Analisi elementare delle serie storiche

Le serie storiche

- Una SERIE STORICA è una serie di valori di una stessa variabile (Y), rilevati nel tempo, cioè in ISTANTI DI TEMPO diversi e, di solito (anche se non obbligatoriamente), tra loro EQUIDISTANTI;
- la "variabile" TEMPO svolge quindi la funzione di "antecedente logico" della variabile Y; quest'ultima è, in qualche modo, DIPENDENTE dal tempo.
- La variabile "tempo" è, in realtà, qualcosa che SOTTINTENDE un insieme di altri fenomeni che influiscono come "cause" sulla variabile Y: se, ad esempio, considerassimo come Y la variabile "arrivi turistici in Italia", la successione dei valori nel tempo di Y sarebbe, di certo, influenzata da una serie di fattori: i fattori climatici, le variazioni nel cambio della Lira (o dell'Euro) rispetto alle altre valute, l'immagine del paese all'estero, la situazione economica dei paesi di origine, la situazione politico-militare internazionale, ecc.

18 e 19 ottobre 2011

Statistica sociale

- Quando si analizza una serie storica, il tempo (che è per sua natura una variabile CONTINUA) viene in qualche modo "trasformato" in una variabile DISCRETA, visto che la variabile viene "colta" nel suo manifestarsi in precisi ISTANTI di tempo. Una serie storica può essere costruita sia a partire da DATI DI FLUSSO che da DATI DI STOCK.
- In questa sede, ci limiteremo a illustrare alcune tecniche elementari di analisi (la cosiddetta "analisi classica").
- Uno degli aspetti più utili delle serie storiche consiste nel "riconoscere", all'interno della dinamica temporale dei dati (in diminuzione oppure in aumento) alcuni "fattori causali" ben riconoscibili, e a volte veri e propri eventi-chiave.
- Mostriamo di seguito la SERIE STORICA DEGLI ARRIVI DI CITTADINI di nazionalità STATUNITENSE alle frontiere italiane, dal 1985 al 1996.

18 e 19 ottobre 2011

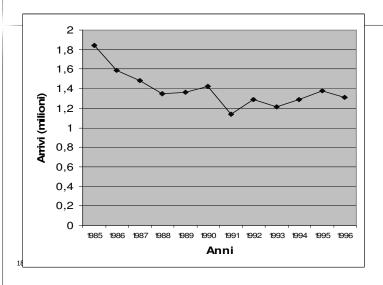
Statistica sociale

3

Anno	Arrivi di cittadini USA alle frontiere (milioni)	Num.indici (base mobile)	Incremento/ decremento	
1985	1,84	100,0	0,0	
1986	1,59	86,4	-13,6	
1987	1,48	93,1	-6,9	
1988	1,35	91,2	-8,8	
1989	1,36	100,7	0,7	
1990	1,42	104,4	4,4	
1991	1,14	80,3	-19,7	
1992	1,29	113,2	13,2	
1993	1,21	93,8	-6,2	
1994	1,29	106,6	6,6	
1995	1,38	107,0	7,0	
1996	1,31	94,9	-5,1	

Statistica sociale



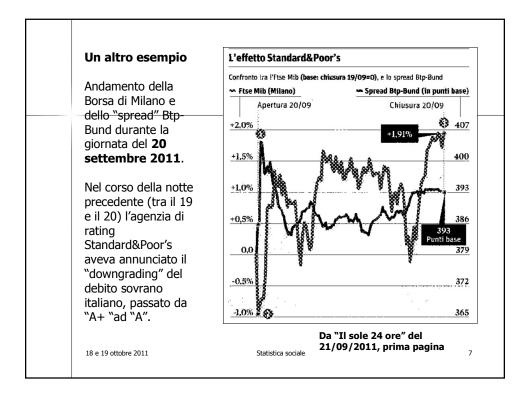


Interpretazione

- Come si può notare, fin dal primo anno considerato (1985) la serie presenta subito una fase di decremento piuttosto accentuato: sono gli anni in cui, dopo che attorno alla metà degli anni '80 il Dollaro si era valutato enormemente nei confronti della Lira (con quote record proprio nel 1985), la divisa americana tornava su binari "normali": si spiega così il decremento, dovuto alla perdita del vantaggio competitivo dell'Italia come meta turistica.
- Inoltre, al dato del cambio lira/dollaro si aggiunge il momento di crisi dell'economia americana che fa seguito al CROLLO DI WALL STREET dell'ottobre 1987.
- Si ha poi una leggera ripresa fino al 1990, seguita poi dal CROLLO (-19,7%) cui si assiste nel 1991, anno della "prima guerra del Golfo", evento che ha agito come deterrente psicologico, spingendo molti cittadini statunitensi a non visitare i paesi dell'Europa meridionale per timore di possibili attentati. Lo stesso timore di attentati può essere stato tra le cause del decremento (-6,2%) del 1993 (anno in cui si sono verificati in Italia alcuni attentati terroristici).

