

ALWAYS
LEARNING

CAPITOLO 7 – L'analisi delle serie storiche per la programmazione delle attività

Paragrafo 7.2 – Previsioni per mezzo dell'analisi delle serie storiche

Paragrafo 7.3 – Le analisi preliminari e la valutazione delle capacità previsiva dei modelli



Paragrafo 7.2

Le previsioni per mezzo dell'analisi delle serie storiche

Argomenti

- L'impiego dell'analisi delle serie storiche nelle previsioni: impostazione logica
- Le fasi di una analisi delle serie storiche a fini descrittivi e previsivi

Paragrafo 7.2.1

L'impiego dell'analisi delle serie storiche nelle previsioni: impostazione logica

Per **serie storica** (o temporale) si intende una successione di dati osservati su un determinato fenomeno (variabile Y) ordinati secondo la variabile tempo t (per $t = 1, 2, \dots, N$).

La cadenza temporale di osservazione (oraria, giornaliera, settimanale, mensile, trimestrale, annuale, ecc.) può fare riferimento ad una successione di istanti temporali o ad intervalli temporali, (equispaziati o no). Nel primo caso si parla di **serie di stato o posizionali** (ad es. gli addetti di una azienda a fine mese), nel **secondo di serie di flusso** (ad es. la produzione giornaliera).

Le serie storiche presentano, in genere, oscillazioni intorno ad un andamento di lungo periodo che, incluso quest'ultimo, sono state denominate componenti ("virtuali") della serie. Queste, soprattutto nel campo economico, sono di quattro tipi principali.

- i) **Trend** (T): movimento tendenziale monotono di fondo, di lungo periodo, che mette in evidenza una evoluzione strutturale del fenomeno dovuta a cause che agiscono in modo sistematico sullo stesso.
- ii) **Ciclo** (C) o movimento (oscillazione) congiunturale: originato dal presentarsi di condizioni più o meno favorevoli, di espansione e contrazione, del contesto economico nel quale si colloca il fenomeno in esame.
- iii) **Stagionalità** (S): oscillazioni originate da fattori climatici (alternanza delle stagioni) e/o di organizzazione sociale
- iv) **Accidentalità** (e) o componente di disturbo: è data da movimenti irregolari, erratici o accidentali provocati da una serie di circostanze ciascuna di entità trascurabile.

In termini logico formali si può quindi verosimilmente definire la seguente ipotetica relazione per una serie storica:

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

dove $t=1, \dots, N$.

Tenendo presente la possibile esistenza di componenti sistematiche e di disturbo, si ricorda che vi sono **due approcci** (vedi i testi citati) all'analisi delle serie storiche: uno, cosiddetto **classico** (o tradizionale), che assume che il processo rappresentato dalla serie, comprenda una parte deterministica, che consente di stimare le componenti virtuali sopra definite, e una componente di disturbo casuale; l'altro, cosiddetto **moderno**, che assume che la serie sia stata generata da un processo stocastico a componenti correlate descrivibile con appositi modelli probabilistici

In questa impostazione, le **previsioni** (proiezioni, estrapolazioni) che si ottengono si propongono semplicemente di fornire una **informazione** sul probabile **valore futuro** di una **variabile**.

Gli strumenti che presentiamo servono perciò a una elaborazione sistematica delle informazioni disponibili e i risultati delle previsioni non devono necessariamente sostituirsi a ciò che pensa il **manager** ma soltanto **aiutarlo a decidere**.

Di fatto, per effettuare previsioni statistiche ci si troverà a scegliere tra tecniche alternative; è quindi opportuno conoscere le ipotesi che costituiscono la struttura portante dei vari metodi.

Paragrafo 7.2.2

Le fasi di una analisi delle serie storiche a fini descrittivi e previsivi

La realizzazione di una previsione tramite l'analisi delle serie storiche deve essere impostata seguendo la logica di qualsiasi ricerca statistica e, quindi, si sviluppa attraverso le **seguenti fasi**:

1. approfondita **analisi del problema** di previsione da affrontare
2. **raccolta dei dati** e verifica della loro qualità
3. **analisi preliminare** dell'andamento e della struttura della serie storica
4. scelta e **stima del modello**
5. valutazione della **bontà del modello** e sua utilizzazione a fini previsivi.

