Capitolo 2 – Disponibilità e produzione delle informazioni statistiche

Concetti generali

Nella sua attività l'impresa ha bisogno di reperire informazioni (dati), classificabili nel seguente modo:

- provenienza:
  - *dati interni* (provenienti dall'impresa)
  - *dati esterni* (provenienti da soggetti terzi)
- significatività:
  - *dati primari* (quasi esclusivamente esterni, commissionati per finalità specifiche: indagini preso esperti, focus group, studi sperimentali, indagini dirette con comunicazione, tecniche di osservazione)
  - *dati secondari* (esterni come i dati ISTAT ed interni come bilanci, contabilità dei costi e controllo qualità collaterali alla normale attività di impresa). Provengono delle fonti statistiche, termine con il quale si può indicare un ente appartenente alla statistica ufficiale (Sistema Statistico Nazionale), l'indagine/rilevazione specifica o la pubblicazione/mezzo di diffusione dei dati
  - *dati per analogia* (casi di studio, simulazioni, previsioni)

C'è una distinzione (anche se spesso possono coincidere) fra i concetti di:

- unità di rilevazione: le unità che compongono il campione, dalle quali si vuole ottenere informazioni sulle unità di analisi
- unità di analisi (unità statistica):gli individui/entità sui quali vengono rilevate le singole informazioni

Le informazioni statistiche, i dati, si classificano in:

- dato elementare (*dato statistico*): è una misurazione, un dato (valore/modalità), che si riferisce al singolo carattere osservato in una singola unità statistica appartenente ad una popolazione/campione
- microdato *(record*): insieme dei dati elementari rilevati su di una singola unità di analisi (es. età, residenza, genere, ecc)
- macrodato (*dato aggregato* o statistica): deriva dalle aggregazioni di dati elementari, è una funzione di dati elementari nonché un modo di presentare le informazioni contenute nei dati tramite indicatori numerici sintetici
- metadato: è un dato che descrive e definisce dati statistici, con la funzione di:
  - fornire la definizione delle unità di analisi e dei caratteri osservati, nonché la loro significatività
  - specificare la classificazione adottata e la qualità dei dati ottenuti
  - descrivere la metodologia impiegata per giungere alla produzione di statistiche (trasparenza)

Dati sperimentali

I dati sperimentali sono generati da uno *studio sperimentale* (esperimento). Un esperimento è tale quando le unità sperimentali sono sottoposte ad un trattamento allo scopo di misurarne l'effetto su di esse tramite la cosiddetta *variabile risposta* (variabile risultato). I dati sperimentali sono i dati ideali per la ricerca esplicativa (causale).

Gli studi sperimentali hanno una propria nomenclatura:

- unità sperimentale: svolge lo stesso ruolo dell'unità statistica nelle indagini statistiche
- variabile risposta: la variabile che misura l'influenza del trattamento sulle unità sperimentali (es. sintomi)
- variabile esplicativa: la variabile della quale si vogliono misurare gli effetti sulla variabile risposta (es. farmaco A)
- trattamento: la condizione sperimentale applicata alle unità sperimentali (es. farmaco A o placebo)

La più importante differenza fra *dati sperimentali e dati osservazionali* è proprio quella che nei secondi la rilevazione vuole misurare i fenomeni senza intervenire a modificare lo stato delle variabili.

Qualità della statistica e statistica ufficiale

L'informazione statistica è caratterizzata da una buona o da una cattiva qualità. I sei criteri guida per una statistica di elevata qualità, detti *dimensioni della qualità*, sono:

- rilevanza: capacità di rispondere alle esigenze
- accuratezza: vicinanza tra stima e valore vero del carattere misurato
- puntualità e tempestività: corrispondenza fra la data di diffusione e quella prefissata (puntualità) e differenza fra momento di pubblicazione e momento a cui la statistica fa riferimento (tempestività)

- accessibilità e chiarezza: facilità di ottenimento delle informazioni (accessibilità) e capacità di rendere comprensibili e interpretabili le statistiche (chiarezza)
- comparabilità: la possibilità di confrontare due statistiche dello stesso fenomeno in momenti differenti, effettuate con lo stesso metodo di rilevazione
- coerenza: l'informazione desumibile dà una visione univoca del fenomeno

C'è chiaramente un trade-off fra alcune dimensioni della qualità, ad esempio rilevanza-coerenza, rilevanza-comparabilità e comparabilità temporale-comparabilità spaziale.

La produzione di dati ad hoc: le indagini campionarie

Se l'impresa ha esigenze conoscitive non soddisfatte dalla statistica ufficiale o da altre indagini statistiche, è necessario condurre un'indagine ad hoc. Essa può essere:

- completa: effettuata su tutta la popolazione obiettivo (*censimento*)
- campionaria: effettuata su un sottoinsieme rappresentativo della *popolazione obiettivo* detto *campione*

Lo strumento con il quale sono raccolte le informazioni è il questionario, somministrato per via orale, cartacea, informatica.

Punto di partenza per effettuare il campionamento è avere a disposizione un disegno di campionamento (linea metodologica) che individui una lista di campionamento, al cui interno siano ordinate tutte le unità che fanno parte della popolazione obiettivo. Riuscire a stilare una lista di campionamento è nella pratica un'operazione molto difficoltosa, soggetta spesso ad un'approssimazione che fa sì che si parli di popolazione di selezione come una approssimazione della popolazione obiettivo.

Tuttavia la popolazione di selezione non rappresenta ancora la popolazione alla quale si può univocamente riferire la formazione del campione, perché alcune unità potrebbero essere irreperibili o rifiutarsi di rispondere (mancata risposta o mancata risposta totale). Così da una parte al *campione teorico* si affianca il suo sottoinsieme detto *campione effettivo*, e la popolazione si dice popolazione di indagine.

Obiettivo della indagine campionaria è quella di *stimare alcuni parametri della popolazione*. La differenza fra il valore ottenuto dello stimatore del parametro e il valore effettivo nella popolazione è detta errore statistico.

Nell'indagine campionaria l'errore statistico, insito nella scelta di rilevare il carattere in un sottoinsieme della popolazione, è detto *errore campionario*, ed è stimabile se il campione è selezionato con metodi probabilistici.

La *numerosità campionaria* n è legata alla varianza campionaria da una relazione di proporzionalità inversa (compare infatti al suo denominatore). Se l'obiettivo per una statistica qualitativamente affidabile è ottenere una bassa varianza, è naturale che la scelta della numerosità campionaria debba orientarsi sulla massimizzazione del trade-off fra numerosità campionaria e costi connessi alla produzione dell'indagine.

Tipologie di indagini

Le indagini si distinguono in:

- indagini *cross-section* o trasversali: le unità statistiche sono osservate allo stesso tempo t, metodologia utile per una comparazione fra gruppi
- indagini *panel* o longitudinali: le unità statistiche sono osservate in più momenti, metodologia utile per un'analisi dei cambiamenti nel tempo delle singole unità

Un aggregato di rilevazioni ordinate rispetto al tempo è detto serie storica. Si tratta generalmente di macrodati.

Tipologie di campionamento

A seconda del metodo di campionamento scelto si distingue in:

- campioni probabilistici: ogni unità della popolazione ha una probabilità nota di essere estratta:
  - campionamento *casuale semplice*
  - campionamento *sistematico*
  - campionamento *stratificato*
  - campionamento *a grappolo*
- campioni non probabilisti: non è possibile calcolare la probabilità di un'unità della popolazione di essere estratta, perché differente fra le unità secondo ragioni di comodità, praticità, costo:
  - campionamento *di comodo*: il campione è scelto senza alcuna pretesa di rappresentatività della popolazione
  - campionamento *a scelta ragionata*: il campione è scelto selezionando "manualmente" le unità che si ritengono rappresentative (anche detto dei "testimoni privilegiati")
  - campionamento *per quota*: il campione è scelto in modo tale da mantenere una proporzione fissa di questo rispetto all'universo, l'intervistatore ha piena discrezione su chi scegliere

- campionamento *a valanga*: ogni individuo fornisce informazioni sui caratteri suoi e di altre unità statistiche, il numero di unità del campione cresce esponenzialmente

Nel campionamento casuale semplice il campione viene estratto dalla lista di campionamento mediante estrazione con ripetizione o senza ripetizione. Si definisce *frazione di campionamento* f il rapporto fra la numerosità del campione n e la numerosità della popolazione N, che nel ccs è anche la probabilità associata ad ogni unità statistica di far parte del campione.

La stima puntuale della *media della popolazione* si effettua con il corrispondente *stimatore corretto* ovvero la media campionaria: $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$.

La stima puntuale dell'errore standard (*deviazione standard*), se è ignota la varianza σ_0^2 della popolazione, si effettua con lo stimatore corretto dell'errore standard della popolazione, ovvero l'errore standard campionario:

$S_X = \sqrt{(1-f)\frac{S_X^2}{n}}$ utilizzando lo stimatore varianza campionaria ovvero $S_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$.

Si definisce $(1-f)$ *fattore di correzione per popolazioni finite*, utilizzato per correggere il ccs senza reimmissione (senza ripetizione). In caso di ccs con ripetizione o senza ripetizione ma da popolazioni sufficientemente grandi, esso è pari a 1.

In pratica il ccs è poco utilizzato, perché richiede una lista di campionamento completa, alti costi (dispersione territoriale) e trascura informazioni a priori che caratterizzano la popolazione. I suoi vantaggi rimangono la facilità teorica di applicazione a qualsiasi oggetto di indagine.

Nel campionamento sistematico si introduce appunto un "sistema" per effettuare il campionamento e non ci si affida ad un'estrazione. Esso avviene mediante l'ordinamento delle unità all'interno della lista di campionamento secondo un determinato carattere. Si procede poi ad un'unica estrazione casuale di un'unità, dalla quale, secondo un predefinito *passo di campionamento*, si selezionano le successive. Le stime dei parametri della popolazione si effettuano come esposto in precedenza.

Nel campionamento stratificato si utilizzano le informazioni a priori sulle caratteristiche interne della popolazione per suddividerla in *classi* (strati) da cui estrarre con ccs, con l'obiettivo di *ridurre la variabilità entro gli strati* ed ottenere così stime più precise dei parametri, aiutati in ciò dalle correlazioni fra il carattere osservato e il carattere su cui si basa la stratificazione. È necessario operare una scelta riguardante il numero di strati e la frazione di campionamento in ognuno.