18 e 19 ottobre 2011

Statistica sociale



LE COMPONENTI DI UNA SERIE STORICA

- Una serie storica si può considerare come risultato di un INSIEME COMPLESSO di fenomeni sottostanti. Anche se la presenza di questo aspetto è reale, è molto difficile darne concretamente conto in termini empirici.
- Ciò che si può fare agevolmente è analizzare la DINAMICA INTERNA della serie storica, esaminandone alcuni "comportamenti" tipici rispetto al tempo.
- Un aspetto va precisato: un certo PERIODO di tempo può essere considerato "lungo" o "breve" a seconda della disciplina e/o del contesto al quale i dati fanno riferimento. Se, ad esempio, avessimo a che fare con dati relativi ai mercati finanziari (quotazioni, ecc.), l'arco di una giornata sarebbe già considerato "medio periodo", mentre una settimana sarebbe già un "lunghissimo periodo". Si considera generalmente "breve periodo" un periodo di due-tre anni, mentre per parlare di "lungo periodo" bisogna superare i dieci anni.

LE COMPONENTI DI UNA SERIE STORICA

- Vale in generale il principio per cui, in una serie storica, è presente una componente di fondo che esprime la TENDENZA di lungo periodo. Questa "tendenza di fondo" si dice **TREND** della serie storica. Ciò che "rimane" della serie storica una volta escluso il trend ha generalmente una forma "ciclica", ossia esprime fenomeni che hanno nel tempo un andamento altalenante ("onde").
- Tra gli andamenti di tipo CICLICO, si può infine distinguere un "ciclo" di breve-brevissimo periodo, legato all'andamento delle stagioni (e più in generale dei periodi) ALL'INTERNO DELL'ANNO. Quest'ultima componente, detta STAGIONALITA', è molto importante nell'ambito dei fenomeni turistici.
- Infine e questo è un aspetto comune a tutti i dati statistici vi è nel dato della serie storica una COMPONENTE ERRATICA o CASUALE, dovuta a fattori casuali INTRINSECI AI DATI STESSI.

18 e 19 ottobre 2011 Statistica sociale

9

- Riassumendo, un dato di una serie storica, Y(t), relativo all'istante t, si può considerare quale RISULTATO delle seguenti COMPONENTI:
- T(t) TREND
- C(t) CICLO
- S(t) STAGIONALITA'
- **■** | E(t) COMPONENTE ERRATICA (CASUALE)
- In molte analisi empiriche, soprattutto di tipo economico, può risultare difficile o impossibile distinguere con precisione tra "trend" e "ciclo", o perché si ha a che fare con cicli lunghi che superano la durata del periodo che si sta studiando, o perché non si è certi della durata, o della stessa esistenza, del ciclo. Si parlerà allora un'unica componente denominata *TREND-CICLO*.

Metodi di stima delle componenti: il metodo decompositivo

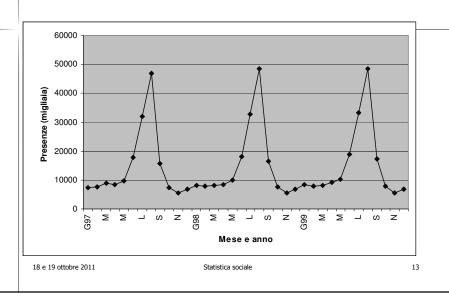
Prendiamo in esame la seguente serie storica, relativa alle PRESENZE TURISTICHE, nel corso degli anni 1997, 1998 e 1999, di TURISTI ITALIANI in Italia, calcolati sulla base delle registrazioni negli esercizi ricettivi (dati Istat). Si tratta quindi di un'analisi che riguarda il cosiddetto "turismo domestico".