L'analisi del problema di previsione da affrontare è certamente indispensabile, anche per **avere informazioni a priori sul comportamento evolutivo** che in genere presenta il fenomeno oggetto di studio.

La fase 2 riguarda la possibilità di utilizzare i dati già disponibili in azienda o da raccogliere ex novo sul fenomeno di interesse. E' evidente che in ogni caso occorre **valutare bene la qualità dei dati disponibili** (definizioni, metodi di rilevazione, ecc.) e la loro **comparabilità nel tempo** (per la trattazione di tutti questi problemi rinviamo al Capitolo 2).

La fase 3 riguarda l'analisi della serie dal punto di **vista grafico** e con **indici descrittivi** al fine di evidenziare l'eventuale presenza delle oscillazioni di interesse (trend, ciclo, stagionalità).

La fase 4 ha l'obiettivo di **individuare il modello più adeguato** per la stima delle componenti virtuali della serie e di stimare il modello scelto.

La fase 5 riguarda i **metodi** e gli **indici** per valutare la **bontà del modello** utilizzato e delle eventuali **previsioni** che si desidera effettuare.

Paragrafo 7.3

Le fasi di una analisi delle serie storiche a fini descrittivi e previsivi

Argomenti

- Analisi grafiche preliminari e correlogramma
- La valutazione della bontà del modello e della sua capacità previsiva

Paragrafo 7.3.1

Analisi grafiche preliminari e correlogramma

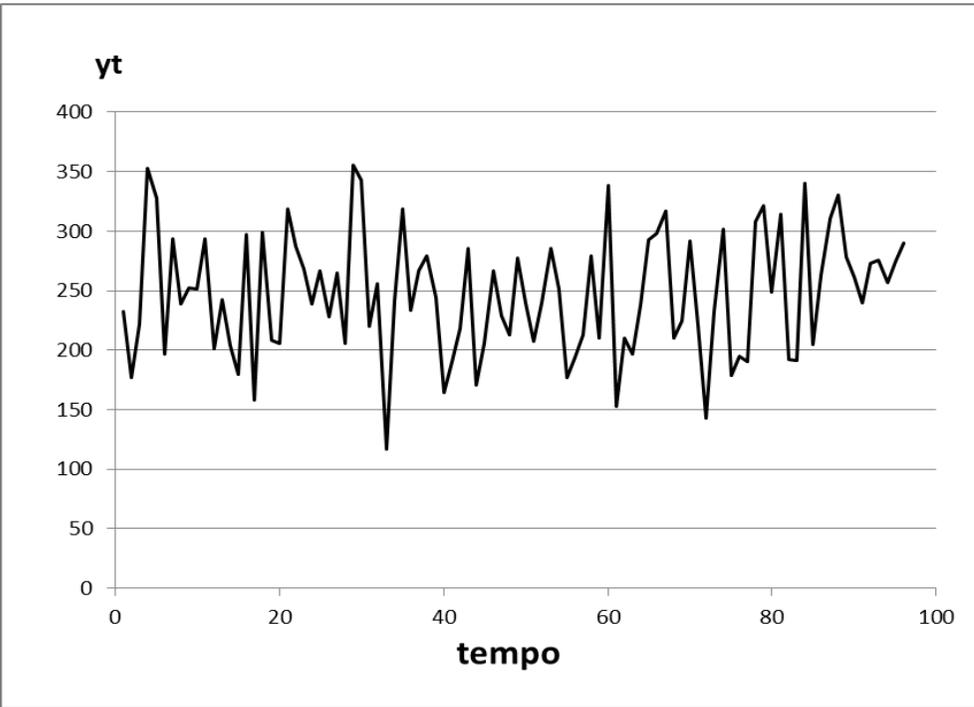
La prima cosa importante da fare quando vogliamo analizzare una serie storica per **verificare l'andamento** e l'eventuale presenza delle **oscillazioni/componenti**, richiamate nel paragrafo precedente, è quella di predisporre un' opportuna **rappresentazione grafica**.

