CAP. 16 SERIE STORICHE

Una serie storica e' una successione di valori di un fenomeno osservati in relazione al tempo.

S.S. tempi → $t = 1, 2, \dots, s$; valori del fenomeno y_t

S.S. discrete / continue , S.S. di stato / di flusso. → per tutto il periodo considerato
↳ periodo preciso

SCOPI DELLE ANALISI DELLE S.S.

- Descrizione dell' andamento storico del fenomeno osservato
- Obiettivo della PREVISIONE
- Scoprire della struttura interna dell' andamento di un fenomeno osservato.

Individuare una funzione matematica → UN MODELLO — estrapolazione / interpolazione

Ci sono degli indici che tengono conto dell' ORDINE DI SUCCESSIONE che si ha nelle S.S.

INDICE DI OSCILLAZIONE → misura l'intensita' assoluta delle oscillazioni consecutive

$$O_s = \frac{\sum_{t=1}^{s-1} |y_{t+1} - y_t|}{s-1}$$

INDICE DI EVOLUZIONE → serve a rappresentare la tendenza del fenomeno a crescere o a diminuire

$$E_v = \frac{y_s - y_1}{s-1}$$

Approcci metodologici di analisi delle serie storiche :

- Approccio TRADIZIONALE o CLASSICO
- Approccio STOCASTICO o MODERNO (si richiama alla nozione di processo stocastico)
- Approccio di tipo FREQUENZIALE (si basa sul presupposto che una serie temporale si possa considerare la risultante della somma di infinite serie periodiche con periodo, ampiezza e fasi diverse → serie di FOURIER)

APPROCCIO CLASSICO

I valori osservati si considerano esprimibili, per ciascuno dei tempi considerati, come la risultante di una componente sistematica (deterministica) di una componente aleatoria. Si assume che la parte sistematica del modello sia scindibile in una componente tendenziale di lungo periodo (TREND) in una componente congiunturale (CICLO) e, se i dati sono rilevati con cadenza inferiore all' anno, in una componente STAGIONALE.

E' una impostazione fondata sul presupposto che le singole osservazioni siano il risultato dell' azione variamente combinata di movimenti non direttamente osservabili che si sommano e/o si moltiplicano nei singoli tempi considerati, movimenti che e' legittimo tentare di scomporre e misurare.

ANALISI COMPONENTI S.S.

TREND : e' la tendenza di fondo che caratterizza l'evoluzione del fenomeno nel lungo periodo.

CICLO : curve sinusoidali di lungo periodo, non hanno necessariamente stessa durata e ampiezza.

Componente STAGIONALE : movimento che ricorre pressappoco sotto la stessa forma ogni anno

Componente OCCASIONALE : e' connessa a fattori perfettamente individuabili che si manifestano in maniera episodica.

Componente CASUALE : Racchiude effetti di natura strettamente accidentale o erratica.

MODELLO ADDITIVO → $y_t = T_t + C_t + S_t + O_t + E_t$

MODELLO MOLTIPLICATIVO → $y_t = T_t \cdot C_t \cdot S_t \cdot O_t \cdot E_t$

MODELLO MISTO → $y_t = (T_t + C_t + O_t) \cdot S_t \cdot E_t$

Metodo dei residui

→ scomposizione della serie storica nelle varie componenti elementari.

Individuata, con opportuna procedura, una prima componente della serie originaria, la serie DEPURATA di quella componente si ottiene per differenza nel modello additivo o per rapporto nel modello moltiplicativo.

ELIMINARE LA COMPONENTE OCCASIONALE

Dati mensili → correggere i dati

I dati di ciascun mese devono essere divisi per il numero effettivo di giorni del mese e moltiplicati per il numero medio di giorni in un mese (365 : 12 o 366 : 12). Durata convenzionale mese → 30 gg.

S.S. STAZIONARIA → Quando alcune grandezze statistiche - media, varianza, ecc... - calcolate sui dati di una s.s. in un certo intervallo di tempo, NON SI MODIFICANO, quando si considerano periodi di tempo ad esso precedenti o successivi.

S.S. EVOLUTIVE → caratterizzate da TREND, componenti cicliche e stagionali accentuate e immancabili oscillazioni di tipo casuale.

→ DETRENDIZZARE → consiste nell' individuare prima di tutto il tipo di trend dalla s.s. originaria evolutiva, poi nel toglierlo dai dati della s.s. (sottraendo nell' additiva o dividendo nella moltiplicativa) dai valori effettivi quelli corrispondenti al trend, in modo da ricavare una SERIE RESIDUA, oscillante attorno ad un VALORE MEDIO COSTANTE.

