

MODELLI PER SERIE STORICHE

Vedremo



- **cos'è una serie storica**
- **l'Analisi esplorativa: grafici per Serie storiche**
- **un semplice modello per Serie storiche**

Introduzione: Serie storiche e R



- **library(fpp2):** *Forecasting: Principles and Practice*, Rob J Hyndman e George Athanasopoulos
- **library(astsa):** *Time Series Analysis and Its Applications With R Examples*, Robert H. Shumway e David S. Stoffer

Introduzione: Serie storiche



Quando si descrivono le relazioni esistenti fra variabili, la natura dei dati raccolti ha un impatto importante sulla scelta del modello econometrico più appropriato. Innanzitutto è importante distinguere:

- osservazioni in **Cross-section**: dati relativi ad un certo numero di unità economiche osservate in un particolare istante temporale
- osservazioni in **Serie storica**:
 - è probabile che le osservazioni di serie temporali su una data unità economica, osservate in un certo numero di periodi di tempo, siano correlate
 - i dati delle serie temporali hanno un ordinamento naturale in base al tempo

Serie storiche

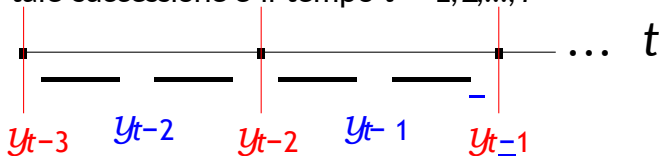


Serie storica (o serie temporale)

Una **Serie storica** y_t 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 (**Variabili di flusso**)
- istanti di tempo consecutivi, generalmente equispaziati (**Variabili di stock**)

Il criterio per ordinare tale successione è il tempo $t = 1, 2, \dots, T$



Andamento di una Serie storica

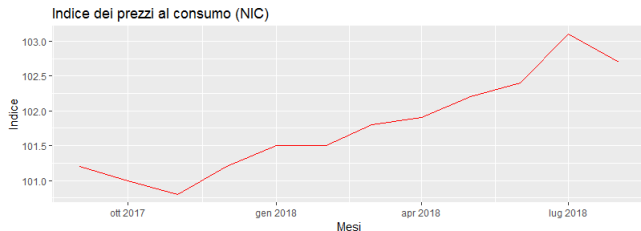


Per la rappresentazione grafica di una Serie storica si usa il diagramma cartesiano, dove le coppie ordinate (t, y_t) , $t = 1, 2, \dots, T$, sono disegnate e unite per interpretare l'andamento della serie

Sull'asse delle ascisse è riportato il tempo t , mentre sull'asse delle ordinate le osservazioni y_t

t	<i>label</i>	y_t
1	<i>set 2017</i>	101.2
2	<i>ott 2017</i>	101.0
:	:	:
$T - 1$	<i>lug 2018</i>	103.1
T	<i>ago 2018</i>	102.7

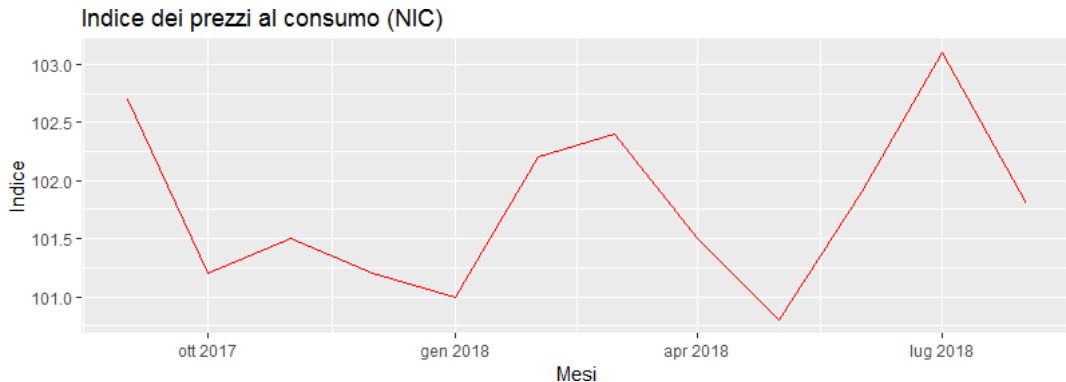
Indice dei prezzi al consumo (NIC - senza tabacchi)



Ordine naturale rispetto al tempo



Una **Permutazione** dell'indice dei prezzi al consumo NIC senza tabacchi - Settembre 2017 / Agosto 2018



Come creare un oggetto Serie storica in R?