18 e 19 ottobre 2011 Statistica sociale 11

Mese e anno	Presenze (migliaia)	Mese e anno	Presenze (migliaia)	Mese e anno	Presenze (migliaia)
Gennaio '97	7438,51	Gennaio '98	8041,63	Gennaio '99	8341,79
Febbraio	7700,7	Febbraio	7770,64	Febbraio	7790,81
Marzo	9015,85	Marzo	8014,09	Marzo	8073,29
Aprile	8262,26	Aprile	8500,27	Aprile	9047,96
Maggio	9580,14	Maggio	10063,17	Maggio	10321,19
Giugno	17732,34	Giugno	18149,79	Giugno	18821,77
Luglio	32028,86	Luglio	32816,46	Luglio	33352,34
Agosto	46822,59	Agosto	48420,47	Agosto	48499,22
Settembre	15792,88	Settembre	16502,49	Settembre	17250,92
Ottobre	7220,36	Ottobre	7529,05	Ottobre	7815,40
Novembre	5484,41	Novembre	5556,65	Novembre	5584,53
Dicembre '97	6777,15	Dicembre '98	6901,37	Dicembre '99	6747,56

Dando una prima occhiata ai dati appare in modo molto chiaro la forte presenza di una *COMPONENTE STAGIONALE*.

GRAFICO relativo alla serie storica



- Abbiamo visto che in questa serie la stagionalità è presente; dobbiamo trovare un modo rigoroso per vedere, in primo luogo, QUALI COMPONENTI sono realmente presenti nella serie storica, in secondo luogo per "isolare" la componente che ci interessa (in questo caso la STAGIONALITÀ), e infine, in terzo luogo, per valutare l'ENTITÀ di tale componente.
- A tale scopo, adotteremo quindi il cosiddetto **METODO DECOMPOSITIVO**, un metodo, cioè, che ci permetterà di scomporre la serie storica nelle varie componenti presenti.
- In particolare, se si è interessati a isolare nella serie la componente "trend-ciclo", il procedimento per eliminare la componente stagionale è detto **DESTAGIONALIZZAZIONE**.

18 e 19 ottobre 2011

Statistica sociale

MODELLI DI ANALISI DELLE SERIE STORICHE

- Finora abbiamo esplicitato le componenti che si suppongono *comporre* una serie storica. Il passo successivo che dobbiamo fare ora consiste nell'enunciazione di un *MODELLO STATISTICO* di analisi, ovvero di una "LEGGE DI COMPOSIZIONE" che permetta di mettere *in relazione* fra loro le varie componenti.
- Un modello è, in generale, un enunciato formale che si suppone aderire, nel miglior modo possibile, alla "forma" empirica dei dati di cui si dispone.

18 e 19 ottobre 2011 Statistica sociale 15

MODELLI DI ANALISI DELLE SERIE STORICHE

- Così, per una serie storica, si può supporre che le componenti siano tra loro sommate, e che seguano pertanto una <u>legge</u> <u>additiva</u>; questo è il caso più frequente, e si avrà allora un <u>MODELLO ADDITIVO</u>; oppure, meno frequentemente, si può ipotizzare che le componenti siano tra loro moltiplicate, e che quindi seguano una <u>legge moltiplicativa</u>: si avrà in questo caso un <u>MODELLO MOLTIPLICATIVO</u>.
- La scelta dell'uno o dell'altro modello dipenderà dal contesto empirico di riferimento, cioè dal fatto che si ritenga che gli effetti delle componenti si combinino sommandosi tra loro (additivamente) oppure moltiplicandosi tra loro (moltiplicativamente).
- Si può ipotizzare anche un **MODELLO MISTO**, in parte additivo e in parte moltiplicativo.

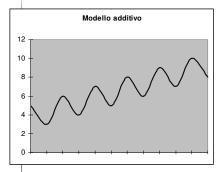
18 e 19 ottobre 2011 Statistica sociale

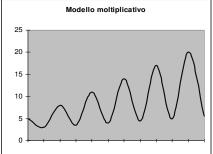
- In termini formali, i tre modelli che abbiamo appena descritto si possono scrivere nel seguente modo:
- Modello additivo:
- | Y(t) = T(t) + C(t) + S(t) + E(t)
- *Modello moltiplicativo:*
- $| Y(t) = T(t) \times C(t) \times S(t) \times E(t)$
- Modello misto:
- $Y(t) = [T(t) \times C(t)] + S(t) + E(t)$

Statistica sociale

17

In termini grafici:





18 e 19 ottobre 2011

Statistica sociale

I due modelli descritti sono formati solo da TREND + CICLO

- Nel primo grafico, possiamo osservare la presenza di un andamento ciclico, ma i cicli RIMANGONO COSTANTI NEL TEMPO, mentre ciò che "cresce" è solo una ipotetica retta che costituisce l'andamento di fondo (trend) della serie.
- Nel secondo grafico, invece, mentre rimane evidente la presenza di un trend crescente, si nota chiaramente come tale trend "moltiplichi" nel tempo anche la dimensione dei cicli, che vanno via via crescendo di intensità al passare del tempo.
- Il primo modello è quindi additivo, perché le componenti (trend + cicli) vi si sommano, mentre il secondo modello è moltiplicativo, perché le componenti (trend+cicli) vi interagiscono moltiplicandosi reciprocamente.

