



## Nota Metodologica

## IL QUESTIONARIO

La stesura del questionario richiede una particolare attenzione nella scelta, nella compilazione e nella successione degli ITEMS che formano lo strumento d'indagine. Per la formulazione del questionario della R.G.S. si è seguito un procedimento logico che, partendo dall'individuazione delle aree problematiche, è approdato alla tipologia delle domande da utilizzare.

Le aree problematiche definite sono le seguenti:

1. Percezione del percorso formativo del campione contattato
2. Percezione e atteggiamento nei confronti del mercato del lavoro
3. Livello d'informazione, mezzi di comunicazione e percezione del Brand di alcune tra le maggiori aziende presenti sul mercato del lavoro

Inoltre si è destinata un'area del questionario ai dati di base (demografici) del campione.

Al fine di ottenere un insieme di informazioni gestibili attraverso una matrice dati, l'equipe dell'RGS ha utilizzato, nella formulazione del questionario, ITEMS con modalità di risposta chiusa e pre-codificata

Inoltre, le domande che hanno composto il questionario sono state principalmente del tipo standard e a batteria

Entrando nello specifico, l'equipe ha utilizzato questa formula per tre motivi, essenzialmente legati alla complessità del questionario:

1. Per utilizzare l'auto-somministrazione dello strumento d'indagine, tecnica che permette di evitare la perturbazione dell'oggetto d'indagine.
2. Per velocizzare la compilazione da parte dell'intervistatore di un questionario molto complesso
3. Per evitare un eccessivo spazio lasciato all'interpretazione sia dell'intervistato che del ricercatore.

Terminata la stesura del questionario e prima di procedere alla somministrazione vera e propria, si è testato (durante il J.M. di Torino tenutosi nel 2004) lo strumento d'indagine per non incorrere in errori d'interpretazione, eventualmente dovuti a:

1. Domande mal poste
2. Respons set
3. Curvilinearità degli items
4. Inutilità del dato reperito

Corretto il questionario, la ricerca ha avuto il suo inizio.

## CAMPIONAMENTO

La società è un insieme di realtà molto complesse.

Lo studio di queste realtà comporta ancora oggi numerose problematiche metodologiche.

In effetti, già il fatto che il ricercatore è parte integrante dell'oggetto studiato pone alla comunità scientifica due dilemmi da superare.

Il primo è la RETROAZIONE, ovvero la concreta possibilità che l'oggetto di studio influenzi il ricercatore (fenomeno frequente soprattutto nelle ricerche macrosociali).

Il secondo è la perturbazione dell'oggetto, cioè la possibilità che sia il ricercatore ad influenzare l'oggetto d'indagine.

Questi dilemmi sono però preceduti da un altro aspetto, vale a dire la possibilità o meno di avere delle rilevazioni complete. Per rilevazioni complete si intende lo studio dell'universo statistico di riferimento (cioè l'insieme delle unità che compongono l'oggetto di studio).

Nelle ricerche macrosociali, come nel nostro caso, la rilevazione completa comporta enormi costi, sia in termini di tempo che in termini monetari. Un esempio concreto è riscontrabile nel censimento della popolazione italiana che l'ISTAT effettua periodicamente: questo, come è a tutti noto, richiede elevati finanziamenti statali e periodi di preparazione, somministrazione, inserimento ed elaborazione molto lunghi.

Per tali motivi, nella quasi totalità delle ricerche socio-statistiche, si usufruisce della possibilità di procedere a rilevazioni campionarie. Per campione s'intende una porzione della popolazione.

E' chiaro che l'attendibilità del dato finale è intimamente connessa al grado di rappresentatività del campione rispetto all'universo di riferimento.

Si dice rappresentativo quel campione che costituisce un'immagine ridotta ma fedele della popolazione da cui proviene.

In termini statistici, tale rappresentatività è determinata dall'assenza di distorsioni sistematiche. Questo costringe ad adottare rigide metodologie d'estrazione che garantiscono a ciascuna unità, dell'insieme di partenza, la stessa possibilità d'essere estratta.

E' opportuno sottolineare che un campione non potrà mai raggiungere la rappresentatività assoluta; proprio per questo motivo, ogni volta che si parlerà di rappresentatività si dovrà far riferimento alle frequenze e ai parametri scelti.