- Istruzione `ts()`

```
x = ts(data = NA, start = c(y,p), frequency = s)
```

- **Data:** un vettore o una matrice di valori delle serie storiche osservate
- **Start:** il periodo della prima osservazione (un singolo numero o un vettore di due numeri interi, che specificano un'unità di tempo naturale). Si può usare **end** nello stesso modo di start
- **Frequency:** il numero di osservazioni per unità di tempo ($s = 1, 4, 12, 52$)

Grafici



- La **rappresentazione grafica** dei dati (non solo di serie storiche) è di fondamentale importanza e deve precedere qualsiasi altro tipo di analisi
- 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
- Il **tipo di dati** determina il **tipo di grafico** da utilizzare
- Per le Serie storiche sono generalmente utilizzati i seguenti grafici:
 1. *time plot*
 2. *seasonal plot*
 3. *sub-series plot*
 4. *lag plot*

Time plot I



Il *time plot* è la rappresentazione delle coppie ordinate $(y_t, t), t = 1, \dots, T$ su un diagramma cartesiano. Può rivelare la presenza di trend, di comportamenti stagionali, di andamenti ciclici o di altre caratteristiche sistematiche

Alcune esemplificazioni:

1. se i dati oscillano intorno ad un **livello medio costante**, la serie ha un andamento orizzontale (né crescente, né decrescente) nel corso del tempo (**Serie stazionarie in media**)
2. 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** (nelle Serie storiche economiche le fluttuazioni sono associate al ciclo economico)

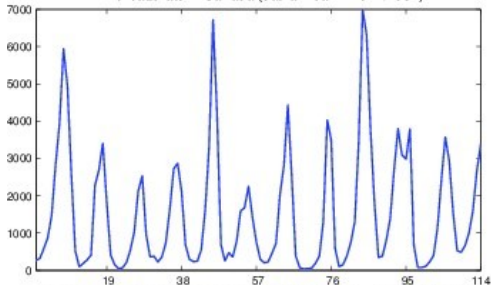
Time plot II



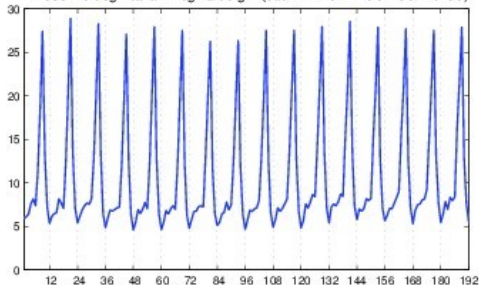
3. se i dati sono influenzati da **fattori stagionali** e tendono a ripetersi quasi esattamente ogni anno, la serie ha un **andamento stagionale** (la principale differenza tra andamento stagionale e andamento ciclico sta nella lunghezza del periodo (fissa e inferiore all'anno nel primo caso, irregolare e superiore all'anno nel secondo caso))
4. se i dati hanno un andamento di **lungo periodo** crescente o decrescente, la serie mostra un **trend (Serie non stazionarie in media)**
5. se i dati hanno una **dispersione omogenea** attorno al livello medio, la serie è **omoschedastica**; se i dati hanno una **dispersione crescente** con il livello della serie, tale serie è **eteroschedastica**

Esempi di *Time plot*

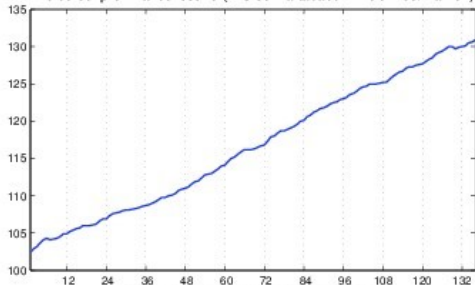
Linci catturate in Canada (dati annuali - 1821/1934)