In genere, si inizia con la costruzione e osservazione di un grafico riguardante tutta la serie, detto anche di lungo periodo o **time plot**, in cui vengono riportati i valori del fenomeno osservato Y (in ordinata) in corrispondenza di ciascun tempo t (in ascissa).

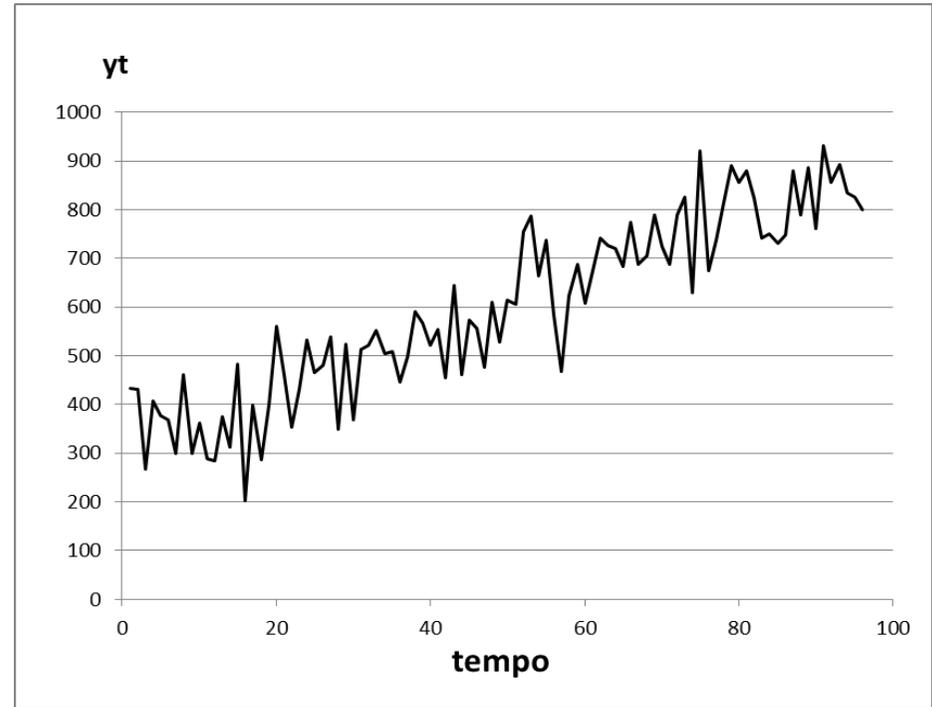
Se nel periodo osservato il livello della serie rimane grosso modo lo stesso ovvero, come solitamente si dice, la **serie è stazionaria in media**, la spezzata del time plot dovrebbe oscillare intorno ad un valore costante uguale alla media della serie (Figura 7.1(a)); viceversa se la **serie è evolutiva** (Figura 7.1(b)), il time plot mette in evidenza il trend (crescente o decrescente, lineare o non lineare).

Figura 7.1 – Esempi di serie storica stazionaria e non stazionaria (evolutiva)

(a) stazionaria



(b) evolutiva



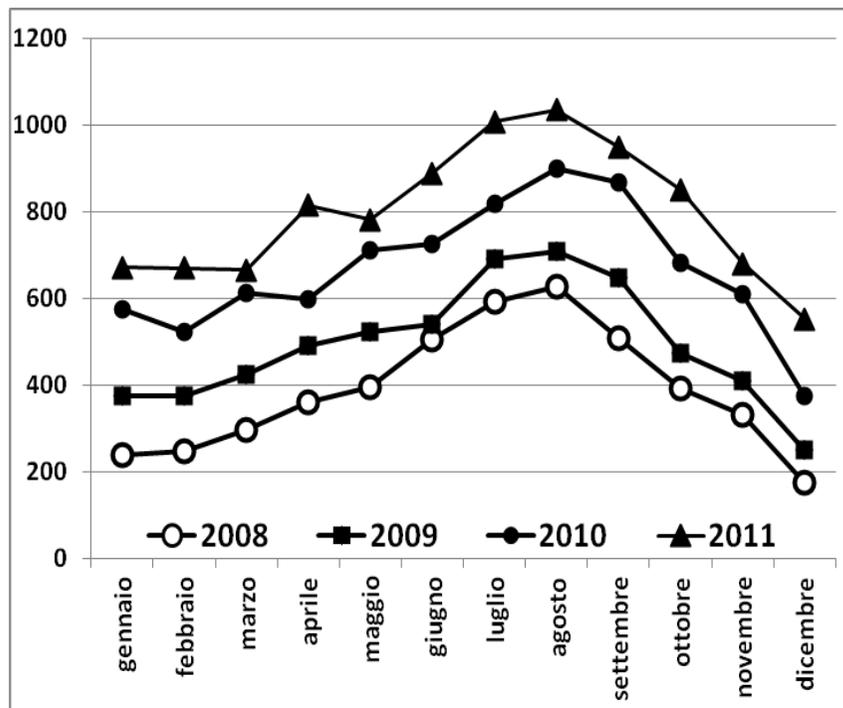
Per verificare la presenza di **oscillazioni stagionali** è utile predisporre il cosiddetto grafico ad anni sovrapposti o **seasonal plot**, che consiste nella rappresentazione dei valori della serie (in ordinata) con riferimento ad un solo periodo annuale (in ascissa) scansionato nei sottoperiodi mensili o trimestrali, ecc. (ma anche giornalieri o settimanali).

