



BIBLIOTECA DEI SAPERI

Modelli per Serie storiche

Introduzione (Slide 2)

Benvenuti!

In questa lezione vedremo altri tipi di Dati strutturati, in particolare oggetti per cui il tempo è fondamentale.

Quindi descriveremo:

- che cosa è una Serie storica
- e quali rappresentazioni grafiche possano essere utilizzate per esplorarne le caratteristiche
- infine, discuteremo un semplice modello che può essere utilizzato per l'analisi delle serie storiche

Slide 3 Introduzione: Serie storiche e R

R base ha già alcune funzioni che permettono di analizzare serie storiche. Tuttavia, sono stati sviluppati dei pacchetti specifici per l'analisi delle Serie storiche, quali:

- fpp2 di cui esiste un libro on-line **Forecasting: Principles and Practice** di Rob J Hyndman e George Athanasopoulos
- e un libro **Time Series Analysis and Its Applications with R Examples** di Robert H. Shumway e David S. Stoffer che è accompagnato dal pacchetto `astsa`

Slide 4 È importante distinguere i diversi approcci per tipi di dati diversi, perché quando si descrivono le relazioni esistenti fra variabili, la natura dei dati raccolti ha un impatto importante sulla scelta del modello econometrico più appropriato. Quindi ricordiamo la distinzione tra dati cross-section (dati relativi ad un certo numero di unità economiche osservate in un particolare istante temporale) e dati in Serie storica per cui è probabile che le osservazioni di serie temporali su una data unità economica, osservate in un certo numero di periodi di tempo, siano correlate e per cui i dati hanno un ordinamento naturale in base al tempo

Slide 5 Una Serie storica y_t (notiamo il pedice t anziché i) è una successione ordinata di osservazioni su un certo fenomeno y che si evolve nel tempo registrate in corrispondenza di intervalli di tempo consecutivi, generalmente della stessa lunghezza (ovvero variabili di flusso) o in istanti di tempo consecutivi, generalmente equispaziati (ovvero variabili di stock).

In ogni caso il criterio per ordinare tale successione è il tempo t che va da 1 a T .

Slide 6 Andamento di una Serie storica

Anche per la rappresentazione grafica di una Serie storica si può usare un diagramma cartesiano. Tuttavia, in questo caso sull'asse delle ascisse è riportato il tempo, mentre sull'asse delle ordinate le osservazioni $\$y_t$. In tal modo tutte le coppie ordinate (t, y_t) per t che va da 1 a T possono essere rappresentate. Unendo con una linea i punti otteniamo il grafico della serie utile per interpretare il suo andamento come nell'esempio in cui viene rappresentato l'indice dei prezzi al consumo per il periodo settembre 2017- agosto 2018.

Slide 7 Ordine naturale rispetto al tempo



L'ordine temporale è fondamentale nel caso delle Serie storiche, perché se permutassimo le osservazioni precedenti otterremmo un andamento completamente diverso.

Slide 8 Come creare un oggetto Serie storica in R?

Gli oggetti che abbiamo visto adesso in R avevano una struttura organizzata in righe e colonne. Anche la Serie storica potrebbe avere una struttura simile, tuttavia, per le analisi successive, è necessario passare l'informazione ad R che l'oggetto considerato è una Serie storica.

Per questo si utilizza la funzione `ts()` che richiede tre input principali:

- 1) il vettore o la matrice di valori delle serie storiche osservate (data)
- 2) il periodo della prima osservazione (un singolo numero o un vettore di due numeri interi, che specificano un'unità di tempo naturale) (start) (si può usare end nello stesso modo di start)
- 3) il numero di osservazioni per unità di tempo $s=1$ dati annuali, $s=4$ per dati trimestrali, $s=12$ per dati mensili e $s=52$ per dati settimanali (frequency)

Slide 9 Grafici

Abbiamo già sottolineato l'importanza delle rappresentazioni grafiche in statistica. La rappresentazione grafica delle Serie storiche è ancora di maggiore importanza e deve precedere qualsiasi altro tipo di analisi, perché i grafici illustrano le caratteristiche principali del fenomeno oggetto di studio, mettono in evidenza eventuali osservazioni anomale, suggeriscono possibili spiegazioni di particolari andamenti e variazioni nei dati. Inoltre, per le serie storiche esistono diversi tipi di grafici quali il time plot, il seasonal plot, il sub-series plot, il lag plot.