In base alla frazione di campionamento all'interno degli strati si distingue in:

- campionamento stratificato proporzionale: stessa frazione di campionamento per ogni strato
- campionamento stratificato non proporzionale: frazioni diverse, usato per strati non omogenei fra loro in termini di numerosità di unità e o di variabilità interna del carattere

Lo stimatore per la media della popolazione in caso di campionamento stratificato di H strati è la *media ponderata delle medie campionarie* ottenute per ciascuno strato:

$\bar{X}_{ST} = \sum_{h=1}^{H} W_h \bar{X}_h = \sum_{h=1}^{H} W_h * \frac{1}{n_h}\sum_{i}^{n_h} X_{hi}$ dove $W_h = \frac{N_h}{N}$ è la *proporzione della popolazione nello strato h*

Nel campionamento a grappoli la popolazione viene divisa in *grappoli* (cluster) in base a esigenze organizzative e/o strutturali. I singoli cluster sono, fra loro, molto simili, e in un certo senso potrebbero già essere un sottoinsieme rappresentativo della popolazione: all'interno del cluster le unità statistiche sono molto differenti fra loro in base al carattere oggetto dell'indagine. Tuttavia vengono estratti in maniera casuale più cluster, e faranno parte del campione tutte le unità statistiche appartenenti ai cluster estratti.

Si può anche effettuare un campionamento a grappoli a due stadi: dai cluster estratti si estraggono casualmente dei campioni le cui unità statistiche faranno parte del campione finale utilizzato per l'indagine.

Strati e cluster si possono pensare come raggruppamenti antitetici:

- a livello interno lo strato è omogeneo mentre il cluster no
- a livello esterno gli strati sono diversi fra loro mentre i cluster sono omogenei
- lo strato presuppone la conoscenza a priori della caratteristica di stratificazione (che altrimenti va indagata preliminarmente) mentre i cluster sono spesso unità organizzative preesistenti che semplificano la raccolta delle liste di campionamento

In conclusione i campionamenti di tipo probabilistico consentono una *valutazione oggettiva degli stimatori* in termini di precisione e quindi di qualità dei risultati (e dunque estenderne i risultati alla popolazione), mentre i campionamenti di tipo non probabilistico sono valutabili solo in base a considerazioni soggettive.

Valutazione dei risultati di un'indagine campionaria

Si era già definito l'errore statistico, o *errore totale*, la distanza del vero valore del parametro dal valore assunto dal suo stimatore, distanza che fornisce un'indicazione sulla accuratezza della rilevazione. Tale errore si divide in due fattispecie di errore distinte:

- errore campionario: l'errore non eliminabile derivate dall'estrazione di un campione
- errore non campionario: dovuto agli aspetti organizzativi della rilevazione e legato alle dimensioni del campione, può derivare da:
  - errori di copertura: la lista di campionamento non individua efficacemente la popolazione obiettivo
  - errori da mancate risposte: generano non corrispondenza fra tra popolazione obiettivo e popolazione effettiva
  - errori di misurazione: errori più "tecnici" (quesiti formulati male, errori dell'intervistatore, ecc)

Capitolo 3 – Interpretazione e comparazione dei dati riferiti a fenomeni aziendali

Rapporti statistici

Il modo più intuitivo per confrontare come si distribuisce un fenomeno rispetto a determinati caratteri è quello del *rapporto*. Si distingue fra:

- *rapporto di composizione*: il rapporto fra un'intensità (frequenza) parziale e l'intensità (frequenza) totale
- *rapporto di coesistenza*: il rapporto che misura lo "squilibrio" fra due intensità coesistenti o in corrispondenza della stessa modalità (rapporto esportazioni/importazioni)
- *rapporto di densità*: il rapporto fra l'intensità di un fenomeno e un'aggregazione statistica dimensionalmente differente (ricchezza pro-capite, spesa per famiglie)
- *rapporto di derivazione*: il rapporto fra un'intensità di un fenomeno e l'intensità di un fenomeno generante o presupposto logico (quozienti demografici: rapporto nati/popolazione, ecc)

I numeri indici semplici: definizioni

I *numeri indici* sono un tipo particolare di rapporto statistico che misurano le variazioni di un fenomeno quantitativo nel tempo (numeri indici temporali) e nello spazio (numeri indici spaziali). Si tratteranno quelli del primo tipo.

Si distingue fra:

- *numeri indici semplici* (elementari): misurano le variazioni relative a grandezze elementari o globali
- *numeri indici sintetici*: misurano le variazioni di fenomeni complessi

Dato un fenomeno quantitativo X distribuito in una serie temporale con intensità $x_0, x_1, \dots, x_n$ un generico numero indice semplice che misura la variazione da t=0 a t=1 è definito dal rapporto:

N. INDICE SEMPLICE $$ {}_0I_1 = \frac{x_1}{x_0} $$

Il pedice a sinistra indica la base temporale rispetto alla quale è valutata la variazione del fenomeno, mentre il pedice a destra specifica il momento temporale nel quale stiamo effettuando la valutazione detto situazione corrente.

Il numero indice semplice è legato alla variazione relativa dalla formula $\frac{(x_1 - x_0)}{x_0} = {}_0I_1 - 1$

Preso sempre il fenomeno X, può essere significativo confrontare l'intensità del fenomeno anno dopo anno sia con l'intensità riferita ad un anno base, sia all'intensità riferita all'anno precedente. Si parla in questi casi di:

- serie storica a base fissa $ {}_0I_1, {}_0I_2, \dots, {}_0I_n $
- serie storica a base mobile $ {}_0I_1, {}_1I_2, \dots, {}_{n-1}I_n $

I numeri indici semplici: proprietà

Si possono enunciare le seguenti proprietà dei numeri indici elementari:

- *identità*: il numero incide calcolato per il periodo base è uguale a 1
- *reversibilità delle basi*: $ {}_tI_s = \frac{1}{{}_sI_t} = \frac{x_s}{x_t} $
- *transitività delle basi*: $ {}_tI_r = {}_tI_q * {}_qI_r = \frac{x_r}{x_t} $

- commensurabilità: l'indice semplice non varia se muta l'unità di misura del fenomeno

Ad esempio reversibilità e transitività permettono, a partire da una serie storica a base fissa, di effettuare un *cambiamento di base* dividendo tutti i membri della serie per il numero indice riferito all'anno che si vuole prendere a riferimento, detto *coefficiente di conversione*.

I numeri indice sintetici: indici sintetici dei prezzi

Fra gli indici economici, una posizione rilevante è occupata dagli indici temporali sintetici dei prezzi, che misurano la variazione del valore di un paniere di k beni. Tale indice può essere costruito concettualmente seguendo due percorsi:

- approccio *aggregato*: come rapporto fra due aggregati economici (valore del paniere al tempo t)
- approccio *matematico*: come media aritmetica ponderata di indici elementari dei prezzi, tenendo conto della diversa importanza dei beni contenuti all'interno dell'aggregato.

Iniziando con l'approccio aggregato, si definisce il valore della spesa (del paniere) al tempo 0 come $v_0 = \sum_{i=1}^{k} p_{i0} q_{i0}$ Definendo similmente il valore della spesa al tempo generico t, si indica il numero indice sintetico come il rapporto tra di essi, ovvero $\frac{v_t}{v_0} = \frac{\sum_{i=1}^{k} p_{it} q_{it}}{\sum_{i=1}^{k} p_{i0} q_{i0}}$.

Tale rapporto risulta però influenzato sia dalla variazione dei tempi sia dalla variazione delle quantità dei k beni che costituiscono in paniere in 0 e in t. Se si vuole stimare la variazione della spesa imputabile esclusivamente ai prezzi si dovrà *mantenere uguali le quantità* fra i due panieri, ovvero valutare panieri uguali in due tempi diversi, mantenendo la *composizione del paniere del tempo base*. L'indice così costruito è detto numero indice dei prezzi di Laspreyes:

N. INDICE LASPREYES $${}_0^p I_t^L = \frac{\sum_{i=1}^{k} p_{it} q_{i0}}{\sum_{i=1}^{k} p_{i0} q_{i0}}$$ dove la quantità del bene i-esimo q_{i0} è comune

Un'alternativa è di *mantenere la composizione del paniere al tempo corrente*. L'indice così costruito è detto numero indice dei prezzi di Paasche:

N. INDICE PAASCHE $${}_0^p I_t^P = \frac{\sum_{i=1}^{k} p_{it} q_{it}}{\sum_{i=1}^{k} p_{i0} q_{it}}$$ dove la quantità del bene i-esimo q_{it} è comune

Tali indici, come detto, possono essere costruiti da un "punto di vista" matematico, come medie ponderate.

I coefficienti di ponderazione sono definiti come $w_{k0} = \frac{p_{k0} q_{k0}}{\sum_{i=0}^{k} p_{i0} q_{i0}}$, ovvero la quota di spesa totale al tempo base 0 del bene generico k.

I numeri indici semplici per ogni bene del paniere generico k sono definiti come ${}_0 i_t^k = \frac{p_{kt}}{p_{k0}}$.

Il numero indice di Laspreyes (tenuto conto che il coefficiente di ponderazione è stato riferito alle quantità al tempo base 0), si può calcolare come media ponderata di:

N. INDICE LASPREYES(bis) $${}_0^p I_t^L = \sum_{i=1}^{k} w_{i0\,0} i_t^i = \sum_{i=0}^{k} \frac{p_{i0} q_{i0}}{\sum_{i=0}^{k} p_{i0} q_{i0}} \frac{p_{it}}{p_{i0}} = \frac{\sum_{i=1}^{k} p_{it} q_{i0}}{\sum_{i=1}^{k} p_{i0} q_{i0}}$$

Ugualmente il numero indice di Paasche (tenuto conto che il coefficiente di ponderazione è stato riferito alle quantità al tempo corrente generico t), si può calcolare come media ponderata di:

N. INDICE PAASCHE(bis) $${}_0^p I_t^P = \sum_{i=1}^{k} w_{it\,0} i_t^i = \sum_{i=0}^{k} \frac{p_{i0} q_{it}}{\sum_{i=0}^{k} p_{i0} q_{it}} \frac{p_{it}}{p_{i0}} = \frac{\sum_{i=1}^{k} p_{it} q_{it}}{\sum_{i=1}^{k} p_{i0} q_{it}}$$

Confrontando i due indici, si arriva alla conclusione che in caso di *correlazione negativa fra le variazioni dei prezzi e le variazioni delle quantità* (come ci si può logicamente aspettare), l'indice di Laspreyes tende a dare valori più alti dell'indice di Paasche. Questo perché entrambi descrivono situazioni estreme del comportamento del consumatore: nessuna variazione nella composizione del paniere (consumatore conservatore) nell'indice di Laspreyes, variazione immediata e aderenza del paniere alla struttura dell'anno corrente (consumatore speculatore) nell'indice di Paasche.