- PER INDIVIDUARE IL TREND → perequazione grafica MEDIE MOBILI

RAPPRESENTARE IL TREND CON UNA FUNZIONE MATEMATICA

— la *retta* (sviluppo aritmetico)
$$T_t = a + bt\ ,$$
— la *funzione esponenziale* (sviluppo geometrico)
$$T_t = \alpha\, \beta^t\ ,$$
— la *parabola di secondo grado*
$$T_t = a + bt + ct^2\ ,$$
— l'*esponenziale di parabola di secondo grado*
$$T_t = \alpha\, \beta^t\, \gamma^{t^2}\ ,$$
— la *funzione esponenziale generalizzata*
$$T_t = a + b\, c^t\ ,$$
— la *funzione di Gompertz*
$$T_t = \alpha\, \beta^{c^t}\ ,$$
— l'*iperbole*
$$T_t = \frac{1}{a + bt}\ ,$$

→ le più "famose" sono:

PARABOLA $T_t = a + bt + ct^2$

ESPONENZIALE $T_t = a\beta^t$

RETTA $T_t = a + bt$

METODO ANAMORFOSI

Ciascuna delle funzioni del trend, in base ad una opportuna TRASFORMAZIONE, diventa una RETTA

La funzione del trend che meglio rappresenta la s.s. e' individuata da quella trasformazione della s.s. che da' luogo ad una s.s. con trend all' incirca rettilineo.

TREND RETTILINEO

$$T_t = a + bt$$

$$T_{t+1} - T_t = b \qquad \frac{b}{y}$$

b: VARIAZIONE ASSOLUTA DEL TREND DEL FENOMENO NELL'UNITÀ DI TEMPO.

a: VALORE DEL TREND DEL FENOMENO AL TEMPO ZERO

LE VARIAZIONI DI T_t SONO PROPORZIONALI ALLE VARIAZIONI DEL PERIODO DI TEMPO

→ per trovare i parametri a e b si suole traslare il tempo $t' = t - \frac{s+1}{2}$

il centro dell' intervallo di osservazione corrisponde a ZERO

$$T_t = a + b\left(t - \frac{s+1}{2}\right) \rightarrow T_t = a' + bt'$$

$$a' = a - b \cdot \frac{(s+1)}{2}$$

Con il metodo dei minimi quadrati si ottiene

$$\begin{cases} b = \dfrac{12 \sum \left(t - \frac{s+1}{2}\right) y_t}{s(s^2 - 1)} \\ a = \dfrac{\sum y_t}{s} - \dfrac{b(s+1)}{2} \end{cases}$$

$$\frac{\sum y_t}{s} = \bar{y}$$

$$\boxed{\begin{cases} a' = \dfrac{\sum y_t}{s} \\ b = \dfrac{12 \sum \left(t - \frac{s+1}{2}\right) y_t}{s(s^2 - 1)} \end{cases}}$$

PARABOLA

$$T_t = a + bt + ct^2$$

$$T_{t+1} - T_t = (b + c) + 2ct$$

SE LE DIFFERENZE TRA I VALORI DELLA SERIE SONO ALLINEATE → TREND PARABOLICO

C → ACCELERAZIONE O DECELERAZIONE DELLA VARIAZIONE DEL FENOMENO AL VARIARE DEL TEMPO

$$y_t - T_t = C_t + S_t + E_t$$

$$y'_t = y - T_t$$

$$\hookrightarrow y'_t = C_t + S_t + E_t$$

TREND ESPONENZIALE

$$T_t = \alpha\beta^t$$

$\alpha > 0$ VALORE DEL FENOMENO AL TEMPO 0
$\beta > 0$ RAPPORTO COSTANTE TRA I SUCCESSIVI VALORI DEL TREND

$$\frac{T_{t+1}}{T_t} = \frac{\alpha\beta^{t+1}}{\alpha\beta^t} = \beta$$

$\beta - 1$ VARIAZIONE RELATIVA COSTANTE TRA UN TEMPO E IL SUCCESSIVO

→ si trasforma in rettilineo

$$\log T_t = \log \alpha + t \log \beta$$

$Y_t = \log T_t$ $\quad a = \log \alpha$ $\quad b = \log \beta$

$Y_t = a + bt$ $\qquad Y_t = a' + bt$

$$a' = \log a' = \frac{\sum \log Y_t}{S}$$, media geometrica

$$b = \log \beta = \frac{12 \sum_1^S \left(t - \frac{S+1}{2}\right) \log Y_t}{S(S^2-1)}$$

OSSERVAZIONI CHE SI RIFERISCONO A FRAZIONI DI ANNO

Dopo la componente tendenziale si ELIMINA la componente stagionale → oscillazioni perfettamente ricorrenti (stesso periodo, stessa ampiezza)
↳ LE MEDIE MOBILI con un numero di termini pari alla lunghezza del periodo eliminano del tutto le oscillazioni

$$y_t^* = \frac{\frac{1}{2}y'_{t-6} + y'_{t-5} + y'_{t-4} + \ldots + y'_t + \ldots + y'_{t+4} + y'_{t+5} + \frac{1}{2}y'_{t+6}}{12}$$

per $t = 7, 8, \ldots, s-6$

→ numero di termini pari al numero dei mesi nell'anno. si ottiene una serie perequata centrata.