Slide 10 Time plot I

Il time plot è la rappresentazione grafica che abbiamo già visto in precedenza delle coppie ordinate y_t, t , su un diagramma cartesiano. Può rivelare la presenza di trend, di comportamenti stagionali, di andamenti ciclici o di altre caratteristiche sistematiche. Ad esempio, se i dati oscillano intorno a un livello medio costante, la serie ha un andamento orizzontale (né crescente, né decrescente) nel corso del tempo, le serie sono dette stazionarie in media.

Time plot II

Invece, se i dati mostrano alternanza tra andamenti crescenti e andamenti decrescenti, che non si ripetono con una periodicità costante, la serie ha un andamento ciclico (nella serie storiche economiche le fluttuazioni sono associate al ciclo economico). Slide 11 Inoltre, se i dati sono influenzati da fattori stagionali che tendono a ripetersi quasi esattamente ogni anno, la Serie ha un andamento stagionale (la principale differenza tra andamento stagionale e andamento ciclico sta nella lunghezza del periodo (fissa e inferiore all'anno nel primo caso, irregolare e superiore all'anno nel secondo caso). Le Serie che hanno un andamento di lungo periodo crescente o decrescente sono non stazionarie in media.

Time plot III

Infine, se i dati hanno una dispersione omogenea attorno al livello medio, la Serie è omoschedastica rispetto a serie eteroschedastiche che hanno una dispersione crescente con il livello della Serie.

Slide 12 Esempi di time plot

La Serie in alto a sinistra mostra, ad esempio, un andamento ciclico, mentre quella in basso a sinistra un andamento stagionale. La Serie in alto a destra mostra una forte tendenza crescente che maschera ogni altra caratteristica, mentre quella in basso a destra oltre alla componente di trend ha anche una forte eteroschedasticità perché la dispersione aumenta con il livello della Serie



Slide 13 Le linee rosse aggiunte ai grafici in basso dovrebbero chiarire ancora di più il concetto di eteroschedasticità.

Slide 14 Il time plot fornisce molte informazioni sull'andamento della Serie e consente di descrivere i dati, di stabilire un'eventuale trasformazione dei dati, di scegliere l'approccio adeguato all'analisi e di formulare un modello.

Slide 15 Come rappresentare una Serie storica in R?

I grafici precedenti possono essere prodotti mediante due funzioni in R base `ts.plot()` e `plot.ts()`. La differenza consiste che per Serie storiche multiple `ts.plot` sovrappone le Serie in un unico grafico, mentre `plot.ts` crea un grafico per ciascuna Serie. Nella libreria `fpp2` esiste anche la funzione `autoplot`.

Nel nostro esempio creiamo un vettore `x` che trasformiamo in Serie storica con `ts(x, frequency=12, start=c(2017,9))` e aggiungiamo al grafico un titolo con `ggtitle` e delle label per gli assi con `xlab("Mesi")` e `ylab("Indice")`.

Slide 16 Seasonal e sub-series plot

Per Serie storiche stagionali possono essere utili altre rappresentazioni grafiche capaci di evidenziare meglio le loro caratteristiche.

Consideriamo una Serie storica Y_t di lunghezza T per cui osserviamo n anni completi con frequenza s ($s=4$ per serie trimestrali e $s=12$ per serie mensili). Nel `seasonal plot` le osservazioni registrate nei diversi anni sono rappresentate in maniera sovrapposta

Slide 17 Nel `sub-series plot`, invece, le osservazioni sono rappresentate cambiando la loro sequenza temporale. In particolare per creare un `sub-series plot` è necessario;

- 1) raggruppare le osservazioni in s gruppi (sub-series), uno per ciascun mese o trimestre
- 2) calcolare la media di ciascun gruppo
- 3) e rappresentare, lungo l'asse delle ascisse, i diversi gruppi e le loro medie

Slide 18 Nel grafico a sinistra mostriamo il `Seasonal plot` delle presenze alberghiere, da cui notiamo la forte stagionalità perché per tutti gli anni abbiamo che il mese di agosto registra il valore più alto rispetto agli altri mesi. Il `sub-series` a sinistra evidenzia la media più alta di tutti i mesi di agosto ed inoltre per tutti i mesi consente di avere un'informazione sulla presenza di eventuali valori anomali.

Slide 19 Gli stessi grafici possono essere prodotti per la Serie che nel `time-plot` presentava una forte eteroschedasticità. Nel grafico a sinistra mostriamo il `seasonal plot` da cui notiamo la forte stagionalità nei mesi estivi insieme alla forte componente di trend. Il `sub-series` a sinistra evidenzia le stesse caratteristiche per ciascun mese.

