindi è un insieme di regole che devono essere necessariamente rispettate dai dati; tra le regole bisogna distinguere tra quelle *deterministiche* e *non-deterministiche* e tra *eventi impossibili* ed *eventi improbabili*, ad esempio che un novantenne potrebbe essere sposato con una diciottenne, è improbabile ma comunque possibile. Di regole simili in un questionario ce ne sono milioni.

In realtà il questionario su carta è solo una forma di rappresentazione, di solito viene sempre rappresentato in modalità grafica, cioè domande e risposte sono degli oggetti su un disegno, la prima cosa da fare è tracciare una freccia tra le domande che sono collegate tra di loro, ad esempio sesso, professione, età, titolo di studio, etc. (ciò che è connesso o collegabile con...). Quindi di solito tutto il questionario viene elaborato a blocchi, come per le sezioni, omogenee al loro interno impattanti su tutto il resto, in seguito vengono fatti dei grafici all'interno di ciascuna sezione per vedere se si è fatto in modo che la sezione possa impattare solo al suo interno. Successivamente tramite un'equazione matematica, dopo aver fatto i grafici, analizziamo quali dati sono collegati tra loro ricaviamo la regola di compatibilità.

Questo insieme di vincoli, di regole cui devono sottostare i dati sono uno strumento fondamentale per il *controllo e la correzione dati*, nel senso che un dato per essere buono deve rispettare tutte queste regole, poi può anche essere sottoposto a menzogna, anche se questo problema viene risolto tramite delle *domande di controllo*, molto numerose all'interno di un questionario.

Un sottoinsieme di queste regole del piano di compatibilità generale viene di solito applicato già in fase di caricamento dati, di solito questi software non hanno tutto il piano di compatibilità, e nel sottoinsieme mettiamo solo le regole deterministiche a livello di caricamento dati, ne mettiamo circa un centinaio, la persona che carica i dati non se ne deve neanche accorgere e quindi solo quelle veramente importanti, perché se viene bloccato colui che carica i dati bisogna rivedere il questionario e diventa laborioso il processo; infine questo milione di regole serve alla fine per correggere la stima.

Il lavoro più complesso è proprio il *controllo e la correzione dati*, ecco il motivo per cui passa molto tempo prima della pubblicazione di un censimento.