Il comportamento effettivo del consumatore è più verosimilmente quello di variare le proprie abitudini di consumo più lentamente che al variare dei prezzi. Si introduce pertanto l'indice sintetico di Fisher:

N. INDICE FISHER $_0^p I_t^F = \sqrt{_0^p I_t^L * {}_0^p I_t^P}$ ovvero la media geometrica dei due indici esposti sopra

In conclusione l'*indice di Laspreyes* presenta:

- vantaggi:
  - richiede la conoscenza della composizione del paniere (e dei pesi) del solo tempo base, e richiede esclusivamente una rilevazione dei prezzi correnti
  - ha un significato economico immediato legato al "paniere fisso"
- svantaggi:
  - rapido invecchiamento del sistema di ponderazione (logoramento della base), ovvero il paniere fisso non è più rappresentativo dei consumi reali
  - tendenziosità positiva (tendenzialmente più alto dell'indice di Paasche)

L'ISTAT rileva annualmente tre importanti indici della variazione dei prezzi al consumo:

- indice nazionale dei prezzi al consumo per l'intera collettività (NIC): l'indice più importante, utilizzato per la misura del *tasso di inflazione* dell'intero sistema economico nazionale
- indice nazionale dei prezzi al consumo per le famiglie di operai e impiegati (FOI): indice utilizzato per le rivalutazioni pensionistiche o degli affitti
- indice dei prezzi al consumo armonizzato per i paesi della UE (IPCA): indice utilizzato per l'armonizzazione delle rilevazioni dell'inflazione su base europea

Dal 1999 i tre indici sono stati integrati fra loro in termini di rilevazione, rappresentatività territoriale, metodo di calcolo e campione di prodotti. In particolare il metodo utilizzato è quello degli *indici dei prezzi concatenati* (serie storiche a base mobile) *di tipo Laspreyes*.

Tasso di inflazione

Il tasso di inflazione indica la variazione relativa del valore di un paniere di beni dal tempo base al tempo corrente. Tale tasso si può misurare anche utilizzando gli indici di prezzi. Ad esempio, definiti gli indici di Laspreyes:

- $_0I_{m,t}$ indice NIC-Istat con base 0 e riferito al mese m dell'anno t
- $_0I_{m,t-1}$ indice NIC-Istat con base 0 e riferito al mese m dell'anno t-1

Si possono definire le due tipologie di tassi di inflazione:

- *tasso tendenziale di inflazione* $\frac{{}_0I_{m,t}}{{}_0I_{m,t-1}} - 1 = {}_{m,t-1}I_{m,t} - 1$
- *tasso congiunturale di inflazione* $\frac{{}_0I_{m,t}}{{}_0I_{m-1,t}} - 1 = {}_{m-1,t}I_{m,t} - 1$

Deflazionamento e inflazionamento di aggregati nominali

Confrontare aggregati espressi in valori nominali (PIL, fatturato di un'azienda, ecc) e riferiti ad anni diversi non può esaurirsi nella costruzioni di numeri indice sintetici, in quanto la *variazione a prezzi correnti* non indica quanto il valore degli aggregati sia cambiato per un effettivo aumento del fenomeno o per l'aumento dei prezzi.

Per effettuare confronti significativi, è allora necessario utilizzare le *variazioni a prezzi costanti*, ovvero "depurare" gli aggregati nominali dall'effetto inflattivo utilizzando gli indici di prezzo visti in precedenza.

A seconda che si voglia riportare i dati più recenti a confronto con i dati meno recenti, o viceversa si voglia "aggiornare" i dati più vecchi per confrontarli con quelli più recenti, si ha rispettivamente:

- *deflazionamento*: $_0x_t = \frac{x_t}{{}_0I_t}$ dove si utilizza un indice di prezzo di Laspreyes per riportare l'aggregato x_t ai prezzi passati
- *inflazionamento*: $_tx_{t-1} = x_{t-1} * {}_{t-1}I_t$ dove si utilizza un indice di prezzo di Laspreyes per riportare l'aggregato x_t ai prezzi futuri

Capitolo 4 – Controllo statistico della qualità dei prodotti e dei processi produttivi

Concetti generali

La qualità di un prodotto è sinonimo di "fitness for use". Usando una definizione della ASQC (American Society for Quality Control), si può definire la qualità come la "totalità degli aspetti e delle caratteristiche di un prodotto che influiscono sulla capacità di soddisfare determinate esigenze".

Controllare il processo produttivo e la sua qualità, e dunque controllare la qualità dei prodotti da presentare sul mercato, è una via necessaria per la sopravvivenza competitiva di un'impresa.

La qualità può essere anche interpretata come la coerenza di un prodotto a due categorie di garanzie:

- *garanzie di progettazione*
- *garanzia di tolleranza*

Questo capitolo tratterà la seconda categoria, ovvero l'aderenza di un prodotto ai limiti entro i quali la sua funzionalità è pregiudicata.

Se la qualità di un prodotto è esprimibile attraverso un valore numerico, si definiscono *specifiche* qui valori di misurazione di una caratteristica fisica (peso, lunghezza, resistenza, ecc) oltre il quale il prodotto smette di essere funzionale. Le specifiche di qualità sono esplicitate da una coppia di valori:

- LSL: *Lower Specification Limit*
- USL: *Upper Specification Limit*

LSL e USL indicano il cosiddetto *intervallo di conformità* (intervallo di specifica), ed il prodotto si dirà conforme alle specifiche se il valore x della caratteristica di specifica cade al suo interno.

In particolare si definisce *target* il punto centrale all'intervallo di specifica: $\tau = \frac{LSL + USL}{2}$

Estendendo il ragionamento all'intero processo produttivo, e assumendo l'ipotesi che le osservazioni della caratteristica di specifica X si distribuiscano secondo una *distribuzione normale*, si dice che un processo è sotto controllo quanto tale distribuzione normale rimane costante, ovvero rimangono costanti i parametri che la definiscono (media e varianza), come stimati da una serie di campioni di prodotti selezionati nel tempo.

Eventuali differenze che si osservano in un processo sotto controllo sono da attribuirsi all'azione di una serie continua di piccole azioni di disturbo ineliminabili, o *cause comuni di variazione*. L'effetto cumulativo di questi errori accidentali è comunque espresso dal parametro varianza, che misura la variabilità naturale del processo.

Indici di capacità di processo

Gli indici di capacità di processo vengono introdotti con la funzione di fornire un dato numerico che esprima una valutazione delle prestazioni del processo, nonché segnali se l'eventuale bassa prestazione sia dovuta a un valore troppo alto della varianza e/o alla non centratura della media di processo rispetto al target.

Tali indici si configurano come *indici di efficacia di processo*, e sono utili in confronto con i limiti di specificazione per evidenziare ciò che il processo è in grado di produrre rispetto a ciò che viene richiesto. Per definire la capacità produttiva si usano i limiti di tolleranza naturale (natural tollerance limit), definiti come:

- LNTL: *Lower Natural Tollerance Limit* $LNTN = \mu - 3\sigma$
- UNTL: *Upper Natural Tollerance Limit* $UNTN = \mu + 3\sigma$

Nella distribuzione normale tali valori definiscono un intervallo detto *intervallo di tolleranza*, centrato sulla media, che contiene il 99,73% dei valori della X.

Definiti i limiti di specifica e di tolleranza naturale, possiamo introdurre gli indici di capacità di processo:

- indice C_p: esprime la capacità reale (effettiva) di processo *quando la media coincide con il target*

  INDICE C_p $$C_p = \frac{USL - LSL}{UNTL - LNTL} = \frac{USL - LSL}{(\mu + 3\sigma) - (\mu - 3\sigma)} = \frac{USL - LSL}{6\sigma}$$

  Tale indice confronta al numeratore l'ampiezza dell'intervallo di specifica con quella dell'intervallo di tolleranza, assunto il processo come centrato sulla media. Se $C_p \geq 1$ il processo è detto capace, poiché produce con una certa affidabilità (99,73% del casi), nell'intervallo di specifica. Altrimenti il processo è non capace

- indice C_{pk}: esprime la capacità reale di processo *anche quando la media non coincide con il target*

  INDICE C_{pk} $C_{pk} = min\{C_{pk,L}; C_{pk,U}\}$ dove

  $$C_{pk,L} = \frac{\mu - LSL}{\mu - LNTL} = \frac{\mu - LSL}{\mu - (\mu - 3\sigma)} = \frac{\mu - LSL}{3\sigma} \qquad C_{pk,U} = \frac{USL - \mu}{UNTL - \mu} = \frac{USL - \mu}{(\mu - 3\sigma) - \mu} = \frac{USL - \mu}{3\sigma}$$

  Per definizione è $C_p \geq C_{pk}$ e $C_p = C_{pk} = C_{pk,L} = C_{pk,U}$ quando la media coincide con il target

  Anche per $C_{pk} \geq 1$ il processo è detto capace.

Come detto, se la media è uguale al target i due indici coincidono, e si può utilizzare C_p come misura reale della capacità di processo. Se la media è diversa dal target, i valori differiscono e C_p può essere visto come la misura "potenziale", ovvero la capacità di processo che si potrebbe ottenere migliorando la media fino al target. Dunque l'utilizzo congiunto dei due indici può aiutare a comprendere come migliorare il processo produttivo.

Infine, un processo *capace* è per definizione anche *sotto controllo*, perché i due indici di capacità si riferiscono a processi sotto controllo. Un processo sotto controllo non è detto invece che sia capace.

Obiettivo Sei-Sigma

La filosofia della qualità detta "*Six-Sigma*" mira a realizzare un processo in grado di realizzare una elevatissima proporzione di pezzi conformi e che rimane capace anche se intervengono malfunzionamenti che causano uno shift della media di un certo rilievo, portando il processo fuori controllo. Il mezzo per realizzare questo risultato è il raggiungimento di un bassissimo valore della varianza, talmente basso che i limiti di specificazione LSL e USL vengano a trovarsi a distanza pari a 6σ dal target. Questa situazione, in caso di media coincidente con il target, fa sì che si possa ottenere un C_p di valore 2, considerato il punto di arrivo dei tentativi di miglioramento della capacità di processo. Se anche si verificano degli shift della media di processo dal target, le proporzioni di non conformità rimangono basse.

Il controllo di qualità offline, online, il controllo di accettazione

L'attività di miglioramento della qualità cerca costantemente di controllare e monitorare le fonti di variabilità della distribuzione del carattere intorno al target (dovuta ad elementi umani, tecnologici, caratteristiche ambientali, ecc) per ridurre gli scostamenti delle prestazioni del processo da quelle specificate. In particolare questo obiettivo viene raggiunto mediante:

- prevenzione della non conformità – <u>metodi offline</u>
  Individuano tramite la programmazione di esperimenti i fattori responsabili delle prestazioni, per individuare le condizioni operative ottimali prima di avviare un processo
- diagnosi in corso di produzione – <u>metodi online</u>
  Si tratta di controlli in corso di produzione, con l'obiettivo duplice di effettuare un monitoraggio continuo per comprendere se il processo rimane sotto controllo (segnalando immediatamente eventuali scostamenti) e di consentire di stimare i parametri distributivi del processo nello stato sotto controllo

Oltre a queste due tecniche, esiste anche il cosiddetto <u>controllo di accettazione</u>, che consiste nel controllo di qualità di lotti di materiali, sia da immettere nel processo produttivo sia di prodotti finiti.

Metodi offline e analisi della varianza: terminologia

I metodi offline risultano necessari nella fase di *pianificazione sperimentale* di un processo, prima di passare alla produzione. I principali scopi riguardano l'identificazione di fattori che influenzano le prestazioni del processo (temperatura, ecc) e la misura dei loro effetti in modo da stabilire come aggiustarne il livello in modo da minimizzare la variabilità.