TROVA LA COMPONENTE CICLICA

$Y''_t = Y_t - T_t - C_t$ la serie depurata da trend, occasionale, e ciclicita'.

per trovare la S_t si fa le medie delle Y''_t degli stessi mesi, divise per quante volte c'e' lo stesso mese

$Y''_t - S_t = E_t$

Destagionalizzazione s.s. con dati trimestrali → medie mobili a 5 termini

$$y_t^* = \frac{\frac{1}{2}y'_{t-2} + y'_{t-1} + y'_t + y'_{t+1} + \frac{1}{2}y'_{t+2}}{4};$$

per $t = 3, 4, \ldots, s-2$

DESTAGIONALIZZAZIONE DATI MENSILI

MODELLO ADDITIVO

$\Delta_{12} y'_t = y'_t - y'_{t-12}$ per $t = 13, 14, \ldots, s$

MODELLO MOLTIPLICATIVO

${}_{t-12}R_t = \frac{y'_t}{y'_{t-12}}$ per $t = 13, 14, \ldots, s$

DIFFERENZE O RAPPORTI TRA I TERMINI DELLA SERIE SLITTATI DI UN ANNO

$y_t = T_t + C_t + S_t + \cancel{O_t} + E_t$ SERIE STORICA

$y'_t = C_t + S_t + E_t$ VALORI DETRENDIZZATI

$y^*_t = C_t + \cancel{E_t}$ VALORI PEREQUATI (MEDIE MOBILI)

$y^*_t = C_t$ < DATI INFRANNUALI / DATI ANNUALI

$y'_t - y^*_t = S_t + E_t$

MEDIE MENSILI → S_t

IN UNA S.S. CON DATI MENSILI

LA COMPONENTE CICLICA SI OTTIENE DALLA PEREQUAZIONE DELLA SERIE DETRENDIZZATA

$$y^*_t = C_t$$

Individuazione della componente stagionale

↳ SERIE RESIDUA = Comp. STAGIONALE + comp. ERRATICA

metodo delle medie mensili: si effettuano le medie aritmetiche di tutti i valori della serie residuale che si riferiscono a ciascun mese (allo stesso mese) → le 12 medie così ottenute esprimono la componente stagionale

Una serie storica STAZIONARIA (o resa tale) può presentare altri elementi di regolarità suscettibili di essere utilizzati con finalità INVESTIGATIVE (previsione).
Il CORRELOGRAMMA costituisce uno strumento utile per accertare se la serie storica è caratterizzata da qualche REGOLARITÀ → consente di individuare le COMPONENTI PERIODICHE NASCOSTE di una SERIE STORICA.

ANALISI DI UNA S.S. STAZIONARIA

↳ CORRELOGRAMMA: è la rappresentazione grafica dei coefficienti di autocorrelazione in funzione dei LAG h

COEFFICIENTI DI AUTOCORRELAZIONE: misurano la correlazione tra i termini della serie storica ed i termini della stessa SLITTATI di h unità di tempo. Misurano la CONCORDANZA / DISCORDANZA tra i valori della s.s. e quelli differiti di h unità di tempo, consentendo di analizzare la sua STRUTTURA INTERNA ossia i LEGAMI tra i termini della stessa.

$r_0 = 1$ $\quad r_h$ assume valori tra -1 e $+1$ $\quad h < s/4$

COEFF. DI AUTOCOVARIANZA → $C_h = \dfrac{\sum_{t=1}^{s-h} (y_t - \bar{y})(y_{t+h} - \bar{y})}{s-h}$ $\quad h = 0, 1, 2, \dots, q < s$; $\quad h < s/4$

COEFF. DI AUTOCORRELAZIONE → $r_h = \dfrac{C_h}{C_0} = \dfrac{\sum_{t=1}^{s-h} (y_t - \bar{y})(y_{t+h} - \bar{y})}{(s-h) \sum_{t=1}^{s} (y_t - \bar{y})^2 / s}$ $\quad -1 \leq r_h \leq 1$

Si calcolano i valori sino al LAG $s/4$
es. se $s = 100$ → LAG 25
si calcolano i r_h da $h=1$ a $h=25$

