# PRIMA PARTE: primo tentativo di previsione

X = scan("clipboard")

11849 1316 4712 800 5097 3270 5390 2135 5962 5795

9271 6864 4247 7961 7191 4970 5012 2929 7363 4907

4700 8219 8674 8263 4294 6097 9115 8924 12561 8626

9559 1706 7405 8057 6463 7595 6702 11052 8422 10019

9594 6443 12052 3535 7962 12876 10614 6469 9396 8421

10895 16583 13724 15362 10740 6999 8168 11241 12206 9600

13865 11190 11306 9760 16678 17245 17059 6331 17635 25429

24124 8025 16168 23902 15618 24579 23023 16311 18848 8986

19745 21513 17378 17839 20271 13917 22520 30279 33373 27728

29400 12759 32293 39832 21975 13304 21511 24382 27790 19053

30140 18024 39378 9351 29202 30944 30984 51181 25809 23035

33939 26385 37051 27823 26570 27587 23434 40944 30355 37954

21883 28773 37127 41490 40786 48097 42233 31002 44818 52006

44235 47597 41311 29891 65657 46764 56831 45358 42840 22724

54484 60902 59253 30692 54913 49463 61164 53140 53932 61780

76395 25181 52878 59362 36356 22600 69768 62622 48934 43957

46525 47652 68435 23165 65990 60518 38106 21206

accessori.auto = ts(X, frequency=12, start=c(1995,1))

ts.plot(accessori.auto)



Come sappiamo dall’acf e da decompose, questa serie ha poca stagionalità che aiuti nella previsione. Il trend è accentuato ma negli ultimi anni ha smesso di salire, quindi anch’esso potrebbe essere ingannevole.

Ritagliamo i dati dal 2005 o 2006:

Auto.2005 = window(accessori.auto, 2005)

Auto.2006 = window(accessori.auto, 2006)

ts.plot(Auto.2005)



ts.plot(Auto.2006)



La serie dal 2006 sarebbe più omogenea, ma i dati sono davvero pochi e forse c’è un calo che può far pensare ad un ritorno verso il 2005, quindi prendiamo la prima (tutto molto soggettivo).

acf(Auto.2005)



acf(Auto.2006)



C’è una lieve stagionalità, non molto marcata. Che tipo di ripetitività stagionale c’è?

a.2005 = window(Auto.2005,c(2005,1),c(2005,12))

a.2006 = window(Auto.2005,c(2006,1),c(2006,12))

a.2007 = window(Auto.2005,c(2007,1),c(2007,12))

a.2008 = window(Auto.2005,c(2008,1),c(2008,12))

plot(Auto.2005, col=0)

lines(a.2005,col=1)

lines(a.2006,col=2)

lines(a.2007,col=3)

lines(a.2008,col=4)



C’è un picco negativo ad agosto ed un picco negativo nel periodo invernale, ma non sempre nello stesso mese (forse è questo che deturpa la periodicità). Gli ultimi due anni hanno struttura identica. Se ci basiamo solo su di essi, possiamo calcolarne la media e fornirla come previsione:

Prev1=ts(a.2007, frequency=12, start=c(2009,1))

Prev2=ts(a.2008, frequency=12, start=c(2009,1))

Prev = (Prev1+ Prev2)/2

> Prev

 Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec

2009 62340.5 56042.5 55049.0 48548.5 50228.5 54716.0 72415.0 24173.0 59434.0 59940.0 37231.0 21903.0

Per disegnarla, dobbiamo crearci un po’ artificialmente un plot con l’anno 2009:

XX = c(X, Prev)

 X.prev = ts(XX, frequency=12, start=c(1995,1))

a.glob = window(X.prev,2005)

a.2005 = window(a.glob,c(2005,1),c(2005,12))

a.2006 = window(a.glob,c(2006,1),c(2006,12))

a.2007 = window(a.glob,c(2007,1),c(2007,12))

a.2008 = window(a.glob,c(2008,1),c(2008,12))

a.2009 = window(a.glob,c(2009,1),c(2009,12))

plot(a.glob, col=0)

lines(a.2005,col=1)

lines(a.2006,col=2)

lines(a.2007,col=3)

lines(a.2008,col=4)

lines(a.2009,col=6)



Naturalmente possiamo ripetere l’esercizio usando gli ultimi tre anni, o altre varianti.