Nel piano sperimentale lo statistico varia in modo controllato i *livelli* (o modalità o valori) di una o più variabili dette *fattori sperimentali* per determinare i loro effetti sulla *variabile di interesse* (o variabile risposta) misurata sulle *unità sperimentali*. Ogni prova dell'esperimento consiste nel fissare il livello dei fattori, nel condurre l'esperimento e nel misurare la variabile risposta. I fattori sono grandezze che vengono trattate in aggregati su scala nominale (alto, medio basso), anche nel caso di fattori di natura quantitativa (temperatura). Se il fattore sperimentale è uno solo, i livelli coincidono con i *trattamenti* ai quali è sottoposta la variabile risposta.

I processi fondamentali della pianificazione sperimentale sono:

- <u>replicazione</u>: la ripetizione dell'esperimento su più unità sperimentali per ogni trattamento
- <u>casualizzazione</u>: assegnare le unità sperimentali ai diversi trattamenti in modo casuale
- <u>controllo locale</u>: applicare principi secondo schemi alla replicazione e alla casualizzazione, per eliminare l'effetto di fattori disturbanti detti sub-sperimentali

Metodi offline e analisi della varianza: ANOVA a una via

Si illustra adesso l'analisi statistica di dati sperimentali in presenza di *un solo fattore* con K trattamenti e n replicazioni per il trattamento. La tecnica dell'ANOVA si propone di *verificare l'ipotesi nulla* di <u>assenza di effetti del fattore sulla Y</u>, intesa come *indipendenza in media*: al variare del livello del fattore sperimentale, la media della variabile risposta non cambia.

Ai fini dell'ANOVA si assume che il risultato della prova j del trattamento i sia rappresentato dalla variabile risposta $Y_{ij} \sim N(\mu_i, \sigma^2)$ che si distribuisce come una normale in cui rimane costante la varianza (*ipotesi di omoschedasticità*).

L'ANOVA si propone di condurre il seguente test di ipotesi:

- $H_0: \mu_1 = \mu_2 = ... = \mu_K = \mu$ vs $H_1: \exists\ \mu_a \neq \mu_b$

Una generica osservazione Y_{ij} può essere vista come $Y_{ij}=\mu_i+\epsilon_{ij}$, ovvero $\epsilon_{ij}=Y_{ij}-\mu_i$ dove $\epsilon_{ij}\sim N(0,\sigma^2)$. La Y_{ij} può essere vista allora come somma di una componente sistematica pari alla media del trattamento e di un *errore sperimentale* epsilon che si distribuisce con media zero, come disturbo accidentale.
Se H_0 è vera diventa $Y_{ij}=\mu+\epsilon_{ij}$, ovvero $\epsilon_{ij}=Y_{ij}-\mu$.
Se H_0 è falsa, ovvero almeno una media di un trattamento è diversa dal mu "generale", sottraendo μ diventa $Y_{ij}-\mu=(\mu_{ij}-\mu)+\epsilon_{ij}$, con $(\mu_{ij}-\mu)\neq 0$ per almeno un trattamento.
Diversamente dalla situazione precedente, la differenza fra Y_{ij} e μ è dovuta non solo a fattori accidentali come epsilon ma per almeno un trattamento anche a cause sistematiche, l'effetto delle quali è espresso da $(\mu_{ij}-\mu)$.
Sempre in questa situazione la varianza è pari a: $E((Y_{ij}-\mu)^2)=(\mu_i-\mu)^2+E(\epsilon_{ij})^2=(\mu_i-\mu)^2+\sigma^2$
Da ciò si deduce che nel caso in cui almeno un trattamento abbia media diversa da μ la variabilità del risultato dell'esperimento dipende sia dalla varianza dell'errore sperimentale (σ^2) sia dall'effetto dovuto alla causa sistematica.
Traslando il ragionamento nell'analisi effettiva di un esperimento con un fattore a K trattamenti e n replicazioni per trattamento, possiamo indicare:

- media trattamento $\bar{y}_i=\frac{1}{n}\sum_{j=1}^{n} y_{ij}$
- meda "generale" $\bar{y}=\frac{1}{K}\sum_{i=1}^{K}\bar{y}_i=\frac{1}{Kn}\sum_{i=1}^{k}\sum_{j=1}^{n} y_{ij}$

Possiamo allora riprodurre la scomposizione sui valori campionari e stimare la variabilità:

- $y_{ij}-\bar{y}=(\bar{y}_i-\bar{y})+r_{ij}$ dove $r_{ij}=y_{ij}-\bar{y}_i$ è detto *residuo*
- dal momento che per costruzione la somma dei residui è zero,
  $\sum_{j=1}^{n}(y_{ij}-\bar{y})^2=\sum_{j=1}^{n}\left[(\bar{y}_i-\bar{y})+(y_{ij}-\bar{y}_i)\right]^2$, ovvero $\sum_{j=1}^{n}(y_{ij}-\bar{y})^2=n(\bar{y}_i-\bar{y})^2+\sum_{j=1}^{n}(y_{ij}-\bar{y}_i)^2$
  dunque $\sum_{j=1}^{n}(y_{ij}-\bar{y})^2=n(\bar{y}_i-\bar{y})^2+\sum_{j=1}^{n}(r_{ij})^2$

Nuovamente si osserva come la variabilità osservata intorno alla media generale dei risultati relativi al trattamento i è pari alla somma di due elementi:

- una componente dovuta solo all'errore sperimentale ineliminabile (cause accidentali) detta ***devianza interna al trattamento i*** (la somma dei quadrati dei residui)
- una componente dovuta all'errore sperimentale ineliminabile (cause accidentali), e se H_0 è falsa dovuta anche a cause sistematiche, detta ***devianza fra le medie dei trattamenti*** (la componente $n(\bar{y}_i-\bar{y})^2$)

Sommando rispetto a ogni trattamento i, si ricava la scomposizione della devianza totale che sta alla base della tecnica ANOVA:
$\sum_{i=1}^{K}\sum_{j=1}^{n}(y_{ij}-\bar{y})^2=n\sum_{i=1}^{K}(\bar{y}_i-\bar{y})^2+\sum_{i=1}^{K}\sum_{j=1}^{n}(r_{ij})^2$ che si può sintetizzare con $sst=ssb+ssw$ dove:

- "sst" significa *sum of squares total* ed è la devianza totale di tutti i valori
- "ssb" significa *sum of squares between* ed è la devianza fra le medie dei trattamenti
- "ssw" significa *sum of suqares within* ed è la devianza interna ai trattamenti

La verifica dell'ipotesi nulla consiste allora nell'effettuare una statistica test tramite *distribuzione di Fisher* (che si dimostra essere la distribuzione che descrive il rapporto fra msb e msw, stimatori corretti della varianza se H_0 è vera) dove:

TEST ANOVA $$F_{oss}=\frac{\frac{ssb}{(K-1)}}{\frac{ssw}{(Kn-K)}}=\frac{msb}{msw}$$

La F_{oss} (valore osservato sulla distribuzione F) assume valori non negativi dal momento che per definizione numeratore e denominatore sono positivi, e assume valore zero solo quando le medie campionarie sono uguali fra di loro. I termini (K-1) e (Kn-K) sono i *gradi di libertà* associati rispettivamente a ssb e ssw, dipendono dal numero di trattamenti e dal numero di replicazioni. Dividere le devianze between e within per i rispettivi gradi di liberà permette di calcolare i valori degli stimatori corretti della varianza: msb (*mean suqares between*) e msw (*mean suqares within*).

Dunque se H_0 è vera (uguaglianza delle medie), F_{oss} è la determinazione di una variabile casuale F di Fisher con (K-1), (Kn-K) gradi di libertà, che si indica con $F_{K-1,Kn-K}$. Assunto un α errore di primo tipo (livello di significatività), ovvero la probabilità di rifiutare H_0 quando è vera, si individua il valore critico F_α tale che $\alpha = P\left(F_{K-1,Kn-K} > F_\alpha\right)$, rifiutando l'ipotesi nulla se $F_{oss} > F_\alpha$. La statistica test conduce pertanto a un test unilaterale, dove la zona di rifiuto dell'ipotesi nulla è posta per valori superiori al valore critico (o di rifiuto) F_α nella coda destra della F di Fisher. Questo perché, se H_0 è falsa, significa che il numeratore msb ingloba anche l'errore sistematico oltre a quello accidentale, assumendo valori superiori di msw e dunque portando a un valore del rapporto più elevato lungo la coda destra della distribuzione.

L'ANOVA fornisce inoltre uno stimatore corretto della varianza σ^2 sia che H_0 sia vera, sia che sia falsa:

$\sigma^2 \approx msw = \frac{ssw}{Kn-K}$. Lo stimatore a partire dalla devianza between è corretto solo se H_0 è vera.

Esperimento con un fattore sperimentale: analisi post-hoc

Il rifiuto dell'ipotesi nulla dell'ANOVA non costituisce un risultato molto significativo, indicando semplicemente che *almeno una media è diversa dalle altre*, senza dare ulteriori informazioni. È necessario condurre allora ulteriori analisi dette "analisi post-hoc" per indagare le cause che hanno portato al rifiuto di H_0, ovvero determinare i trattamenti con medie diverse. Ciò è fondamentale per decidere quale livello del fattore conviene scegliere.

La via più semplice, ma anche più lunga, è quella di confrontare coppie di medie di trattamento tramite test di ipotesi tramite distribuzione di Student per individuare quelle in cui viene rifiutata l'ipotesi nulla di uguaglianza.

Date le ipotesi $H_0: \mu_i - \mu_j = 0$ vs $H_1: \mu_i - \mu_j \neq 0$ occorre calcolare il seguente valore della t di Student:

$$t_{oss} = \frac{\frac{(\bar{y}_i - \bar{y}_j)}{\sqrt{2\sigma^2/n}}}{\sqrt{\frac{ssw}{(Kn-K)\sigma^2}}} = \frac{(\bar{y}_i - \bar{y}_j)}{\sqrt{\frac{2msw}{n}}}$$ dove msw è appunto lo stimatore corretto della varianza.

Se H_0 è vera, t_{oss} è la determinazione di una variabile casuale t di Student con (Kn-K) gradi di libertà, ovvero t_{Kn-K} da confrontare con il valore critico t_α (determinato per il livello di significatività desiderato) in un test di ipotesi bilaterale.

Nel caso di K trattamenti si avranno $h = K(K-1)/2$ confronti a coppie da effettuare simultaneamente (test multiplo), cosa che può risultare particolarmente laboriosa.

Considerazioni finali sull'ANOVA a una via

Se H_0 non viene rifiutata, ovvero i trattamenti non sembrano portare risultati significativamente diversi sulle osservazioni della variabile sperimentale, ciò può essere dovuto:

- a una *effettiva uguaglianza* delle medie di trattamento, anche se può non voler dire necessariamente che il fattore in sé non abbia effetto sulla variabile, ma che in corrispondenza dei livelli scelti le medie della variabile sperimentale non cambiano
- a una *scarsa potenza* del test, che non è stato in grado di rifiutare l'ipotesi nulla:
  - per una stima troppo elevata della varianza σ^2 dovuta alla bassa precisione delle stimatore per un numero di replicazioni troppo ridotto
  - per l'azione di fattori sistematici che l'esperimento non ha considerato e/o controllato. Se infatti nel modello alcuni fattori intervenienti non sono esplicitati, il loro effetto è inglobato nel residuo, in msw, che sarà falsato

L'ANOVA condotta analizzando più di un fattore è detta ANOVA a più vie.