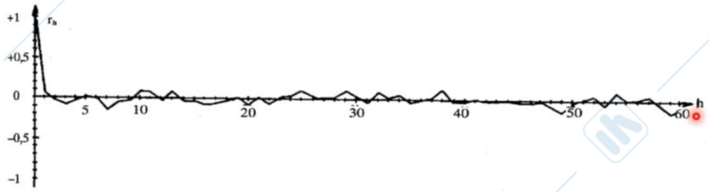
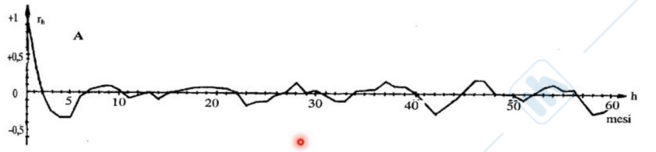


Fig. 16.23.1 - *Correlogramma di una serie di 240 lanci di un dado.*

I VALORI DI r_h OSCILLANO ACCIDENTALMENTE INTORNO ALLO ZERO S.S. COMPLETAMENTE CASUALE : I SUCCESSIVI VALORI SI POSSONO ASSIMILARE AD OSSERVAZIONI INDIPENDENTI (I RISULTATI DÌ UNA SUCCESSIONE DÌ LANCI DÌ UN DADO) E COME TALI SONO TRA LORO NON CORRELATI



AI LAGS PIU' BASSI r_h POSITIVI VIA VIA DECRESCENTI, AI LAGS PIU' ALTI r_h CHE OSCILLANO INTORNO ALLO ZERO : NELLA S.S. VI E' UNA PERSISTENZA DÌ VALORI (UN VALORE ELEVATO E' SEGUITO DA ALCUNI VALORI ELEVATI E UN VALORE BASSO E' SEGUITO DA ALCUNI ALTRI VALORI BASSI)

Serie Territoriali

Le serie territoriali sono esposte in tabelle statistiche le cui modalità sono costituite da unità territoriali, aventi estensioni fissate in base a criteri amministrativi o convenzionali. Talvolta le unità territoriali possono considerarsi di natura puntuale = estensione nulla.

Serie Territoriali tipi:

1) numero dei casi / intensità del fenomeno → unità terr.
2) intensità media fenomeno → unità terr.
3) s.t. doppie: analisi intensità fenomeno in movimento in relazione alle unità terr. di origine e di arrivo
4) s.t. miste

Analisi centrografica

permette di ricavare opportuni centri medi territoriali sotto determinate condizioni. In caso di unità territoriali di tipo non puntuale, serve trovare il baricentro del fenomeno / baricentro demografico / territoriale. Per praticità spesso si usa il CAPOLUOGO

distanza tra 2 punti → $d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$

CENTRI TERRITORIALI:

centro di gravità → dalla condizione che la somma dei quadrati delle sue distanze da tutti gli altri punti è un minimo. → si ottiene $\bar{x} = \frac{\sum_{i=1}^{s} x_i n_i}{N}$ $\quad \bar{y} = \frac{\sum y_i n_i}{N}$

centro mediano → punto in cui sia minima la somma ponderata dei valori assoluti delle sue distanze da tutti gli altri punti →

$\sum_{1}^{s} |d_i| n_i = \sum^{s} \sqrt{(x_i - \bar{x})^2 + (y_i - \bar{y})^2} \cdot n_i = \text{minimo}$

DISPERSIONE TERRITORIALE → grado di dispersione dei valori attorno al centro

indice $\sigma(x,y) = \sqrt{\frac{\sum (x_i - \bar{x})^2 n_i + \sum (y_i - \bar{y})^2 n_i}{N}} = \sqrt{\sigma_x^2 + \sigma_y^2}$

Decomposizione dispersione territoriale → sottogruppi N_k

centri gravità $\bar{x}_k = \frac{\sum x_i n_{ki}}{N_k}$ $\quad \bar{y}_k = \frac{\sum y_i n_{ki}}{N_k}$

scomposizione $Dev(x,y) = \sum_{k=1}^{r} \sum_{i=1}^{s} \left[(x_i - \bar{x}_k)^2 + (y_i - \bar{y}_k)^2 \right] n_{ki} + \sum_{k=1}^{r} \left[(\bar{x}_k - \bar{x})^2 + (\bar{y}_k - \bar{y})^2 \right] N_k$

DISTANZA TRA CIRCOSCRIZIONI DI PARTENZA E DI ARRIVO

$${}^2D^2 = \frac{1}{N} \sum_{i=1}^{s} \sum_{h=1}^{s} \left[(x_i - x'_h)^2 + (y_i - y'_h)^2 \right] n_{ih}$$