1. **ANALISI DEI RESIDUI**

-----inizio recupero dati-----

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))

plot(accessori.auto)

Y = scan("clipboard")

70.6 121.8 165.1 159.9 210.3 188.8 191.1 41.9 141.7 96.4

67.7 41.1 90.7 113.7 137.1 113.0 117.7 122.6 112.0 19.6

85.7 74.0 71.3 48.3 64.7 101.1 117.9 113.0 139.8 123.8

128.7 27.1 89.9 85.7 72.6 48.0 71.1 94.9 112.1 119.2

123.5 117.0 129.0 23.8 111.9 96.8 82.1 58.2 85.8 110.2

142.2 133.2 142.8 140.9 131.2 28.7 101.5 82.1 69.7 56.1

76.7 86.3 115.8 132.1 147.3 139.9 131.2 41.0 105.5 86.6

86.3 51.3 80.3 120.5 172.4 133.0 169.1 168.3 146.8 38.0

105.1 97.6 88.1 63.7 96.9 120.6 154.7 137.3 166.5 158.9

150.1 35.9 119.0 103.6 89.6 60.6 92.1 118.1 140.5 157.8

157.6 153.3 151.3 17.2 106.7 94.1 70.9 57.9 61.6 83.1

113.0 109.2 121.5 122.4 123.2 18.3 103.1 68.2 57.1 33.1

59.3 100.1 115.2 116.0 126.0 109.2 105.1 10.0 75.0

vendite.moto = ts(Y, frequency=12, start=c(2000,1))

plot(vendite.moto)

-----fine recupero dati-----

HW.auto = HoltWinters(accessori.auto)

plot(HW.auto,predict(HW.auto,12))

HW.moto = HoltWinters(vendite.moto)

plot(HW.moto,predict(HW.moto,12))

-----inizio estrazione residui-----

Grazie all’aiuto di alcuni studenti, è stato individuato il comando “residuals”, che abbrevia moltissimo l’estrazione dei residui. Basta quindi fare:

residuals(HW.auto)

A titolo di confronto, ecco il metodo precedentemente spiegato.

**Esercizio** 1. Osservare il risultato del comando HW.auto$fitted. Che anni riguarda, rispetto a accessori.auto?

**Esercizio** 2. Calcolare i residui, differenza tra i valori veri e quelli stimati dal modello; disegnarne il grafico.

stimati = HW.auto$fitted[,1]

> length(X)

[1] 168

residui = X[13: 168]- stimati

ts.plot(residui)

Confrontare i valori numerici prodotti da residuals(HW.auto) con quelli che si leggono in “residui”.

-----fine estrazione residui-----

-----riepilogo prime analisi residui-----

Nella lezione del 21/11 abbiamo già visto delle prime analisi dei residui, relative ai comandi decompose ed stl: Rieseguiamole qui.

**Esercizio** 3. Visualizzare acf dei residui, osservando se somiglia a quella di un white noise. Visualizzare l’istogramma dei residui e della serie storica. Calcolare la varianza spiegata. Visualizzare il qqplot.

par(mfrow=c(1,2))

acf(residui); acf(rnorm(156))

hist(residui,20); hist(X,20)

1-var(residui)/var(X[13: 168])

qqnorm(residui); qqnorm(rnorm(156))

(Ricordiamo inoltre l’uso del comando qqnorm() serve a valutare graficamente la gaussianità: essa è tanto più stretta quanto rettilineo è il grafico di qqnorm)

-----fine riepilogo prime analisi residui-----

-----stima dell’incertezza tramite i residui-----

I residui possono servire per stimare l’incertezza delle previsioni future.

**Esercizio**. 4. Dare un intervallo di confidenza al 90% non parametrico, sulle predizioni future. Prima farlo numericamente sul primo valore futuro. Poi graficamente sui primi 12 valori futuri.

Soluzione. I quantili non parametrici offerti dal software R si ottengono col comando quantile. Ad esempio:

quantile(residui,0.05); quantile(residui,0.95)

Anche se l’algoritmo preciso con cui essi sono calcolati è meno elementare di quello che si potrebbe pensare (come osservato sperimentalmente a lezione), la base concettuale è quella di ordinare i residui:

round(sort(residui),2)

e poi prendere un valore intermedio tra alcuni di essi identificato sulla base della numerosità della stringa e del livello scelto (es. 0.05).

Li si confronti graficamente con l’istogramma.

Questo però è un intervallo centrato nell’origine, non nella predizione futura. Ad esempio, per la predizione del primo mese successivo:

predict(HW.auto,1)+ quantile(residui,0.05); predict(HW.auto,1)+quantile(residui,0.95)

Si confronti con

par(mfrow=c(1,1))

plot(HW.auto,predict(HW.auto,1))

Oppure si disegni

plot(HW.auto,predict(HW.auto ,12))

lines(predict(HW.auto,12)+ quantile(residui,0.05),col= " green ")

lines(predict(HW.auto,12)+ quantile(residui,0.95), col= " green ")

plot(predict(HW.auto ,12))

lines(predict(HW.auto,12)+ quantile(residui,0.05),col= " green ")

lines(predict(HW.auto,12)+ quantile(residui,0.95), col= " green ")

**Esercizio**. 5. Calcolare la deviazione standard dei residui ed utilizzarla per dare un intervallo di confidenza parametrico, al 90%, sulle predizioni future. Confrontare col metodo non parametrico, anche disegnando in verde le bande non parametriche ed in rosso quelle parametriche.

sd(residui)

ampiezza intervallo al 90%:

qnorm(0.95)\*sd(residui)

oppure

qnorm(0.95,0,sd(residui))

Intervallo:

predict(HW.auto,1)- qnorm(0.95)\*sd(residui)

predict(HW.auto,1)+ qnorm(0.95)\*sd(residui)

plot(predict(HW.auto ,12))

lines(predict(HW.auto,12)+ quantile(residui,0.05),col= "green")

lines(predict(HW.auto,12)+ quantile(residui,0.95), col= "green")

lines(predict(HW.auto,12) - qnorm(0.95)\*sd(residui), col= "red")

lines(predict(HW.auto,12)+ qnorm(0.95)\*sd(residui), col= "red")

Eventualmente, invece che imporre d’ufficio che la media è nulla (cosa che produce un intervallo simmetrico attorno al valore previsto), si possono usare i valori

qnorm(0.05, mean(residui),sd(residui))

qnorm(0.95, mean(residui),sd(residui))

1. **METODI REGRESSIVI PER LE SERIE STORICHE**

Nota. Il software R offre tantissime possibilità automatizzate per costruire modelli di natura regressiva, però ispirati alle categorie AR, ARMA, ecc. Un comando molto semplice ed efficace, per chi volesse provare, è ar.ols Nel seguito sviluppiamo solo un modello semplice fatto “a mano”.

> length(Y)

[1] 129

y=Y[13:129]; x1=Y[12:128]; x12=Y[1:117]

Reg = lm(y~x1+x12)

summary(Reg)

a1 = Reg$coefficient[2]; a12 = Reg$coefficient[3]; b = Reg$coefficient[1]

(controllare con Reg)

P=1:141; P[1:129]=Y

for (i in 1:12) {

P[129+i]=a1\*P[129+i-1]+a12\*P[129+i-12]+b

}

Pplus = ts(P, frequency =12, start=c(2000,1))

plot(Pplus)

(si ricordi plot(vendite.moto)

par(mfrow=c(1,2))

plot(HW.moto,predict(HW.moto,12))

plot(Pplus)