Metodi online e tecnica del control chart

Il controllo online si basa sull'utilizzo del control chart, o carta di controllo. Quando si misura una grandezza fisica si parla di *control chart per variabili*, quando si misura la conformità/non conformità di prodotto oppure il numero di difetti all'interno di un campione si parla di *control chart per attributi*.

Gli scopi del metodo online sono:

- *monitoraggio continuo del processo*: in modo da segnalare tempestivamente se il processo va fuori controllo, ovvero se sono in atto cause sistematiche che determinano una variazione del parametro distributivo (media o varianza o entrambe)
- *stima della capacità di processo*: scelto un trattamento tramite uno studio offline (ad esempio quello che garantisce una media vicina al target), è necessario stimare i parametri di processo realizzati nello stato online, dal momento che il dato offline si riferisce esclusivamente ad un contesto sperimentale. Tale stima dei parametri deve provenire da campioni di un processo effettivamente tenuto sotto controllo

Control chart per variabili: monitoraggio della media di processo

Si tratterà ora prima il monitoraggio della media del processo e solo successivamente la stima dei parametri del processo, nonostante la seconda procedura debba logicamente precedere la prima.

Si continua a fare riferimento ad una misura di qualità di tipo continuo X che in una situazione sotto controllo si distribuisce come una normale. Controllare la media significa allora effettuare ricorrenti *test di ipotesi*, sottoponendo a verifica la media campionaria di campioni casuali di n prodotti usciti dal ciclo produttivo, estratti sistematicamente. La statistica test è basata sul continuo confronto fra la media campionaria (stimatore della media di processo) e la media di processo (stimata con i metodi che verranno proposti in seguito), tramite un semplice test di ipotesi tramite distribuzione normale, dato un livello di significatività α.

Il control-chart è il risultato dell'effettuazione continua di questi test di ipotesi. Per il controllo della media è anche chiamato con il nome di *x-bar chart*, con riferimento alla simbologia per la media campionaria. L'x-bar chart è costruito ponendo in ascissa la sequenza dei campioni estratti, e in ordinata il valore delle medie campionarie. In ordinata è poi riportato un primo valore significativo detto *linea di controllo* (control line), che coincide con il valore $\mu_0 = E(\bar{X}_i | \mu = \mu_0)$.

Si indicano poi sempre in ordinata i valori critici assunti dalla distribuzione, significativi nel test bilaterale, detti *limiti di controllo*:

- lower control limit (LCL)
- upper control limit (UCL)

La scelta di tali valori critici è basata sul livello di significatività α, ovvero la probabilità di *falso allarme* (bloccare il processo ritenendolo fuori controllo quando in realtà è sotto controllo). Quale probabilità deve essere tenuta molto bassa, si tende quindi a ritenere ampiamente soddisfacente un livello pari a 3σ (che come si ricorda esclude nel test bilaterale solo lo 0,27% della distribuzione).

Si definiscono allora: $LCL = \mu_0 - \frac{3\sigma_0}{\sqrt{n}}$ e $UCL = \mu_0 + \frac{3\sigma_0}{\sqrt{n}}$

Control chart per variabili: monitoraggio della variabilità di processo

Nel caso di distribuzione normale è necessario effettuare anche il monitoraggio dell'altro parametro distributivo, ovvero la variabilità del processo. Si tende a monitorare la deviazione standard più che la variabilità per due motivi: è espressa nella stessa unità di misura di X (mentre la varianza è una misura quadratica) e compare nelle formule dei limiti di controllo dell'x-bar chart.

Come nel monitoraggio della media, per monitorare la variabilità tramite la deviazione standard si deve scegliere prima uno stimatore corretto di σ_0, quindi calcolare la linea centrale (valore atteso dello stimatore sotto H_0), e infine calcolare i limiti di controllo.

Il control chart per la variabilità, chiamato *S-chart* (dalla simbologia usata per indicare la deviazione standard campionaria, stimatore corretto della deviazione standard), utilizza appunto come stimatore:

$S_i = \sqrt{\frac{1}{n-1}\sum_{j=1}^{n}(X_{ij} - \bar{X}_i)^2}$ dove i indica il generico campione i di numerosità n.

Come si ricorda, è uno stimatore distorto per una bassa numerosità campionaria, così che è necessario correggerlo per una costante tabulata chiamata c_4 che assume valori sempre più vicini ad 1 all'aumentare della ampiezza campionaria.

Dunque $E(S_i | \sigma = \sigma_0) = c_4\sigma_0$ e $Var(S_i | \sigma = \sigma_0) = (1 - c_4^2)\sigma_0^2$, e allora la linea centrale e limiti di controllo sono: $LC = c_4\sigma_0$, $LCL = max\left(0 ; c_4\sigma_0 - 3\sigma_0\sqrt{(1-c_4^2)}\right)$ e $UCL = c_4\sigma_0 + 3\sigma_0\sqrt{(1-c_4^2)}$.

Si nota infatti come in alcuni casi LCL possa risultare anche negativo (cosa impossibile, si sceglie 0).

Per il test di ipotesi si utilizza la statistica test basata sulla t di Student a n-1 gradi di libertà.

Interpretazione dei control chart

La lettura del control chart non si esaurisce nella sola applicazione del test, ma si rivela importante anche all'interno dei limiti di controllo, per evidenziare alcune conformazioni che evidenzino uno *shift* di uno dei parametri della distribuzione. Ad esempio una tendenza alla crescita del valore assunto dallo S-chart (e uno scostamento monotonico rispetto al valore atteso) può indicare una variabilità in aumento.

È poi importante sottolineare come si debba *visionare prima l'S-chart* e solo *successivamente l'x-bar chart*. Occorre infatti verificare che la variabilità del processo non sia cambiata prima di interpretare l'x-bar chart, dal momento che la deviazione standard è utilizzata per il calcolo dei limiti di controllo di quest'ultimo.

Infine, dal momento che l'informazione contenuta in un control chart è basata su dati campionari, si assiste a un certo grado di incertezza. Il metodo di controllo online non è in grado di segnalare piccoli scostamenti del valore del parametro.

Quando il control chart non evidenzia punti al di fuori i valori dei controllo o andamenti sospetti, allora è più corretto parlare di *processo sotto controllo statistico*, volendo sottolineare come il processo appaia sotto controllo (in un certo intervallo di tempo) in base agli strumenti di monitoraggio utilizzati, o meglio non c'è ragionevole evidenza che i parametri si siano modificati.

Costruzione dei control chart

La costruzione dei control chart coinvolge importanti decisioni riguardanti:

- la dimensione campionaria: più alto è n (a parità di livello di significatività del test e dell'ampiezza dello shift minimo che si vuole individuare) maggiore è la *potenza del test*
- i criteri di estrazione dei campioni: ogni campione è assimilabile a un ccs ma con *cadenza ragionata*, e i campioni usati nei control chart sono detti *sottogruppi razionali* (rational subgroups). La cadenza di estrazione deve favorire che eventuali shift dei parametri avvengano fra campioni successivi e non all'interno dello stesso campione

Il secondo punto, in particolare, suggerisce che sia funzionale raccogliere le unità dello stesso campione nello stesso istante, o comunque il più vicino possibile.

Stima dei parametri di processo

Come detto, si è finora supposto di conoscere di partenza i valori μ_0 e σ_0 del processo sotto controllo. Si analizza ora come effettuarne una stima, basata anch'essa sull'utilizzo dei "chart".

È necessario procedere prima alla stima della deviazione standard e successivamente della media. Il procedimento è il seguente, basato su un elevato numero di campioni (in genere intorno a 30) di una numerosità prefissata n, proprio come se si stesse effettuando un controllo online:

1. si calcola la deviazione standard interna di ogni campione s_i
2. si ricava una stima preliminare di σ_0 come $\sigma_0 \approx \hat{\sigma}_0 = \frac{\bar{s}}{c_4}$ dove $\bar{s} = \frac{1}{m}\sum_{i=1}^{m} s_i$
3. si calcolano LC, LCL e UCL usando la stima $\hat{\sigma}_0$, costruendo il cosiddetto *trial S-chart* con i valori delle deviazioni standard interne degli m campioni, disegnando la spezzata
4. si visiona il grafico cercando punti al di fuori dei limiti di controllo o andamenti sospetti, concludendo altrimenti che gli m campioni sono stati estratti da un processo statisticamente sotto controllo per quanto riguarda la variabilità, validando la stima ottenuta di σ_0

Le fasi della stima della media sono analoghe:

1. si calcola la media interna di ogni campione $\bar{x}_i$
2. si ricava una stima preliminare di μ_0 come $\mu_0 \approx \hat{\mu}_0 = \bar{\bar{x}}$ dove $\bar{\bar{x}} = \frac{1}{m}\sum_{i=1}^{m} \bar{x}_i$
3. si calcolano LC, LCL e UCL usando la stima $\bar{\bar{x}}$ e la stima validata di σ_0, costruendo il cosiddetto *trial x-chart* con i valori delle medie interne degli m campioni, disegnando la spezzata
4. si visiona il grafico cercando punti al di fuori dei limiti di controllo o andamenti sospetti, concludendo altrimenti che gli m campioni sono stati estratti da un processo statisticamente sotto controllo per quanto riguarda la media, validando la stima ottenuta di μ_0

Capitolo 6 – Misura della relazione fra variabili per le decisioni aziendali

Nella gestione dell'impresa capita spesso di avere la necessità di studiare la sussistenza o meno di relazioni fra due variabili di interesse. Si introducono quindi due importanti strumenti:

- l'analisi statistica della correlazione
- i modelli di regressione

Analizzando il legame fra due sole variabili (una indipendente, una dipendente) si parlerà di correlazione e regressione *semplici*, in contrapposizione alla regressione *multipla*, in cui operano due o più variabili indipendenti.

Analisi e misura della correlazione semplice

L'analisi della correlazione fra due variabili quantitative si conduce inizialmente con la costruzione di un *diagramma di dispersione* (scatter plot), dove la disposizione dei punti (dati) può già indicare intuitivamente una qualche correlazione positiva o negativa.

Per avere una misura più precisa della correlazione è utile calcolare l'*indice di Pearson*, utile se le variabili sono continue, distribuite normalmente e hanno una relazione di tipo lineare L'indice è pari al rapporto fra la covarianza delle due variabili e il prodotto delle loro deviazioni standard:

INDICE PEARSON $$corr(x,y)=\rho_{xy}=\frac{cov(x,y)}{\sigma_x * \sigma_y}=\frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y})^2}}$$

L'indice di Pearson può assumere valori fra -1 e 1, dove questi due estremi rappresentano rispettivamente situazioni di correlazione negativa e correlazione positiva.