18 e 19 ottobre 2011 Statistica sociale 19

SCOMPORRE UNA SERIE STORICA: LE MEDIE MOBILI E LA STIMA DELLA STAGIONALITÀ

- Uno dei metodi empirici più utilizzati per "scomporre" una serie storica, e soprattutto per depurarla della componente stagionale, consiste nell'applicazione alla serie di una MEDIA MOBILE.
- La media mobile è una MEDIA ARITMETICA che viene applicata ai dati, per un periodo (cioè per un certo numero di termini della serie su cui andrà poi calcolata la media mobile) della serie storica che corrisponde al "ciclo" stagionale; per tornare al nostro esempio, poiché il picco stagionale è su base annuale e riguarda dati a cadenza mensile, si dovrà utilizzare una media mobile A 13 TERMINI.

SCOMPORRE UNA SERIE STORICA: LE MEDIE MOBILI E LA STIMA DELLA STAGIONALITÀ

- Se invece avessimo avuto a che fare con una serie storica giornaliera, che descrivesse un fenomeno dotato di stagionalità settimanale, avremmo allora dovuto utilizzare una media mobile a sette termini.
- Una volta applicate le medie mobili ai dati, questi ultimi andranno infine sostituiti con le medie mobili stesse, e si otterrà così l'effetto voluto di "lisciamento" (SMOOTHING) della serie storica, che non avrà più (o avrà meno) quelle "spigolosità" dovute alla presenza della stagionalità.

18 e 19 ottobre 2011 Statistica sociale 21

A seconda dell'effetto, più o meno accentuato, che si vuole imprimere a tale "lisciamento", una media mobile – relativa al valore t-esimo della serie – può essere semplice, e in questo caso assume la forma:

$$\overline{y}_{t} = \frac{y_{t-k} + \dots + y_{t-1} + y_{t} + y_{t+1} + \dots + y_{t+k}}{2k+1}$$

■ Si noti che il numero a denominatore (2*k*+1) è il doppio del *raggio* della media mobile, sommato di uno, proprio perché si deve tener conto anche del *centro* della media stessa, che è poi quel valore al quale la media si dovrà sostituire.

Una media mobile può invece essere *ponderata*, nel senso che si "aggiunge" alla media mobile una certa STRUTTURA DI PONDERAZIONE, più accentuata attorno al centro della media; una tipica media mobile ponderata può ad esempio essere la seguente:

$$\overline{y}_{t} = \frac{y_{t-k} + ... 2y_{t-2} + 2y_{t-1} + 3y_{t} + 2y_{t+1} + 2y_{t+2} ... + y_{t+k}}{(2k+1) + 2 + 4}$$

■ In questo caso la struttura di ponderazione che è stata applicata alla media mobile è del tipo [2,2,3,2,2]; in alternativa ad una struttura di questo tipo si può applicare alla serie una ponderazione diversa, tale da enfatizzare soprattutto i termini centrali, cioè del tipo [2,3,4,3,2] oppure [2,3,4,5,4,3,2].

18 e 19 ottobre 2011 Statistica sociale 23

- Se la media mobile è ponderata, il numero di termini riportato al denominatore deve essere pari al numero di elementi *realmente* presenti nella serie, cioè deve essere pari a (2*k*+1) più il numero di elementi presenti *in più* rispetto a quelli che si avrebbero in caso di media non ponderata (nel nostro esempio, è facile verificare che gli elementi da sommare sono *sei*).
- In generale, una media mobile ponderata, dando maggiore peso agli elementi della serie più vicini al centro, è **PIÙ CONSERVATIVA** rispetto a una media mobile semplice, nel senso che la serie storica che si ottiene come "output" è più simile alla serie originaria, mentre una media mobile semplice può trasformare anche notevolmente i dati originari.