Presenza del trend e stagionalità possono essere messe in evidenza anche da un grafico dove in ascissa sono riportati gli anni e in ordinata i valori della serie relativi ai periodi sub-annuali per ciascun anno.

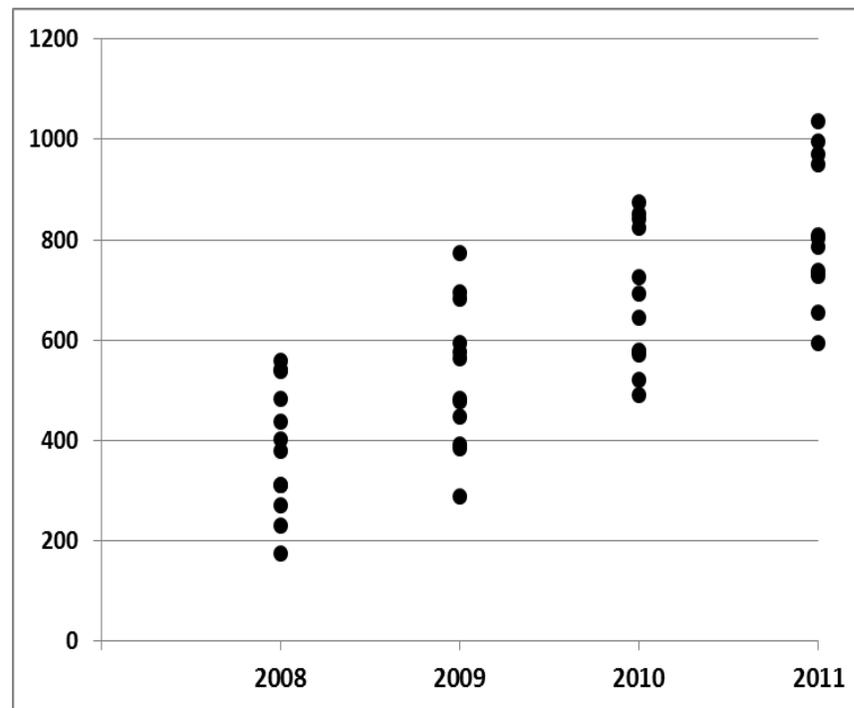
Due esempi di queste rappresentazioni sono illustrati nei grafici che seguono, rispettivamente Figura 2a e Figura 2b.