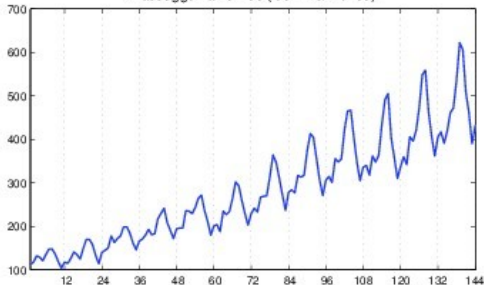
Presenze degli italiani negli alberghi (dati in milioni - Gen. 90/Dic. 05)



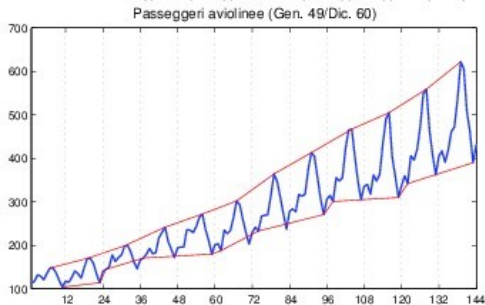
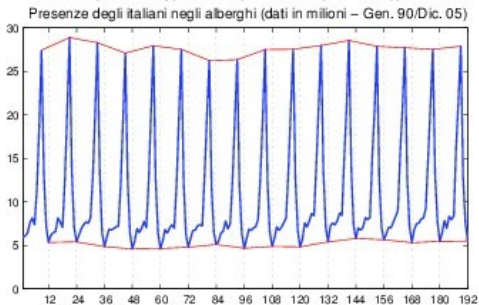
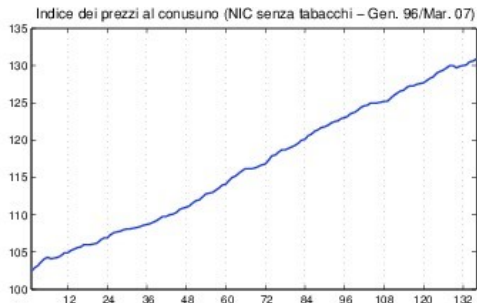
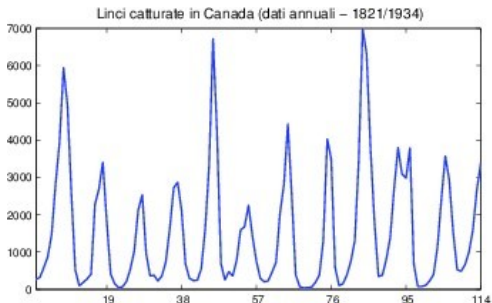
Indice dei prezzi al consumo (NIC senza tabacchi - Gen. 96/Mar. 07)



Passeggeri aviolinee (Gen. 49/Dic. 60)



Esempi di *Time plot*



Obiettivi di un *Time plot*



La disamina del *Time plot* è fondamentale:

- per descrivere i dati
- per stabilire l'eventuale trasformazione dei dati
- per scegliere l'approccio adeguato per l'analisi
- per formulare un modello

Come rappresentare una Serie storica in R?

- Istruzione `ts.plot()`
- Istruzione `plot.ts()`

► Per serie storiche multiple, `ts.plot` sovrappone le serie in un unico grafico, `plot.ts` crea un grafico per serie

```
library(fpp2)
```

```
library(ggfortify)
```

```
x<-c(101.2,101.0,100.8,101.2,101.5,101.5,101.8,101.9,  
102.2,102.4,103.1,102.7)
```

```
ts.x <- ts(x,frequency=12,start=c(2017,9))
```

```
autoplot(ts.x,ts.colour="red") +
```

```
  ggtitle("Indice dei prezzi al consumo (NIC)") + xlab("Mesi")
```

```
  +
```

```
  ylab("Indice")
```

Seasonal e sub-series plot



Per **Serie storiche stagionali** possono essere utili altre rappresentazioni grafiche capaci di evidenziare meglio le loro caratteristiche stagionali
Sia data una serie storica Y_t di lunghezza:

$$T = n \times s$$

dove n è il numero degli anni e s è il numero delle osservazioni registrate in un anno ($s = 4$ per serie trimestrali e $s = 12$ per serie mensili)

- Nel **Seasonal plot** le osservazioni registrate nei diversi anni sono rappresentate in maniera sovrapposta. Per creare un **Seasonal plot** occorre rappresentare n sottoserie, sovrapponendole