Nel nostro caso, il campione adottato è stato estratto rispettando le proporzioni delle variabili **SESSO, AREA GEOGRAFICA, LAUREA ACQUISITA** riferite all'universo statistico.

La tecnica utilizzata per estrarre il campione verrà descritta in seguito.

L'aspetto che ora occorre affrontare è quella della numerosità campionaria, ovvero il numero di unità che devono formare il campione per renderlo valido.

Dal momento che non conosciamo l'universo statistico abbiamo la necessità di utilizzare, per la nostra indagine, delle stime.

Per stima s'intende la possibilità di prevedere la presenza-assenza di un carattere fissando la probabilità d'errore (**INTERVALLO DIFIDUCIA**). Ponendo come base tale stima, è possibile determinare la numerosità campionaria attraverso un processo controllabile.

Riguardo alla dimensione del campione occorre anzitutto procedere accertando in anticipo quanto siamo disposti a sbagliare.

Nel caso in cui si ipotizzi di accettare un errore del 5%, occorrerà adottare la formula per determinare la numerosità del campione attraverso la stima di una frequenza con un intervallo di fiducia pari al 95%.

E' chiaro che ogni qualvolta si affronta un'operazione di stima, occorre avere dei punti fermi (dati, frequenze, parametri conosciuti). Partendo da questi, avremo la possibilità di stimare, quindi di determinare quante unità ci occorreranno per far sì che le nostre affermazioni scaturite dalle indagini siano attendibili nel 95% dei casi.

In questa ricerca, il punto di partenza è la variabile sesso (con la sua frequenza distribuita nell'universo che abbiamo conosciuto attraverso i dati del MURST).

Per maggiori chiarimenti su quanto detto, è consigliabile approfondire l'argomento **STIMA PUNTUALE E PER INTERVALLO** alla base del procedimento di determinazione della numerosità campionaria.

Determinata la numerosità campionaria, per ottimizzare la qualità dei risultati, si è proceduto all'estrazione di un campione stratificato; esso consiste nell'individuare degli strati della popolazione rispetto ad uno o più caratteri noti.

Come detto in precedenza, l'equipe della RGS ha individuato come caratteri noti e probanti il sesso, l'area geografica e il tipo di laurea acquisita.

Rispettando le proporzioni di questi caratteri della popolazione, si è frazionato il campione in sottocampioni, e in ciascuno di essi si è proceduto a contattare i neolaureati attraverso una somministrazione sistematica svoltasi durante i J.M. che si sono tenuti lungo tutto il territorio nazionale.

Un aspetto fondamentale per determinare la correttezza di una survey è la procedura di reperimento del dato.

Trattandosi di una ricerca basata su dati primari (attraverso l'utilizzo del questionario come strumento d'indagine) occorre evidenziare le modalità di somministrazione con le quali l'equipe ha proceduto.

### SOMMINISTRAZIONE

Come già detto, l'RGS è una survey che si pone come obiettivo l'analisi dei fenomeni su base nazionale. Per tale motivo il reperimento dati è stato suddiviso in più fasi determinate dall'opportunità di utilizzare momenti di alta concentrazione dell'universo di riferimento.

Si è proceduto in questo modo per avere la possibilità di contattare la totalità del campione in un periodo di tempo contingentato.

Il contingentamento è stato necessario per la particolarità stessa del fenomeno indagato.

In effetti, avendo come obiettivo quello di esplorare l'atteggiamento dei neolaureati nei confronti del mondo del lavoro, e per far sì che tale risultato non fosse obsoleto già al termine dell'indagine, si è reso necessario svolgere la ricerca nel più breve tempo possibile; poiché il campione da contattare consisteva in un universo sommerso (nota: i neolaureati non hanno momenti aggregativi che permettono facilmente la rilevazione del dato, come accade invece per gli studenti durante l'orario didattico, e per gli occupati durante l'orario lavorativo; per questo motivo essi sono più difficili da rintracciare), si è reso necessario ricercare momenti di emersione dell'universo stesso.

L'opportunità si è presentata all'interno dei Job Meeting.