Figura 7.2 – Grafici stagionali di una serie mensile

(a) seasonal plot
(mesi ad anni sovrapposti)



(b) mesi di ciascun anno

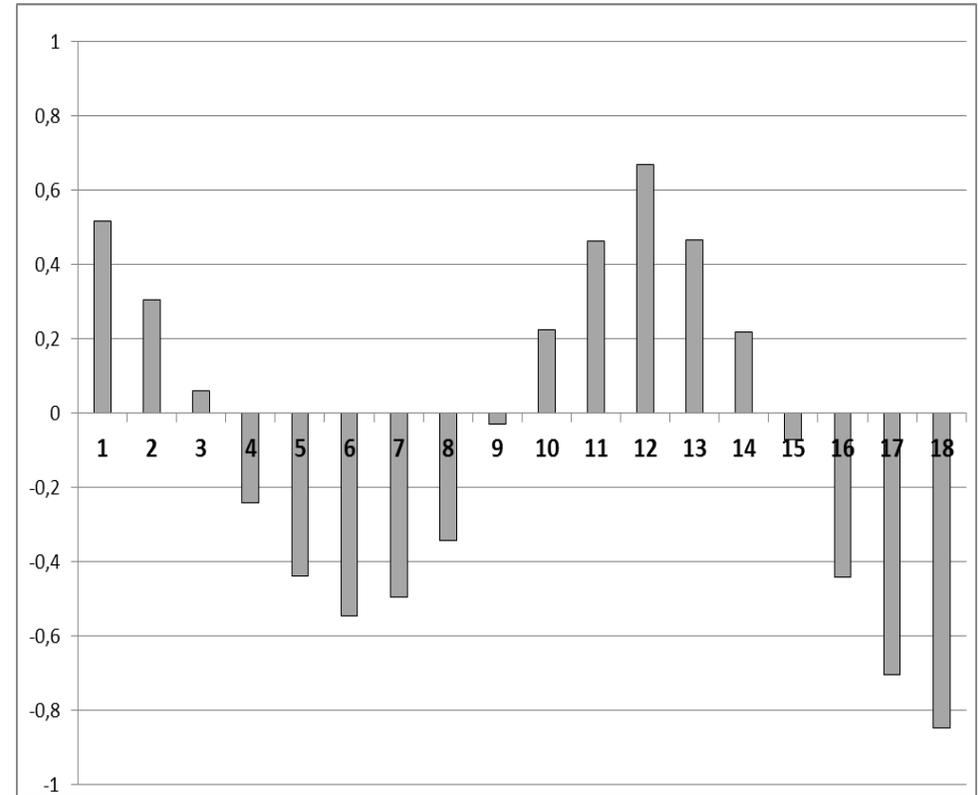
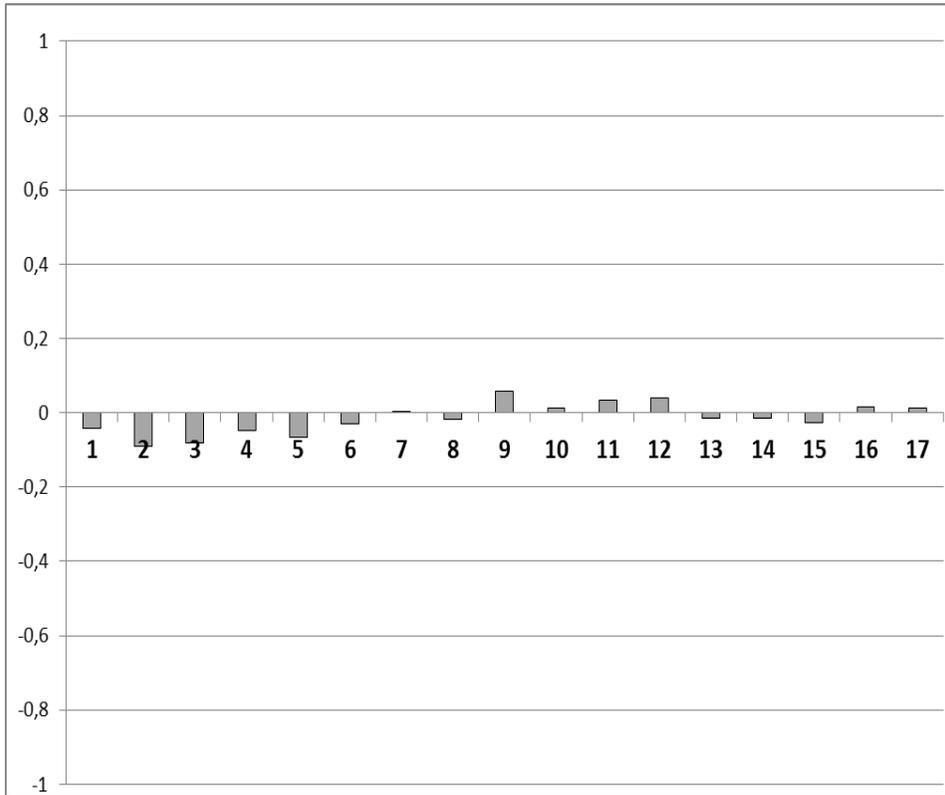