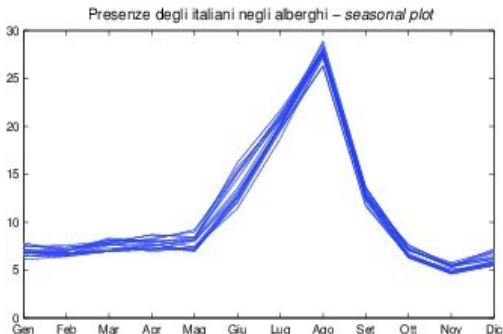
Seasonal e Sub-series plot



Nel *Sub-series plot* le osservazioni sono rappresentate cambiando la loro sequenza temporale. Per creare un *sub-series plot* è necessario:

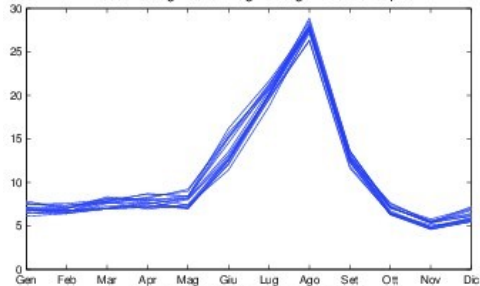
1. raggruppare le osservazioni in s gruppi (\equiv sub-series), una per ciascun mese o trimestre
2. calcolare la media di ciascun gruppo
3. rappresentare, lungo l'asse delle ascisse, i diversi gruppi e le loro medie

Esempi di *Seasonal* e *Sub-series plot*

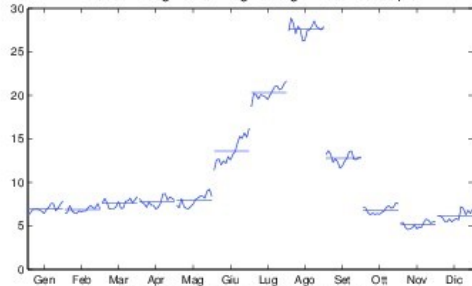


Esempi di *Seasonal* e *Sub-series* plot

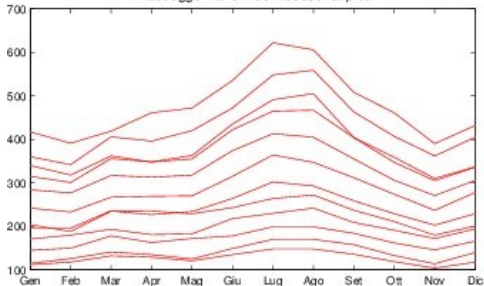
Presenze degli italiani negli alberghi – *seasonal plot*



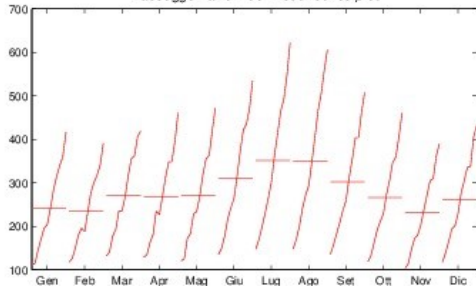
Presenze degli italiani negli alberghi – *sub-series plot*



Passeggeri aviolinee – *seasonal plot*



Passeggeri aviolinee – *sub-series plot*



Seasonal plot in fpp2



```
ggseasonplot(a10, year.labels=TRUE, year.labels.left=TRUE)  
  ylab("$ million") +  
  ggtitle("Seasonal plot: antidiabetic drug sales")
```

```
ggseasonplot(a10, polar=TRUE) +  
  ylab("$ million") +  
  ggtitle("Polar seasonal plot: antidiabetic drug sales")
```