I J.M. sono fiere presenti in Italia da oltre 15 anni che consentono ai neolaureati di contattare direttamente un certo numero di aziende.

Queste occasioni d'incontro hanno consentito all'equipe una somministrazione che avrebbe altrimenti comportato tempi estremamente più lunghi.

La somministrazione all'interno di ciascun J.M. è avvenuta seguendo un iter standard, tale da rendere omogenea la rilevazione del dato; all'interno delle università dove avvenivano i J.M., la RGS disponeva di uno spazio apposito dove poter effettuare la rilevazione.

L'equipe, formata da due ricercatori e due operatori (formati ad hoc), somministrava i questionari ai ragazzi inviati dalla reception del J.M., dopo che la stessa ne aveva constatato l'appartenenza all'universo di riferimento.

Illustrata la finalità della ricerca e consegnato il questionario, i ragazzi compilavano lo stesso in un apposito spazio, impiegando un tempo medio di 30 minuti.

Terminata la compilazione, i ragazzi consegnavano il questionario agli operatori, i quali, controllata la correttezza e completezza della compilazione stessa, procedevano all'archiviazione.

Nel contempo, sia gli operatori che i ricercatori erano a disposizione dei ragazzi per eventuali chiarimenti.

Terminata la somministrazione e l'ordinamento dei questionari, l'equipe della RGS ha proceduto all'inserimento dei dati su supporto magnetico avvalendosi di un software dedicato.

### PULIZIA DEI DATI

Conclusa la fase d'inserimento dati che si è riproposta dopo ogni J.M., si è avviata la fase di pulizia dei dati: questa fase consiste nel controllo della correttezza dell'inserimento. Infatti, durante quest'ultimo, accade spesso di incorrere in errori dovuti all'elevato numero di questionari trattati.

Questo controllo viene effettuato attraverso due passaggi:

- il primo si fonda su un'analisi di frequenza finalizzata all'individuazione dei "valori selvaggi": es. (M/F), valori che non sono stati previsti in nessuna modalità della variabile, o che si rilevano poco plausibili.
- il secondo verte sull'individuazione del numero del questionario corrispondente all'errore inserito (ovviamente i questionari sono numerati in ordine progressivo) e, controllato sullo stesso il dato esatto, si procede alla correzione della matrice.

Prima di inoltrarsi nell'analisi statistica dei dati, occorre verificare che i dati ottenuti dall'indagine risultino ottimali; questo significa smussare, ove possibile, l'eccessiva presenza di frequenze basse. Tale procedura implica l'osservazione, mediante analisi monovariata, delle frequenze ottenute. Nel caso in cui si presentino frequenze basse all'interno di variabili che permettono l'accorpamento logico delle modalità, si interverrà con una ricodifica:

Es: Molto, abbastanza, poco, per nulla.

Tale intervento permetterà di ottenere un livello più alto di significatività della successiva analisi bivariata e soprattutto della analisi multivariata.

Fatto ciò, si procederà all'analisi dei dati.

L'analisi monovariata è la forma più semplice di rappresentazione.

Come si evince dal termine stesso, consiste nell'analisi di una variabile alla volta, senza relazionarla alle altre variabili presenti in matrice.

L'analisi monovariata, oltre a ricoprire le funzioni imprescindibili prima citate (controllo, pulizia e aggregazione dei dati) è finalizzata ad illustrare al fruitore le piattaforme, teoriche ed empiriche, sulle quali verranno strutturate le analisi successive.

E' necessario sottolineare che la maggior parte delle variabili ottenute è di natura "categoriale" (sorte cioè usufruendo di scale nominale e ordinali) e per questo motivo definite più correttamente "mutabili", es... Questa condizione ha comportato l'utilizzo di indici statistici appropriati.

#### ANALISI BIVARIATA

Al termine dell'osservazione mono-variata dei dati ha inizio l'analisi vera e propria.

Infatti, i dati archiviati raccontano eventuali realtà non tanto attraverso la distribuzione delle singole frequenze quanto attraverso il loro incrocio.

L'incrocio dei dati può essere effettuato a due a due (analisi bivariata) o fra tre variabili e più (analisi multivariata). In questo paragrafo affronteremo il primo dei due casi, l'analisi bivariata.