Modello di regressione lineare semplice

La correlazione descrive tuttavia una relazione biunivoca, senza evidenziare (se esiste) una relazione di dipendenza di una variabile da un'altra indipendente. Se si crede possa esistere una dipendenza lineare di una variabile Y dipendente da una variabile X indipendente, il *modello di regressione lineare semplice* ha l'obbiettivo di stimare ed esprimere la relazione lineare fra le due variabili che si ipotizza avere una funzione analitica del tipo: $Y_i=\alpha+\beta X+\epsilon$.

Dove:

- α : intercetta della retta di regressione "ideale", parametro
- β : pendenza della retta di regressione "ideale", parametro
- ϵ : componente accidentale o di disturbo

Guardando la struttura della funzione, si vede come ad una parte "sistematica" e deterministica si aggiunga una parte "accidentale" casuale.

Il modello di regressione lineare è un modello *probabilistico*, nel senso che, dato un certo valore della variabile deterministica X (variabile indipendente), si può considerare il valore assunto dalla variabile casuale Y secondo il modello di regressione lineare come il valore medio (valore atteso) ovvero $E(Y|X=x_i)=\mu_i=\alpha+\beta x_i$.

Riassumendo vanno assunti alcuni *assiomi*:

- che fra X e Y esista una qualche relazione lineare
- che tale relazione sia influenzata da errori non correlati con valore atteso pari a 0 e stessa varianza σ^2
- che i valori di X siano dati senza errori

Il successo del modello di regressione lineare è dovuto in particolare ai fattori di:

- semplicità: la funzione lineare è una funzione semplice
- sintesi
- effettiva linearità: molti fenomeni di correlazione fra variabili sono lineari o ben approssimati da una relazione lineare
- trasformazioni: spesso si può approssimare relazioni non lineari con relazioni lineari in maniera piuttosto soddisfacente (ad esempio utilizzando le trasformazioni logaritmiche)
- limitatezza dell'intervallo: in un intervallo di piccole dimensioni la regressione lineare è una buona approssimazione anche di una relazione non lineare

Il primo strumento utilizzato nel modello è quello della retta dei minimi quadrati ordinari, ovvero la retta che, fra tutte quelle possibili, *minimizza la somma dei quadrati dei residui* (scarti) *fra valori osservati e valori interpolati della variabile risposta tramite il modello di regressione*.

La retta MQO è la retta dove, dati $\beta \approx b = \frac{ss_{xy}}{ss_x}$ e $\alpha \approx a = \bar{y} - b_1 \bar{x}$ stimatori dei parametri e dove

$ss_{xy} = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$ codevianza di xy e $ss_x = \sum_{i=1}^{n} (x_i - \bar{x})^2$ devianza di x,

questi sono tali ci sia soluzione: $b = \sum_{i=1}^{n} y_i h_i$, dove $h_i = \frac{(x - \bar{x})}{ss_x}$

Si ottiene così una retta di regressione che minimizza le distanza fra valori osservati della Y e valori "interpolati" o previsti della Y, detta anche scarto:

RETTA MQO $\hat{y}_i = a + bx_i$

I residui possono essere interpretati come quelle "differenze" in eccesso o in difetto dei valori osservati rispetto ai valori previsti dal modello, quelle parti del fenomeno "non spiegate" dalla regressione e dunque da minimizzare.

Si dimostra con il teorema di Gauss-Markov che la retta dei minimi quadrati fornisce, fra gli stimatori corretti dei parametri della relazione lineare, gli stimatori più efficienti (meno variabili).

Si possono elencare alcune proprietà della retta dei minimi quadrati:

- la somma dei valori "teorici" è uguale a quella dei valori osservati $\sum_{i=1}^{n} \hat{y}_i = \sum_{i=1}^{n} y$
- dalla proprietà precedente discende che $\bar{\hat{y}} = \bar{y}$
- e che la somma dei residui è pari a 0 $\sum_{i=1}^{n} e_i = \sum_{i}^{n} (y_i - \hat{y}_i) = 0$
- nello scatter plot la retta di regressione passa sempre nel punto di coordinate $(\bar{x}\,;\,\bar{y})$

Bontà di adattamento del modello MQO

Si può calcolare una "misura" dell'adattamento del modello mediante l'*indice di determinazione* R^2, uguale al rapporto fra devianza di regressione e devianza totale della variabile dipendente:

INDICE DI DETERMINAZIONE R^2 $$R^2 = \frac{dev\,reg}{dev\,tot} = 1 - \frac{dev\,res}{dev\,tot} = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$$

Tale indice può assumere valori tra 0 e 1:

- se $R^2=0$: il modello non spiega affatto la variabile risposta, la devianza di regressione è nulla e la variabilità di y non dipende da una relazione con x, almeno non di tipo lineare
- se $R^2=1$: il modello spiega perfettamente la variabile risposta, i punti sono allineati sulla rette di regressione

Nella realtà questi sono due casi limite: il significato dell'indice di determinazione consiste nel misurare la percentuale della devianza totale "spiegata" mediante la retta di regressione.

Riassumendo, la *scomposizione della devianza* si attua nelle seguenti parti:

- *devianza totale*: $\sum_{i=1}^{n} (y_i - \bar{y})^2$ la devianza totale dei valori della variabile dipendente
- *devianza di regressione*: $\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$ la devianza dei valori stimati con la regressione
- *devianza residua*: $\sum_{i=1}^{n} (y_i - \hat{y}_i)^2$ la devianza dei residui

Dove $y_i = \hat{y}_i + e_i$ ovvero $dev\,tot = dev\,reg + dev\,res$ o $DEV(Y) = DEV(\hat{Y}) + DEV(E)$.

Dunque la devianza totale dei valori osservati della variabile Y (variabilità degli scarti dei valori osservati dalla media) è il risultato del contributo di due componenti: la devianza di regressione (variabilità degli scarti dei valori stimati dalla media) e la devianza dei residui (variabilità degli scarti dei valori osservati dai valori stimati). Nell'indice di determinazione allora, si cerca di vedere quanta parte della devianza totale è spiegata dalla devianza di regressione, ovvero dal modello. Tanto più il rapporto $DEV(\hat{Y}) / DEV(Y)$ è vicino a 1, tanto più il modello "spiega" la relazione lineare.

L'indice di determinazione spiega l'adattamento di un modello di regressione lineare, può essere uguale a 0 nonostante fra le due variabili ci sia una relazione, sebbene non lineare.

Stima dell'errore standard (deviazione standard degli ε)

Ciascun residuo può essere considerato una stima dell'errore ε. Si può dimostrare che lo scarto quadratico medio dei residui corretto per il denominatore n-2 è uno stimatore corretto della *deviazione standard degli* ε_i:

$$S_{YX}=\sqrt{\frac{\sum_{i=1}^{n} e_i^2}{n-2}}=\sqrt{\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n-2}}=\sqrt{\frac{DEV(E)}{n-2}}$$

La deviazione standard del fattore erratico misura la variabilità dei valori osservati della Y intorno alla retta di regressione. Tale valore è detto *errore standard*, ed è una caratteristica della retta di regressione.

Inferenza sul coefficiente angolare

Fare inferenza sul valore del coefficiente angolare b della retta di regressione significa verificare tramite test di ipotesi ex post la "validità" dell'assioma che riguardava l'esistenza di una relazione lineare fra una variabile indipendente X e una variabile dipendente Y. Questo perché, data una qualunque serie di osservazioni su uno scatter plot, è sempre possibile ricavare una retta di regressione. La prova dell'esistenza di una relazione lineare non è ovviamente dipendente "quantitativamente" dal valore assunto dal coefficiente angolare, poiché potrebbe esistere una relazione lineare "forte" anche per coefficienti non troppo elevati come 0,5 o 0,1. L'indice di determinazione R^2 può fornire una prima indicazione sulla bontà di adattamento, ma è ancora insufficiente perché non ingloba la variabilità interna al campione delle osservazioni.

È necessario allora fare inferenza, ovvero *condurre un test di ipotesi* che, si spera, rifiuti l'ipotesi nulla $H_0:\beta=0$ vs $H_1:\beta\neq 0$ in un *test di Student*.

È necessario introdurre preliminarmente il valore dell'*errore standard della stima del coefficiente angolare "b"*:

$$S_b=\frac{S_{XY}}{\sqrt{DEV(X)}}=\frac{S_{XY}}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}}$$

Si può quindi effettuare il test di ipotesi calcolando il quantile sulla *t di Student a n-2 gradi di libertà*:

$$t_{n-2}=\frac{b-\beta}{S_b}=\frac{b}{S_b} \qquad per\ H_0:\beta=0$$

Trattandosi di un test bilaterale sottoposto alla variabile dei gradi di libertà, e in assenza di specificati livelli di significatività richiesti, possiamo affermare comunque che:

- si rifiuta sicuramente H_0 se: $t_{n-2}<-4 \wedge t_{n-2}>4$
- si accetta sicuramente H_0 se: $-1.5\leq t_{n-2}\leq 1{,}5$

Capitolo 7 – Analisi delle serie storiche per la programmazione delle attività

Considerazione generali

L'utilizzo di tecniche previsionali è un elemento fondante della pianificazione gestionale di un'impresa, in termini di produzione, gestione delle scorte, budgeting, ecc.

I modelli di previsione si possono dividere in:

- quantitativi: a partire da dati statistici adeguati, si utilizzano modelli per l'analisi e la previsione
  - analisi delle serie storiche
  - metodi esplicativi
- qualitativi: si basano su giudizi di esperti

In questo capitolo verranno presentati i metodi di previsione quantitativi basati sull'utilizzo delle serie storiche.

Previsioni per mezzo dell'analisi delle serie storiche: impostazione

Per *serie storica* si intende una successione di dati osservati su un determinato fenomeno (variabile Y) ordinati secondo la variabile temporale t. La cadenza temporale di questa variabile può fare riferimento ad istanti temporali (*serie di stato* o posizionali) o a intervalli temporali (*serie di flusso*), a seconda della variabile di interesse.

Le serie storiche presentano oscillazioni intorno a un andamento di lungo periodo determinato da fattori di medio o breve periodo. Questi concetti sono riassumibili nella definizione delle 4 componenti "virtuali" (perché non osservabili direttamente ma rappresentabili soltanto attraverso una simulazione) della serie storica tipica dell'ambito economico:

- **trend (T)**: il movimento tendenziale di lungo periodo, di fondo, mette in evidenza un'evoluzione strutturale del fenomeno causata da fattori che agiscono a livello sistematico
- **ciclo (C)**: detto anche movimento (o oscillazione) congiunturale, determinato spesso da condizioni favorevoli o sfavorevoli del ciclo economico, con persistenza nel medio periodo
- **stagionalità (S)**: oscillazioni originate nel breve e brevissimo periodo da fattori climatici, si presentano con una certa regolarità e sistematicità all'interno dell'anno dando luogo ad andamenti ricorrenti a periodo fisso
- **accidentalità (e)**: componente di disturbo data da oscillazioni irregolari provocate da circostanze accidentali, è l'unica componente non sistematica ma casuale

Si ipotizza in genere che i valori assunti dalla componente di disturbo e_i siano una successione di variabili casuali (tipicamente con distribuzione normale) con media pari a 0 e varianza finita, con parametri costanti nel tempo e incorrelate fra loro.