Subseries plot in R



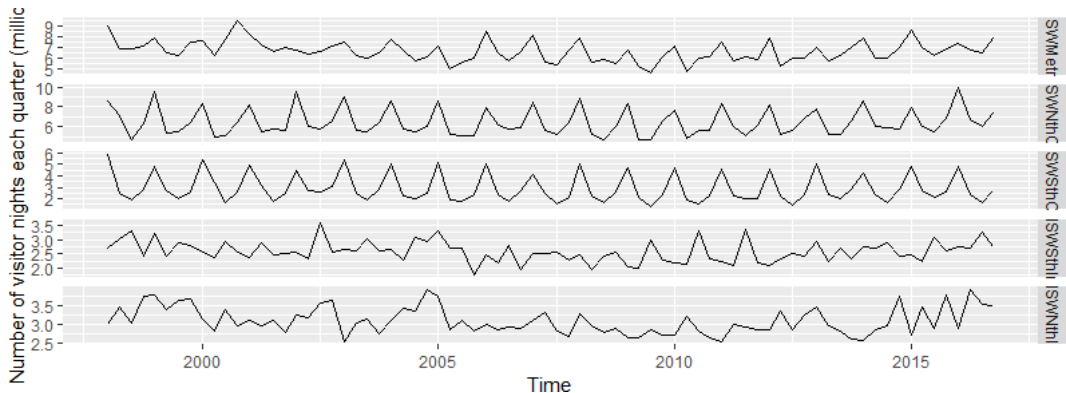
```
library(fpp2)
ggsubseriesplot(a10) +
  ylab("\$ million") +
  ggtitle("Seasonal subseries plot: antidiabetic drug sale
```

```
library(seasonal)
monthplot(AirPassengers)
m <- seas(AirPassengers)
monthplot(m)
```

Confronto tra andamenti in fpp2



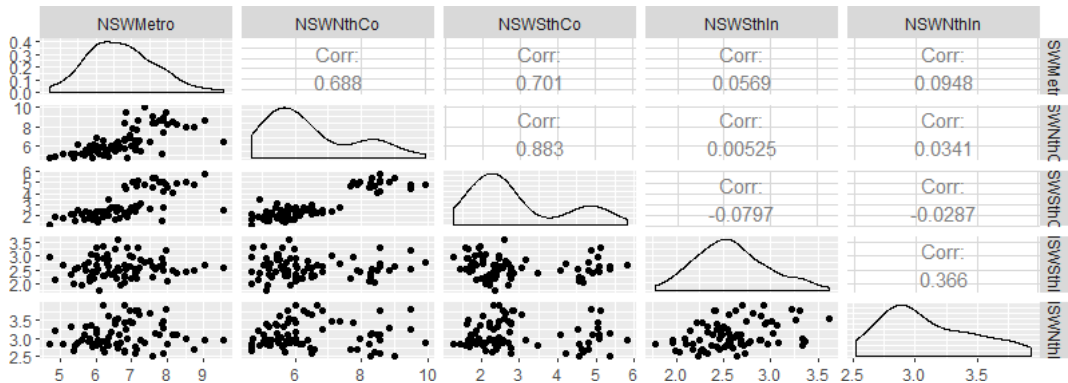
```
autoplot(visnights[,1:5], facets=TRUE) +  
  ylab("Number of visitor nights each quarter (millions)")
```



Correlazione tra serie



```
library(GGally)
ggpairs(as.data.frame(visnights[,1:5]))
```



Rappresentazione analitica

Un certo fenomeno è funzione del tempo
MODELLO STATISTICO

- **Modelli deterministici**

Modelli di scomposizione della Serie storica con componenti di natura deterministica (\equiv funzioni non stocastiche del tempo), ad eccezione di quella irregolare:

$$Y_t = f(t, a_t)$$

- **Modelli ARIMA**

Permettono di approssimare una classe molto ampia di processi stocastici con un numero ridotto di parametri:

$$Y_t = f(Y_{t-1}, Y_{t-2}, \dots, a_t)$$

- **Modelli strutturali (Harvey, 1990)**

Approccio classico I

La serie storica y_t è concepita come la somma o il prodotto delle seguenti componenti:

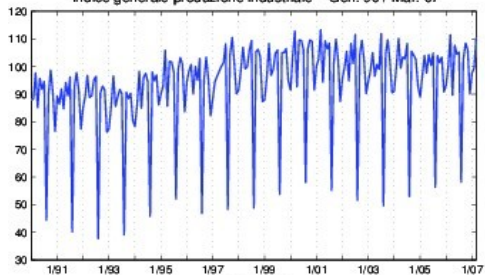
- **Trend T_t :** componente tendenziale che esprime la dinamica di lungo periodo
- **Stagionalità S_t :** componente periodica che si ripete (quasi) sistematicamente ogni anno
- **Ciclo C_t :** componente di breve periodo, ricorrente ma non periodica, con durata superiore all'anno
- **Componente irregolare a_t :** componente puramente casuale e quindi non prevedibile dalle osservazioni passate