- Definiamo ora DUE INDICATORI utili per conoscere, in una serie storica, l'intensità del fenomeno di stagionalità.
- Poiché abbiamo adottato un METODO DECOMPOSITIVO, la generica media mobile calcolata in corrispondenza del tempo *t* dovrà intendersi quale stima delle componenti che rimangono se escludiamo la componente stagionale e quella erratica, entrambe almeno approssimativamente eliminate dall'applicazione delle medie mobili.
- Pertanto, trend e ciclo al tempo t sono STIMATI dalla relativa media mobile:

$$\overline{y}_t \cong T_t + C_t$$

Statistica sociale

25

■ Di conseguenza, la stagionalità è stimata in via approssimativa dalla differenza tra il valore empirico della serie al tempo te la relativa media mobile. Questa differenza è detta **INDICE LORDO DI STAGIONALITÀ**:

Indice lordo di stagionalità =
$$y_t - \overline{y}_t \cong S_t + E_t$$

Una misura della stagionalità si può ottenere, oltre che dalla differenza, anche dal *rapporto* tra valore misurato e media mobile. Si parla in questo caso di *QUOZIENTE LORDO DI STAGIONALITA*:

Quoziente lordo di stagionalità =
$$\frac{y_t}{y_t} \cong S_t + E_t$$

18 e 19 ottobre 2011

Statistica sociale

- Lo studio degli indici (o dei quozienti) lordi di stagionalità, e il loro confronto da un anno all'altro è molto utile, perché permette di individuare il TIPO DI STAGIONALITÀ presente nella serie storica.
- Le FLUTTUAZIONI STAGIONALI, infatti, possono mantenersi costanti o variare di forma e/o di intensità nel corso degli anni.
- Se la stagionalità non varia, si parla di MODELLO DI STAGIONALITÀ COSTANTE O RIGIDO; se si ha una variazione di intensità ma non di forma (ad esempio, meno presenze ma sempre nello stesso periodo), si avrà un MODELLO DI STAGIONALITÀ SEMI-RIGIDO; infine, se variano sia forma che intensità (meno presenze, in un periodo diverso), si parla di MODELLO DI STAGIONALITÀ VARIABILE.

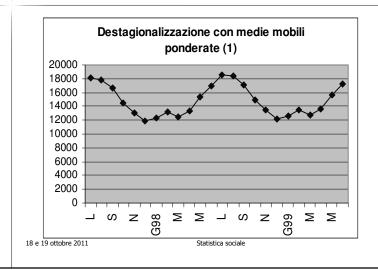
18 e 19 ottobre 2011 Statistica sociale 27

UN ESEMPIO DI DESTAGIONALIZZAZIONE

Riprendiamo ora i dati sulle presenze di turisti di nazionalità italiana in Italia nel triennio 1997-1999. Prima di procedere a un'analisi accurata della stagionalità, mostriamo nei tre grafici che seguono gli effetti che si hanno sulla serie applicando ai dati, rispettivamente:

PRIMO GRAFICO:

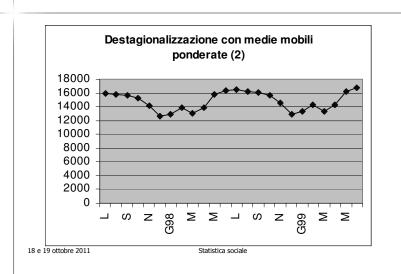
media mobile ponderata, con ponderazione nei *sette* termini centrali, nei quali: con ponderazione 3 nei tre centrali e con ponderazione 2 nei 4 termini rimanenti (1)



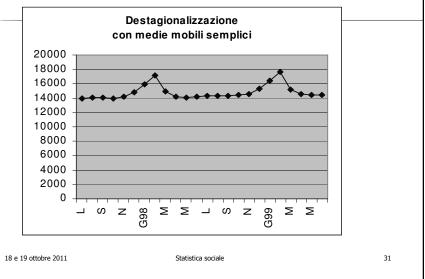
29

SECONDO GRAFICO:

media mobile ponderata, con ponderazione 2 nei *sette* termini centrali (2)



TERZO GRAFICO: media mobile semplice a 13 termini



- Si noti come la componente stagionale rimanga pressoché inalterata nel primo grafico, risulti ancora molto presente nel secondo, e venga completamente eliminata soltanto nell'ultimo grafico, che rappresenta i dati "trattati" con MEDIE MOBILI NON PONDERATE.
- È quindi chiaro che per i nostri dati il metodo più efficace risulta essere l'applicazione di MEDIE MOBILI SEMPLICI.