Slide 20 Seasonal plot in fpp2

La libreria `fpp2` rende molto semplice produrre un `seasonal plot` utilizzando la funzione `ggseasonplot` a cui è necessario passare l'oggetto Serie storica in cui abbiamo i nostri dati (in questo caso l'oggetto `a10`)

Slide 21 Subseries plot in fpp2

Altrettanto semplice è in `fpp2` produrre un `subseries plot`. La funzione è `ggsubseriesplot`. In alternativa esiste un altro pacchetto `seasonal`.



Slide 22 Confronto tra andamenti in fpp2

Molto utile per confrontare tra loro l'andamento di diverse Serie è l'opzione `facet=TRUE` in `autoplot()` che consente di fare il grafico di più Serie (nel nostro esempio le prime 5 colonne dell'oggetto `visnights`) su assi multipli.

Slide 23 Correlazione tra Serie

La funzione `ggpairs()` del pacchetto `GGally` permette di costruire una grande matrice di grafici a dispersione. I grafici a dispersione di ciascuna coppia di variabili numeriche sono disegnati nella parte sinistra della figura. La correlazione di Pearson è visualizzata a destra. La distribuzione di ciascuna variabile è disponibile sulla diagonale.

Slide 24 Rappresentazione analitica

Concludiamo questa sezione sulle Serie storiche accennando ai diversi approcci che possono essere utilizzati per analizzare un certo fenomeno in funzione del tempo:

- si possono utilizzare **Modelli deterministici** (approccio classico all'analisi delle Serie storiche) che sono modelli di scomposizione della Serie storica con componenti di natura deterministica, ovvero funzioni non stocastiche del tempo ad eccezione di quella irregolare
- oppure Modelli chiamati ARIMA (approccio moderno all'analisi delle Serie storiche) che permettono di approssimare una classe molto ampia di processi stocastici con un numero ridotto di parametri, studiando la correlazione (autocorrelazione) delle osservazioni
- ed infine Modelli strutturali introdotti da Harvey nel 1990

Slide 25 Approccio classico I

Nell'approccio classico, la Serie storica y_t è pensata come la somma o il prodotto di alcune componenti. In particolare:

- il **trend** T_t , la componente tendenziale che esprime la dinamica di lungo periodo
- la **componente stagionale** S_t , la componente periodica che si ripete (quasi) sistematicamente ogni anno
- e **quella di ciclo** C_t , una componente di breve periodo, ricorrente ma non periodica, con durata intorno all'anno
- infine, è introdotta la **componente irregolare** a_t , una componente puramente casuale e quindi non prevedibile dalle osservazioni passate

La scomposizione può essere **additiva**, in cui le componenti introdotte precedentemente vengono sommate tra di loro, oppure **moltiplicativa** in cui invece la serie osservata è il prodotto delle diverse componenti.

Slide 26 Approccio classico II

In questo grafico, ad esempio, si presenta una possibile scomposizione dell'indice di produzione industriale in cui in alto a sinistra abbiamo il grafico della Serie originale, e in senso orario il ciclo trend, la componente irregolare e quella stagionale. In questo esempio le componenti di trend e ciclo sono state aggregate tra di loro.

Slide 27 Trend e Stagionalità

Per concludere le analisi delle Serie storiche è possibile utilizzare il **modello di regressione** che abbiamo visto precedentemente.

In particolare, utilizzando la funzione `tslm` e le variabili esplicative `trend` and `seas` è possibile stimare la component di trend con una funzione deterministica del tempo (in questo esempio un trend lineare) e delle variabili dummy per tener conto della componente stagionale.

Slide 28 Previsioni



Una volta stimato il Modello con tslm e dopo aver salvato l'output della stima nell'oggetto fit è possibile utilizzare il modello a fini previsivi. Utilizzando la funzione forecast, a cui è necessario dare come input l'oggetto in cui è stata salvata la stima, è possibile produrre un grafico delle previsioni per 20 periodi in avanti ($h=20$).

Conclusioni Slide 29

Bene, siamo giunti alla fine di questa video lezione.

Ti ricordo che abbiamo parlato degli altri tipi di Dati strutturati per i quali il tempo è fondamentale.

In particolare abbiamo visto:

- che cosa è una Serie storica
- e quali rappresentazioni grafiche possano essere utilizzate per esplorarne le caratteristiche
- infine, abbiamo descritto un semplice modello che può essere utilizzato per l'analisi delle serie storiche

Grazie per l'attenzione!