L'incrocio tra due variabili, nel nostro caso quasi sempre di natura categoriale, è chiamato tabella di contingenza e così noi l'abbiamo definita in tutto il rapporto.

Occorre sottolineare che statisticamente nell'analisi bivariata intervengono maggiori elementi e indici rispetto all'analisi di frequenza (a seconda della natura della relazione). Questo poiché la complessità dell'approccio aumenta in modo direttamente proporzionale al valore esplicativo dell'intervento d'indagine.

Per questo, nel corso dell'analisi sono stati utilizzati diversi coefficienti statistici per determinare la significatività o meno delle tabelle di contingenza (chi quadro, phi, coefficienti di cograduazione, etc...).

Per significatività si intende il caso in cui in una tabella di contingenza le variabili che la costituiscono siano legate tra loro da una qualche relazione che determina il cambiamento dell'una in relazione al modificarsi dell'altra.

Nell'intero rapporto sono state considerate e studiate in profondità solo le tabelle significative (a differenza delle frequenze che compaiono nella loro completezza). Le altre non sono state considerate affatto data la loro inutilità euristica.

#### ANALISI MULTIVARIATA

Se, come detto in precedenza, l'analisi monovariata rappresenta la forma più semplice di rappresentazione dei dati, l'analisi multivariata è la più complessa.

Ancora più dell'analisi bivariata, essa prevede una molteplicità di tecniche relative alla natura dei dati in possesso del ricercatore.

Ma che cosa è l'analisi multivariata?

E' una tecnica di studio statistico delle variabili. Questo metodo permette di superare lo scoglio presente nelle prime due (mono e bivariata) che consiste nel non poter analizzare simultaneamente le influenze e le relazioni che ogni variabile ha sulle restanti.

Con l'analisi multivariata è possibile ottenere dei modelli che permettono di non perdere di vista "l'insieme".

Quindi è possibile considerare questa tecnica come una sintesi dei caratteri rilevati.

Per fare un esempio figurativo, è come se, dovendo rappresentare una persona che suona uno strumento utilizzando tempi swing o bossanova, reinterpretando standards musicali, sintetizzassimo il tutto con il termine jazzista.

La dimensione ottenuta non sarebbe altro che l'influenza che intercorre tra i caratteri in nostro possesso.

Come prima detto, il ricercatore, per ottenere queste dimensioni, deve effettuare una scelta delle tecniche da utilizzare.

Nel caso specifico, l'analisi che è stata utilizzata, per studiare i dati ottenuti dalla RGS è del tipo a CORRISPONDENZE MULTIPLE, data la natura categoriale delle variabili in possesso.

Detta anche FATTORIALE, l'analisi a corrispondenze multiple elabora una particolare tabella di contingenza mediante l'utilizzo del CHI QUADRO.

L'estrazione dei fattori (dimensioni sintetiche) si ricava dal calcolo dell'inerzia delle variabili attive. Per variabili attive si intendono quelle variabili che concorrono direttamente, per scelta del ricercatore, all'individuazione degli assi fattoriali.

Inoltre, per meglio esplorare, descrivere e sintetizzare i fattori ottenuti, vengono utilizzate altre variabili, chiamate supplementari o illustrative, che permettono una migliore interpretazione dello spazio dimensionale (senza contribuire direttamente all'estrazione).

Tendenzialmente, le variabili attive immesse nell'analisi corrispondono alle variabili assunte in ipotesi come dipendenti, mentre quelle illustrative sono le variabili assunte in ipotesi come indipendenti.

Altro tipo di analisi multivariata utilizzata nello studio è la loglineare.

Dopo aver esplorato lo scenario ed essere venuti a conoscenza delle relazioni tra variabili attive e illustrative, con l'analisi loglineare si è in grado di comprendere il tipo di interazione che intercorre tra le stesse.

Il concetto alla base dell'analisi loglineare consiste nel passare dalle frequenze ai logaritmi delle frequenze. Questo passaggio comporta la possibilità di trattare delle variabili categoriali come fossero metriche e quindi permette l'analisi di regressione anche su variabili che altrimenti non la permetterebbero. In particolare, l'equazione di regressione invece che avere come unità gli individui è costituita dalle celle (logaritmo della frequenza di cella).