Mese e anno	Medie mobili	Indici.di stagionalit à	Mese e anno	Medie mobili	Indici.di stagionalit à	Mese e anno	Medie mobili	Indici.di stagional ità
Gen. '97	-	-	Gen. '98	15960,18	-7918,6	Gen. '99	16421,2	-8079,4
Febbraio	-	-	Febbraio	17221,07	-9450,4	Febbraio	17627,57	-9836,8
Marzo	-	-	Marzo	14888,75	-6874,7	Marzo	15229,91	-7156,6
Aprile	-	-	Aprile	14253,08	-5752,8	Aprile	14561,67	-5513,7
Maggio	-	-	Maggio	14125,10	-4061,9	Maggio	14412,09	-4090,9
Giugno	-	-	Giugno	14234,09	3915,7	Giugno	14503,7	4318,1
Luglio	13992,13	18036,7	Luglio	14354,45	18462,0	Luglio	-	-
Agosto	14017,68	32804,9	Agosto	14335,16	34085,3	Agosto	-	-
Settembre	14041,78	1751,1	Settemb re	14358,44	2144,1	Settemb re	-	-
Ottobre	14002,12	-6781,8	Ottobre	14437,97	-6908,9	Ottobre	-	-
Novembre	14140,66	-8656,3	Novembre	14578,04	-9021,4	Novembr e	-	-
Dic. 97	14799,86	-8022,7	Dic. '98	15251,78	-8350,4	Dic. '99	-	-

La serie delle medie mobili da noi calcolata ci permette di ravvisare nei dati la presenza di UN TREND PRESSOCHÉ COSTANTE, con una LEGGERISSIMA CRESCITA appena percettibile.

Statistica sociale

33

Per quanto riguarda la natura della STAGIONALITÀ presente, come si può notare osservando gli INDICI DI STAGIONALITÀ si può senz'altro ipotizzare per il fenomeno da noi investigato un MODELLO DI STAGIONALITÀ QUASI COSTANTE, con una leggera tendenza alla crescita (soprattutto nel mese di agosto).

Mese e anno	Medie mobili	Quoz.di stagional ità	Mese e anno	Medie mobili	Quoz.di stagional ità	Mese e anno	Medie mobili	Quoz.di stagional ità
Gen. '97	-	-	Gen. '98	15960,18	0,50	Gen. '99	16421,2	0,51
Febbraio	-	-	Febbraio	17221,07	0,45	Febbraio	17627,57	0,44
Marzo	-	-	Marzo	14888,75	0,54	Marzo	15229,91	0,53
Aprile	-	-	Aprile	14253,08	0,60	Aprile	14561,67	0,62
Maggio	-	-	Maggio	14125,10	0,71	Maggio	14412,09	0,72
Giugno	-	-	Giugno	14234,09	1,28	Giugno	14503,7	1,30
Luglio	13992,13	2,29	Luglio	14354,45	2,29	Luglio	-	-
Agosto	14017,68	3,34	Agosto	14335,16	3,38	Agosto	-	-
Settembre	14041,78	1,12	Settembre	14358,44	1,15	Settembre	-	-
Ottobre	14002,12	0,52	Ottobre	14437,97	0,52	Ottobre	-	-
Novembre	14140,66	0,39	Novembre	14578,04	0,38	Novembre	-	-
Dic. '97	14799,86	0,46	Dic. '98	15251,78	0.45	Dic. '99	-	-

Statistica sociale

35

■ Le stesse considerazioni si possono fare osservando i QUOZIENTI DI STAGIONALITÀ, anche se ovviamente i quozienti permettono meno – rispetto alle differenze (indici di stagionalità) – di ravvisare nella stagionalità cambiamenti di lieve entità come quelli presenti nella nostra serie storica.

■ Perché questo?

18 e 19 ottobre 2011

- Perché la scelta di utilizzare gli indici o, viceversa, i quozienti di stagionalità dipende dal MODELLO DI ANALISI che si è scelto all'inizio:
- Se il modello scelto è quello ADDITIVO, è opportuno utilizzare gli indici di stagionalità;
- Se, invece, il modello scelto è quello MOLTIPLICATIVO, è opportuno utilizzare i <u>quozienti di stagionalità</u>.

METODI DI STIMA DELLE COMPONENTI: IL METODO ANALITICO

 Per ora, non abbiamo ancora gli strumenti metodologici per affrontare questo argomento, del quale quindi parleremo più avanti.

18 e 19 ottobre 2011

Statistica sociale