In questo capitolo si affronta l'analisi delle serie storiche mediante l'*approccio classico*, che parte dalla postulazione dell'esistenza di una parte deterministica (che consente di stimare le componenti virtuali) e di una parte di disturbo casuale, in contrapposizione all'approccio moderno che utilizza modelli stocastici. L'approccio classico si basa sul fatto che le componenti presentino nel passato delle regolarità di comportamento (*pattern sistematico*) che si possono estrapolare e tradurre in termini formali e deterministici, ritenendo che si possano presentare anche in futuro, e dunque le proiezioni non saranno altro che "informazioni" sul probabile valore futuro della variabile di interesse

Le fasi di un'analisi delle serie storiche sono tipicamente:

- *analisi del fenomeno* e del problema di previsione da affrontare
- *raccolta dei dati* (già presenti o raccolta ex novo) e verifica della loro qualità e comparabilità temporale
- *analisi preliminare* sull'andamento della serie storica per mezzo di grafici e indici descrittivi
- *scelta e stima del modello* per la valutazione delle componenti virtuali
- *valutazione della bontà del modello* (valutazione della capacità previsiva) e sua utilizzazione ai fini previsivi

Analisi grafiche preliminari

Riportando la serie storica su un *time plot* (scatter plot con in ascissa il tempo), si riesce a cogliere "a vista", se presente, la componente di *trend* di lungo periodo. Se la serie oscilla intorno ad un valore costante (*stazionaria in meda*) allora si è di fronte ad una serie storica priva di trend, se invece si riesce a cogliere un generico "aumento" (o "diminuzione" dai valori) si dirà che la serie è *evolutiva* con trend rispettivamente crescente o decrescente, di tipo lineare o non lineare. Con il time plot, se non ben definite, è difficile riconoscere la presenza di oscillazione cicliche o stagionali.

Per verificare la presenza di *oscillazioni stagionali* è invece utile predisporre il cosiddetto *seasonal plot*, ovvero un time plot con intervallo di ascissa annuale e scansionato in periodi temporali infra-annuali. In ordinata saranno rappresentate le serie storiche (mensili, trimestrali, ecc) di un singolo anno. Si osserverà una stagionalità se le spezzate delle serie si muovono, lungo l'anno, più o meno nello stesso modo.

Modelli di composizione e scomposizione della serie e i metodi di stima delle componenti

Come detto l'approccio classico all'analisi delle serie temporali ipotizza che la serie sia funzione di componenti sistematiche (trend, ciclo, stagionalità) e la componente accidentale di disturbo, ovvero:

$$Y_t = f(T_t, C_t, S_t, e_t)$$

Per quanto riguarda la stima della componente ciclica C_t, l'argomento non vine affrontato dal momento che il ciclo economico non presenta oscillazioni regolari e la sua stima pone numerose difficoltà. Tale componente verrà omessa, o meglio considerata come facente parte della componente di trend (trend-ciclo).

Il primo passo è definire il modo con cui le componenti virtuali interagiscono fra loro per dar luogo alla serie, ovvero è necessario specificare la funzione f.

Le sue due principali forme sono:

- **modello additivo** $y_t = T_t + S_t + e_t$
- **modello moltiplicativo** $y_t = T_t \times S_t \times e_t$

Il modello additivo è appropriato quando l'ampiezza della oscillazione stagionale *non varia al variare* del livello della serie (che si definirà allora "serie additiva").

Il modello moltiplicativo è appropriato quando la fluttuazione stagionale *aumenta/diminuisce proporzionalmente* al livello della serie (che si definirà allora "serie moltiplicativa").

Successivamente occorre definire i metodi da utilizzare per la stima delle componenti, che possono essere:

- **metodi empirici**: in cui lo strumento più utilizzato è quello della media mobile
- **metodi analitici**: in cui si effettua una interpolazione ottenendo una funzione analitica

I due metodi non sono alternativi, e possono essere utilizzati congiuntamente.

Impiego delle medie mobili per eliminare le oscillazione e stimare le componenti sistematiche

Il metodo delle medie mobili consiste nel calcolo di una nuova serie storica i cui termini sono il risultato della media di k termini continui della serie di partenza. Possono presentarsi due casi nella definizione di una serie di medie mobili a k termini:

- k sia *dispari*: il termine della media mobile si riferisce al tempo su cui è stata calcolata, valore che lascia prima e dopo un uguale numero di termini. Si parla di *media mobile semplice*
- k sia *pari*: è necessario calcolare la media di due medie mobili contigue per ottenere un valore "centrato" sul tempo t di riferimento. Si parla in questo caso di *media mobile centrata*

  Il procedimento consiste nel calcolo di una *media ponderata* di k+1 termini (dove k+1 sarà un numero dispari), nella quale è attribuito peso doppio a tutti i termini ad esclusione di quelli marginali

L'applicazione delle medie mobili contente di "smorzare" le oscillazioni, in un processo di *lisciamento della serie* (smoothing). Esse hanno in particolare la proprietà di eliminare o ridurre le oscillazioni con periodo pari al numero dei termini coinvolti nel calcolo della serie. L'applicazione di una media mobile a 12 termini (se i dati sono mensili) o a 4 termini (se i dati sono trimestrali), consente di eliminare le oscillazioni stagionali.

Stima della stagionalità, della serie destagionalizzata e del trend-ciclo

Si analizzerà adesso come si stimano le componenti virtuali di una serie storica impiegando le medie mobili. Le fasi sono le seguenti (serie di dati mensili con stagionalità annuale):

1. calcolo della media mobile centrata a 12 termini

   Dovrebbe eliminare le oscillazioni stagionali e gran parte della componente erratica, rappresentando una prima approssimazione del trend-ciclo che possiamo indicare con $T_t^{(1)}$
2. calcolo della componente di stagionalità mista a errore

   Si confronta la serie originaria y_t con la serie stimata $T_t^{(1)}$, e il confronto consisterà in:
   - una *differenza nel modello additivo* $(S,e)_t = y_t - T_t^{(1)}$
   - un *rapporto nel modello moltiplicativo* $(S,e)_t = \frac{y_t}{T_t^{(1)}}$

   La prima sarà detta differenza lorda di stagionalità, mentre il secondo coefficiente lordo di stagionalità
3. stima della componente stagionale

   L'obiettivo è stimare un coefficiente (additivo o moltiplicativo) per il singolo mese, sotto l'*ipotesi di stagionalità*, ovvero che la stagionalità si presenti costante negli anni, con differenze dovute solo alla componente erratica. Per fare ciò occorre:
   - calcolare la media aritmetica dei termini $(S,e)_t$ riferiti allo stesso mese nei vari anni di osservazione, per eliminare le componente di errore. Si ottengono 12 coefficienti diversi
   - si verifica che questi 12 coefficienti mensili verifichino il "principio di conservazione delle aree", ovvero che le oscillazioni stagionali esauriscano l'effetto all'interno dell'anno. In tal caso:
     - la media dei coefficienti stagionali deve essere uguale a 0 nel modello additivo
     - la media dei coefficienti stagionali deve essere uguale a 1 nel modello moltiplicativo

   Queste operazioni permettono di ottenere i *coefficienti netti di stagionalità*, una stima definitiva dei coefficienti mensili
4. derivazione della serie destagionalizzata D_t

   Si utilizzano i coefficienti netti di stagionalità per eliminare la stagionalità dalla serie originaria, ottenendo la *serie destagionalizzata*: in particolare sottraendo i coefficienti additivi e dividendo per i coefficienti moltiplicativi
5. stima del trend-ciclo

   La serie destagionalizzata non dovrebbe presentare oscillazioni stagionali, e contenere solamente il trend-ciclo e l'effetto di disturbo. Per ottenere una stima definitiva del trend-ciclo si può ridurre le oscillazioni residue con una media mobile con un opportuno numero di termini o una regressione lineare

Si può dunque ricomporre la parte sistematica virtuale della serie storica che contenga le due componenti:

- modello additivo $\hat{y}_t = \hat{T}_t + \hat{\hat{S}}_t$
- modello moltiplicativo $\hat{y}_t = \hat{T}_t \times \hat{\hat{S}}_t$

La valutazione della bontà del modello e della sua capacità previsiva

Costruito il modello previsionale, è necessario analizzare la sua "bontà" di adattamento alla realtà, sia in senso ex-post che ex-ante, distinguendo fra:

- *goodness of fit* : $r_t = y_t - \hat{y}_t$
  La bontà di adattamento "ex-post", effettuata su un primo set di valori osservati, per verificare il funzionamento del modello
- *goodness of forecast* : $ep_t = y_{t+h} - F_{t+h}$
  La bontà di previsione, estensione del modello previsionale nel futuro

Per calcolare r_t e ep_t si utilizzano seguenti indici di errore:

- errore medio (ME) $ME = \frac{1}{m}\sum_{t=1}^{m}(y_t - \hat{y}_t)$
- errore quadratico medio (MSE) $MSE = \frac{1}{m}\sum_{t=1}^{m}(y_t - \hat{y}_t)^2$
- errore medio assoluto (MAE) $MAE = \frac{1}{m}\sum_{t=1}^{m}|y_t - \hat{y}_t|$
- errore medio assoluto percentuale (MAPE) $MAPE = \frac{1}{m}\sum_{t=1}^{m}\frac{|y_t - \hat{y}_t|}{y_t}$

Capitolo 5 – Performance tecnica del processo produttivo: produttività ed efficienza

I concetti di *produttività* ed *efficienza*, sebbene usati a volte come sinonimi nella lingua comune, si riferiscono a due differenti chiavi di lettura della performance di un processo produttivo:

- efficienza: pone l'accento sulla distanza, in termini relativi, del processo stesso dal suo limite potenziale, dipendente dalla tecnologia produttiva esistente (processo benchmark)
- produttività: pone l'accento sul confronto fra il risultato del processo produttivo (output) e i mezzi utilizzati per ottenerlo (input)

Concetto di efficienza

L'efficienza è definita come *la distanza in termini relativi di un processo dal suo potenziale*, ed è quindi un numero puro non espresso in unità di misura. A seconda che si analizzi l'output o l'input si parlerà di:

- ***indice di output-efficienza***
  la misura del massimo output producibile utilizzato un certo quantitativo di input

  OUTPUT-EFFICIENZA $\varphi_k = \frac{output\ prodotto}{output\ producibile}$ (per definizione <1)
- ***indice di input-efficienza***
  la misura del minimo input impiegabile (della frazione λ di input impiegabile rispetto all'utilizzo effettivo) per ottenere un certo quantitativo di output