# SECONDA PARTE: secondo tentativo di previsione

Infine, adottiamo un atteggiamento molto conservativo: immaginiamo che la struttura annuale ora descritta non ci convinca, mentre crediamo nei valori e basta (senza struttura temporale).

Usiamo un metodo elementare: consideriamo i valori della serie come un campione aleatorio (esperimenti indipendenti), troviamo una densità che gli somigli ed prediciamo il futuro con essa (previsione media più intervallo di confidenza).

Usare tutta la serie è poco sensato. Vediamo anche l’istogramma globale:

 hist(X)

hist(X,20)



C’è una predominanza di valori bassi che sono irrealistici per il futuro, riguardano un periodo di tempo troppo lontano.

hist(Auto.2005,10)



Vale

> mean(Auto.2005)

[1] 46683.25

>

Questa è la previsione media. Forniamo l’intervallo di confidenza simmetrico al 90%.

N=length(Auto.2005)

> 5/100\*N; 95/100\*N

[1] 2.4

[1] 45.6

>

Auto.sort=sort(Auto.2005)

> c(Auto.sort [2], Auto.sort [46])

[1] 21883 68435

>

L’intervallo “empirico” [21883, 68435] contiene il 90% dei valori futuri. Un po’ si vede dal seguente istogramma più fine:

hist(Auto.2005,20)

……………….

Nota. Siccome c’è una parvenza di periodicità, questo secondo metodo sembra inutilmente rozzo. Tuttavia, si deve tener presente che alcuni aspetti così marcatamente ripetitivi negli ultimi due anni erano invece molto diversi negli anni precedenti.



Ad esempio, il mese di gennaio era molto più basso e quello di dicembre più alto. Quindi potrebbe valere la pena, se volessimo spingere di più questo metodo, isolare dei sottoperiodi dell’anno solare ed applicare questa metodologia ad essi separatamente. Per esempio, basterebbe escludere agosto (che può essere previsto per conto suo) ed il risultato sarebbe più realistico, come previsione degli altri mesi. Oppure si potrebbero isolare i mesi invernali e fare una loro previsione separata dal resto. E così via.

# TERZA PARTE: terzo tentativo di previsione

SE = HoltWinters(X,beta=FALSE ,gamma=FALSE)

> SE

Holt-Winters exponential smoothing without trend and without seasonal component.

Call:

HoltWinters(x = X, beta = FALSE, gamma = FALSE)

Smoothing parameters:

 alpha: 0.1610812

 beta : FALSE

 gamma: FALSE

Coefficients:

 [,1]

a 45292.56

>

Il softare ha scelto un parametro molto conservativo.

plot(SE)



> predict(SE,12)

Time Series:

Start = 169

End = 180

Frequency = 1

 fit

 [1,] 45292.56

 [2,] 45292.56

 [3,] 45292.56

 [4,] 45292.56

 [5,] 45292.56

 [6,] 45292.56

 [7,] 45292.56

 [8,] 45292.56

 [9,] 45292.56

[10,] 45292.56

[11,] 45292.56

[12,] 45292.56

>

La previsione somiglia a quella media fatta da noi a mano.

plot(SE, predict(SE,12))



Nota: se usassimo il metodo di media mobile, il risultato sarebbe simile.

A titolo sperimentale, vediamo il grafico forzando il parametro alpha:

plot(HoltWinters(x = X, alpha = 0.01, beta = FALSE, gamma = FALSE))

plot(HoltWinters(x = X, alpha = 0.99, beta = FALSE, gamma = FALSE))