Inoltre per verificare la stazionarietà o meno della serie e la sua “struttura” si ricorre al calcolo di **misure di correlazione** che, nel caso specifico, vengono calcolate tra coppie di dati della serie distanti k periodi.

$$\rho_k = \frac{\sum_{t=k+1}^N (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^N (y_t - \bar{y})^2}$$

I valori del coefficiente di autocorrelazione ρ_k al variare di k , vengono rappresentati in un grafico detto **correlogramma** che illustra le caratteristiche delle oscillazioni presenti nei dati.

Figura 7.3 – Esempi di correlogramma



Esempio 7.1

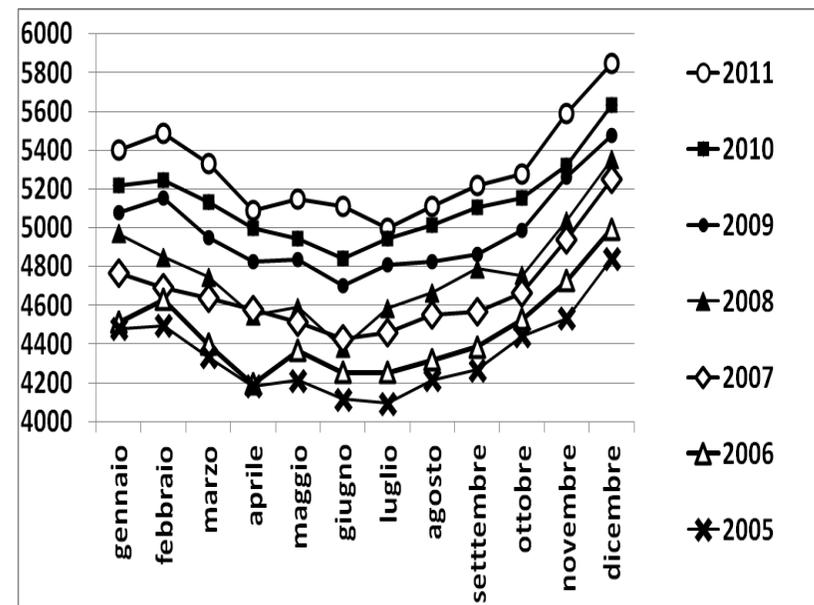
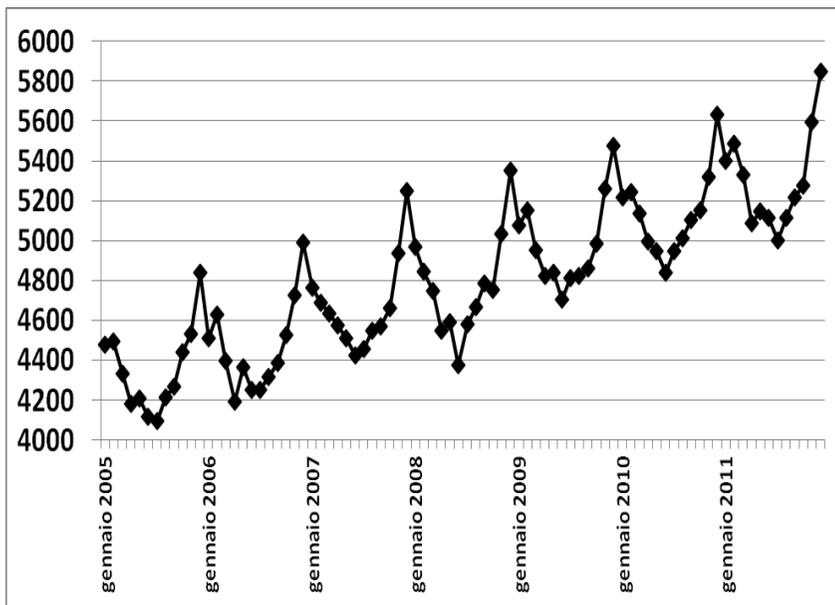
Ipotizziamo che le vendite mensili di un prodotto dell'azienda Alfa siano quelle riportate nella tabella che segue.