  INPUT-EFFICIENZA $\lambda_k = \frac{input\ impiegabile}{input\ impiegato}$ (per definizione <1)

Concetto di produttività

La produttività è definita come *il rapporto fra output prodotto e input utilizzato*, nel caso di un processo che utilizza un solo input per produrre un solo output si avrà:

PRODUTTIVITA' $P = \frac{y}{x}$

Nel caso di processo produttivo con più di un input si distingue fra:

- *produttività parziale*: la produttività calcolata come rapporto fra output e singolo input
- *produttività totale dei fattori* (PTF): la produttività calcolata su tutti gli input, più difficile da misurare per il cosiddetto "problema di aggregazione" delle produttività parziali

Concetti economici della teoria di produzione

Si riportano alcuni concetti propri della teoria di produzione microeconomica:

- aspetto tecnico (input convertito in output in un processo) vs aspetto economico (profitto come differenza fra i ricavi delle vendite degli output ed i costi dell'acquisto di input)

- breve periodo (input fissi e variabili) e lungo periodo (tutti gli input sono considerati variabili)
- *rendimenti di scala*:
  - crescenti aumento di input di una frazione λ causa aumento più che proporzionale di output
  - costanti aumento di input di una frazione λ causa aumento proporzionale di output
  - decrescenti aumento di input di una frazione λ causa aumento meno che proporzionale di output

Funzione di produzione in una tecnologia monoprodotto

Verrà trattato esclusivamente il processo caratterizzato da una *tecnologia monoprodotto*, ovvero caratterizzato da un solo output. Si definisce la funzione di produzione come la funzione che associa a un vettore di input la massima quantità producibile di output secondo la tecnologia esistente:

F. DI PRODUZIONE $$f(\boldsymbol{x}) = \{ max\ y \in \Re | z(\boldsymbol{x}; y) \in \boldsymbol{Z} \}$$

La funzione di produzione è detta anche "frontiera produttiva", con il significato di limite massimo raggiungibile dalla tecnologia.

Il modello analitico più utilizzato in microeconomia per descrivere la relazione fra input ed output di un processo produttivo è il ***modello Cobb-Douglas***, sia per le sue caratteristiche di semplicità analitica (è una funzione log-lineare, linearizzabile mediante trasformazione logaritmica) sia per la sua coerenza con la teoria economica (isoquanti convessi, saggio marginale di sostituzione decrescente, elasticità unitaria costante, identificazione immediata del rendimento di scala).

La funzione di produzione Cobb-Douglas è:

COBB-DOUGLAS $$f_{CD} \qquad \boldsymbol{y} = A \cdot \prod_{j=1}^{q} x_j^{b_j}$$ (nel caso di due input è $\boldsymbol{y} = A \cdot x_1^{\alpha_1} \cdot x_2^{\alpha_2}$)

E se operata una trasformazione logaritmica: $$f_{CDL} \qquad \log(\boldsymbol{y}) = \log(A) \cdot \sum_{j=1}^{q} b_j \cdot \log(x_j)$$

Nel caso di due input l'identificazione dei rendimenti di scala è possibile sommando gli esponenti delle variabili:

- rendimenti di scala crescenti $\alpha_1 + \alpha_2 > 1$
- rendimenti di scala costanti $\alpha_1 + \alpha_2 = 1$
- rendimenti di scala decrescenti $\alpha_1 + \alpha_2 < 1$

Determinare i parametri caratteristici della Cobb-Douglas può essere molto complicato. Prendendo un certo numero di osservazioni di output di un processo (un certo numero di vettori $\boldsymbol{z}(\boldsymbol{x}; y)$), un primo modo potrebbe essere quello di stimare i parametri mediante una regressione lineare. Tuttavia la funzione lineare trovata (retta dei minimi quadrati) non sarebbe coerente con il concetto della "frontiera di produzione", poiché un certo numero di processi darebbero luogo a quantitativi di output superiori al limite posto dalla funzione. Il problema di stima si risolve impostando la risoluzione della regressione lineare (minimizzazione degli scarti quadrati) sotto il vincolo della non positività dei residui.

Misura dell'efficienza ricavata dalla funzione di produzione

Definita la funzione di produzione come una funzione Cobb-Douglas, possiamo recuperare i concetti di efficienza in un'ottica analitica:

- ***output-efficienza***

  Il rapporto fra l'output prodotto dal processo ed il massimo output producibile con lo stesso input, ovvero il valore della funzione di produzione calcolato per quel quantitativo di input

  OUTPUT-EFFICIENZA $$\varphi_k = \frac{y_k}{A \cdot \prod_{j=1}^{q} (x_{kj})^{\alpha_j}}$$ (per definizione <1)

- ***indice di input-efficienza***

  Il valore della frazione λ di input impiegabile rispetto all'utilizzo effettivo per ottenere un certo quantitativo di output, è necessario estrarlo dal valore assunto dal funzione di produzione ovvero

  $y_k = A \cdot \prod_{j=1}^{q} (\lambda \cdot x_{kj})^{\alpha_j}$ ovvero

  INPUT-EFFICIENZA $$\lambda = \left(\frac{y_k}{A \cdot \prod_{j=1}^{q} (x_{kj})^{\alpha_j}} \right)^{\frac{1}{\sum_{j=1}^{q} \alpha_j}}$$ (per definizione <1)

Misura della produttività ricavata dalla funzione di produzione

Recuperando il concetto di produttività esposto in precedenza, si è interessati alla sua misura in particolare per un confronto:

- a *livello temporale*, per evidenziare la variazione della produttività:
  - tra due periodi consecutivi
  - in confronti multiperiodali
- a *livello interaziendale*, per evidenziale i divari di produttività fra aziende:
  - tra due aziende
  - fra più aziende (confronti multilaterali)

La misura della produttività (per uno dei due scopi) in caso di un processo caratterizzato da un solo input e un solo output è semplice, basandosi su *numeri indice di produttività* ovvero il rapporto fra le due produttività degli istanti di tempo considerati, che può anche essere visto come il rapporto fra la variazione relativa di output e di input:

$$ {}_sP_t = \frac{P_t}{P_s} = \frac{y_t / x_t}{y_s / x_s} = \frac{y_t / y_s}{x_t / x_s} $$

Nel caso invece di un processo che impiega due o più input per produrre due o più output, le cose sono più complicate. A differenza dell'efficienza in questo caso si considerano processi produttivi multiprodotto. Se calcolare le produttività parziali in un rapporto output specifico/input specifico non pone problemi, calcolare la *produttività totale dei fattori* (PTF o TFP "total factor productivity") necessita di alcune considerazioni.

Sarà infatti necessario calcolare un rapporto che abbia al numeratore un'aggregazione degli indici delle quantità di output e al denominatore un'aggregazione degli indici delle quantità di input.

Nel caso voglia conoscere ad esempio la variazione di produttività fra due tempi in un'azienda caratterizzata da processo a due input e due output sarà:

$$ TFP = \frac{aggregazione(y_{21}/y_{11}) e (y_{22}/y_{12})}{aggregazione(x_{21}/x_{11}) e (x_{22}/x_{12})} $$

Tale aggregazione viene effettuata utilizzando una *media aritmetica ponderata* degli indici di variazione degli output e degli input, dove i pesi sono rispettivamente i *ricavi di vendita degli output* ed i *costi di acquisto degli input*, valori che sono conoscibili solo se si conoscono i prezzi unitari di vendita e acquisto.

Aggregare medie ponderate di indici semplici porta alla creazione dei già analizzati *indici di Laspeyres e di Paasche*, dove la scelta fra i due dipende (trattandosi questa volta di indici di quantità, e di dover fissare allora un "paniere di prezzi") dalla scelta dei prezzi unitari di vendita e acquisto, ovvero se riferirli al tempo base o al tempo attuale dell'indice.

Per il sistema dei pesi di Laspeyres (sistema dei prezzi riferito al tempo base) avremo:

- ***indice delle quantità di output***

$$ {}_sIY_t^L = \sum_{i=1}^{m} \left(\frac{y_{it}}{y_{is}} \cdot \frac{p_{is} \cdot y_{is}}{\sum_{i=1}^{m} p_{is} \cdot y_{is}} \right) = \frac{\sum_{i=1}^{m} p_{is} \cdot y_{it}}{\sum_{i=1}^{m} p_{is} \cdot y_{is}} $$

- ***indice delle quantità di input***

$$ {}_sIX_t^L = \sum_{i=1}^{m} \left(\frac{x_{it}}{x_{is}} \cdot \frac{p_{is} \cdot x_{is}}{\sum_{i=1}^{m} p_{is} \cdot x_{is}} \right) = \frac{\sum_{i=1}^{m} p_{is} \cdot x_{it}}{\sum_{i=1}^{m} p_{is} \cdot x_{is}} $$

Dove indici e prezzi unitari sono quelli riferiti rispettivamente a output e input.

Ottenuti i due indici sintetici di output ed input per calcolare la produttività totale dei fattori basterà farne il rapporto:

INDICE V. PRODUTTIVITA' TOTALE (Laspeyres) $$ TFP_{st}^L = \frac{{}_sIY_t^L}{{}_sIX_t^L} $$

Per il sistema dei pesi di Paasche (sistema dei prezzi riferito al tempo attuale) avremo:

- ***indice delle quantità di output***

$$ {}_sIY_t^P = \sum_{i=1}^{m} \left(\frac{y_{it}}{y_{is}} \cdot \frac{p_{it} \cdot y_{is}}{\sum_{i=1}^{m} p_{it} \cdot y_{is}} \right) = \frac{\sum_{i=1}^{m} p_{it} \cdot y_{it}}{\sum_{i=1}^{m} p_{it} \cdot y_{is}} $$

- ***indice delle quantità di input*** $${}_sIX_t^P = \sum_{i=1}^{m} \left(\frac{x_{it}}{x_{is}} \cdot \frac{p_{it} \cdot x_{is}}{\sum_{i=1}^{m} p_{it} \cdot x_{is}} \right) = \frac{\sum_{i=1}^{m} p_{it} \cdot x_{it}}{\sum_{i=1}^{m} p_{it} \cdot x_{is}}$$

INDICE V. PRODUTTIVITA' TOTALE (Paasche) $$TFP_{st}^P = \frac{{}_sIY_t^P}{{}_sIX_t^P}$$

Se si vuole ottenere un valore migliore della variazione di produttività totale dei fattori si può utilizzare l'*indice di Fisher*, che essendo la media geometrica dei due indici può essere "trasposto" ai due indici di produttività totale:

INDICE V. PRODUTTIVITA' TOTALE (Fisher) $$TFP_{st}^F = \sqrt{TFP_{st}^L \cdot TFP_{st}^P}$$

In particolare l'indice di Fisher è utile in quanto gode della *proprietà di reversibilità della base*, ovvero invertendo tempo base e tempo attuale si ottiene il reciproco del valore di partenza, ovvero ad esempio:

$${}_sIX_t^F = \frac{1}{{}_tIX_s^F}$$

Né l'indice di Laspeyres né quello di Paasche godono di tale proprietà.