$$y_t = T_t + S_t + C_t + a_t \rightarrow \text{scomposizione additiva}$$

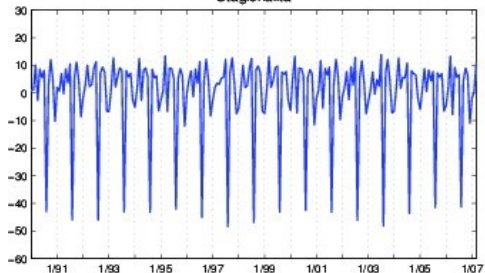
$$y_t = T_t \times S_t \times C_t \times a_t \rightarrow \text{scomposizione moltiplicativa}$$

Approccio classico II - Un esempio di scomposizione

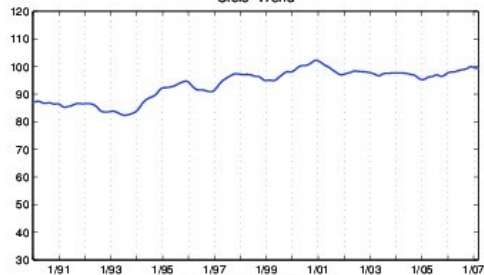
Indice generale produzione industriale - Gen. 90 / Mar. 07



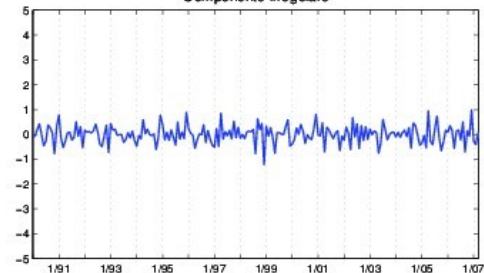
Stagionalità



Ciclo-Trend



Componente irregolare



Trend e Stagionalità

```
> fit <- tslm(AirPassengers ~ trend + season)
```

```
>summary(fit)
```

Coefficients:

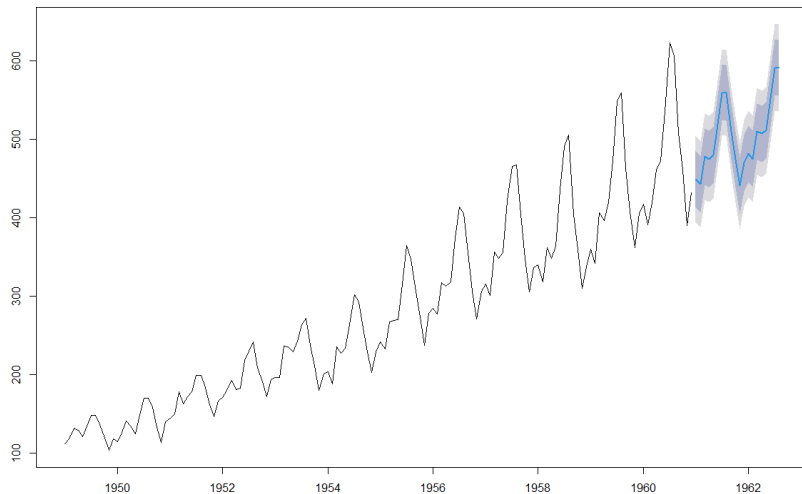
	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	63.50794	8.38856	7.571	5.88e-12	***
trend	2.66033	0.05297	50.225	< 2e-16	***
season2	-9.41033	10.74941	-0.875	0.382944	
season3	23.09601	10.74980	2.149	0.033513	*
season4	17.35235	10.75046	1.614	0.108911	
season5	19.44202	10.75137	1.808	0.072849	.
season6	56.61502	10.75254	5.265	5.58e-07	***
season7	93.62136	10.75398	8.706	1.17e-14	***
season8	90.71103	10.75567	8.434	5.32e-14	***
season9	39.38403	10.75763	3.661	0.000363	***
season10	0.89037	10.75985	0.083	0.934177	
season11	-35.51996	10.76232	-3.300	0.001244	**
season12	-9.18029	10.76506	-0.853	0.395335	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Previsioni



```
> plot(forecast(fit, h=20))
```



Riepilogo e conclusioni finali



- **cos'è una serie storica**
- **l'Analisi esplorativa: grafici per Serie storiche**
- **un semplice modello per Serie storiche**