Tabella 7.1 Vendite mensili di un prodotto dell'azienda Alfa negli anni dal 2008 al 2011

		gennaio	febbraio	marzo	aprile	maggio	giugno	luglio	agosto	settembre	ottobre	novembre	dicembre
anni	2005	4479	4496	4333	4184	4212	4115	4095	4217	4267	4441	4534	4840
	2006	4510	4630	4400	4195	4367	4252	4252	4318	4386	4526	4726	4992
	2007	4765	4689	4634	4575	4513	4427	4458	4549	4569	4663	4939	5252
	2008	4968	4847	4747	4548	4590	4378	4583	4665	4789	4754	5036	5352
	2009	5077	5151	4951	4826	4837	4703	4811	4825	4862	4986	5262	5476
	2010	5218	5245	5134	4995	4946	4839	4945	5012	5105	5154	5318	5630
	2011	5399	5485	5331	5087	5146	5112	5000	5113	5216	5278	5592	5847

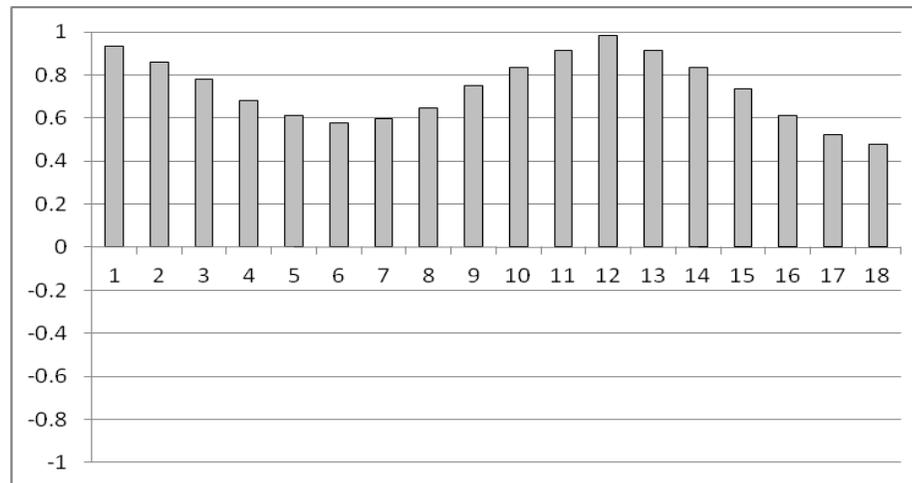
Se i manager della produzione o delle vendite sono interessati a stimare le eventuali componenti della serie e poi ad effettuare delle previsioni, è opportuno in primo luogo svolgere le analisi preliminari sopra indicate e predisporre i grafici time plot, seasonal plot e correlogramma.

Figura 7.4 Time plot e Seasonal plot della serie in Tabella 7.1



Come tutti e tre i grafici mettono chiaramente in evidenza si nota la presenza di trend, stagionalità e accidentalità. Si osservi che il correlogramma, a differenza di quelli prima presentati, mette in evidenza soltanto valori positivi dovuti alla presenza di un trend crescente.

Figura 7.5 Correlogramma della serie in Tabella 7.1



I risultati di queste analisi preliminari indicano quindi che è opportuno stimare le componenti trend e stagionalità che presentano regolarità che dovrebbero essere utili sia per l'organizzazione delle attività di produzione e di gestione delle scorte che per effettuare previsioni delle vendite.

Paragrafo 7.3.2

La valutazione della bontà del modello e della sua capacità previsiva

Nel campo della analisi delle serie storiche a fini previsivi si può parlare di **due tipi di valutazione** riguardanti un modello scelto per rappresentare la serie di interesse.

In primo luogo, si possono stimare (\hat{y}_t), sulla base del modello scelto, i valori teorici della serie e confrontare i dati stimati con i valori osservati verificando come il modello riesce a riprodurre i dati storici. Si parla in questo caso di **goodness of fit** (come per la misura della bontà di adattamento presentata per il modello di regressione nel capitolo 6), e in termini formali si ha che **l'errore di stima** è: $r_t = y_t - \hat{y}_t$

In secondo luogo, interessa verificare come il modello stimato riesce a riprodurre i dati futuri, e si effettua perciò il confronto tra i valori futuri e le previsioni del fenomeno. In questo caso si misura la **goodness of forecast**, che in termini formali indica un **errore di previsione** (ep) al tempo $t+h$ e cioè: $ep_{t+h} = y_{t+h} - F_{t+h}$

Ovviamente **l'errore di previsione**, come sopra definito, si può calcolare solo quando **i dati per i tempi futuri** saranno **disponibili**. Si pone perciò il problema di come si possa **valutare la goodness of forecast in anticipo**, fin dal tempo in cui si fa la previsione, per capire la capacità previsiva del modello.

Nel complesso si avrà:

- una serie di dati disponibili: y_1, y_2, \dots, y_N
- una serie di dati precedenti al tempo n sui quali si adatta il modello da utilizzare per le previsioni: y_1, y_2, \dots, y_m con $m < N$
- una serie di stime per i periodi da 1 a m : $\hat{y}_1, \hat{y}_2, \dots, \hat{y}_m$, che si possono confrontare con i valori osservati y_1, y_2, \dots, y_m per valutare il goodness of fit
- una serie di valori previsti: $F_{m+1}, F_{m+2}, \dots, F_N$, che possono essere confrontati con i valori osservati per lo stesso periodo $y_{m+1}, y_{m+2}, \dots, y_N$ fornendo la possibilità di valutare il goodness of forecast.

La valutazione delle capacità previsive del modello è effettuata per il periodo passato e quindi nell'utilizzare il modello a fini previsivi occorre accettare l'ipotesi, o sperare, che gli errori di previsione abbiano la stessa intensità anche per il futuro.

Si osservi che spesso colui che utilizza le previsioni è interessato a **prevedere** il valore della serie nel **periodo** immediatamente successivo **all'ultimo dato disponibile**. Si tratta della previsione ad un passo (**one-step forecast**) e in questo caso spesso si utilizzano i precedenti $t-1$ dati per effettuare la previsione al tempo t (F_t) e poi successivamente, nel momento in cui sarà disponibile, si aggiunge l'osservazione del tempo t (y_t) per ri-stimare il modello e effettuare la previsione al tempo $t+1$ (F_{t+1}), e così via. Gli errori di previsione vengono in tal modo valutati passo per passo.

Ovviamente interessa avere una valutazione sintetica degli **errori di adattamento (r_t)** e di quelli **di previsione (ep_t)** e a questo fine si utilizzano frequentemente le seguenti misure:

-Errore medio (mean error: ME): media aritmetica degli errori;

$$ME = \frac{1}{m} \sum_{t=1}^m r_t$$

$$ME = \frac{1}{N-m} \sum_{t=m+1}^N ep_t$$

- Errore quadratico medio (*mean square error*: MSE): media aritmetica dei quadrati degli errori

$$MSE = \frac{1}{m} \sum_{t=1}^m r_t^2$$

$$MSE = \frac{1}{N-m} \sum_{t=m+1}^N ep_t^2$$

- Errore medio assoluto (mean absolute error: MAE): media aritmetica degli errori presi in valore assoluto

$$MAE = \frac{1}{m} \sum_{t=1}^m |r_t|$$

$$MAE = \frac{1}{N-m} \sum_{t=m+1}^N |ep_t|$$

Per evitare che le suddette misure dipendano dall'unità di misura della serie, gli errori possono essere trasformati in **errori relativi** (solitamente espressi in percentuale rispetto ai valori osservati), sui quali si calcolano le medie sopra indicate. In particolare, dal MAE si ottiene il MAPE, mean absolute percentage error):

$$MAPE = \frac{100}{m} \sum_{t=1}^m \frac{|r_t|}{y_t}$$

$$MAPE = \frac{100}{N - m} \sum_{t=m+1}^N \frac{|ep_t|}{y_t}$$