



SCUOLA DI ECONOMIA E STATISTICA

STATISTICA SPAZIALE

Prova di Laboratorio – appello del 27/04/2017

Cognome	Nome
Numero di Matricola	

Esercizio 1

Si consideri la variabile PH e le coordinate dei punti di misura contenute nel file soja98 della libreria geoR.

1. Si calcoli il variogramma empirico della variabile PH usando 20 distanze (bin) per la sua costruzione. Se ne produca il grafico e si riportino i valori del variogramma empirico e i valori delle corrispondenti distanze, approssimati al terzo decimale, negli spazi sotto riportati.

Si commenti l'andamento del variogramma ottenuto.

2. Si stimi il variogramma tramite un modello sferico e un modello esponenziale usando i minimi quadrati ponderati con il sistema di pesi di Cressie, fissando la massima distanza considerata pari a 100 e inizializzando la procedura con il valore 0.08 per la soglia parziale, 0.07 per il nugget e 100 per il parametro di range.

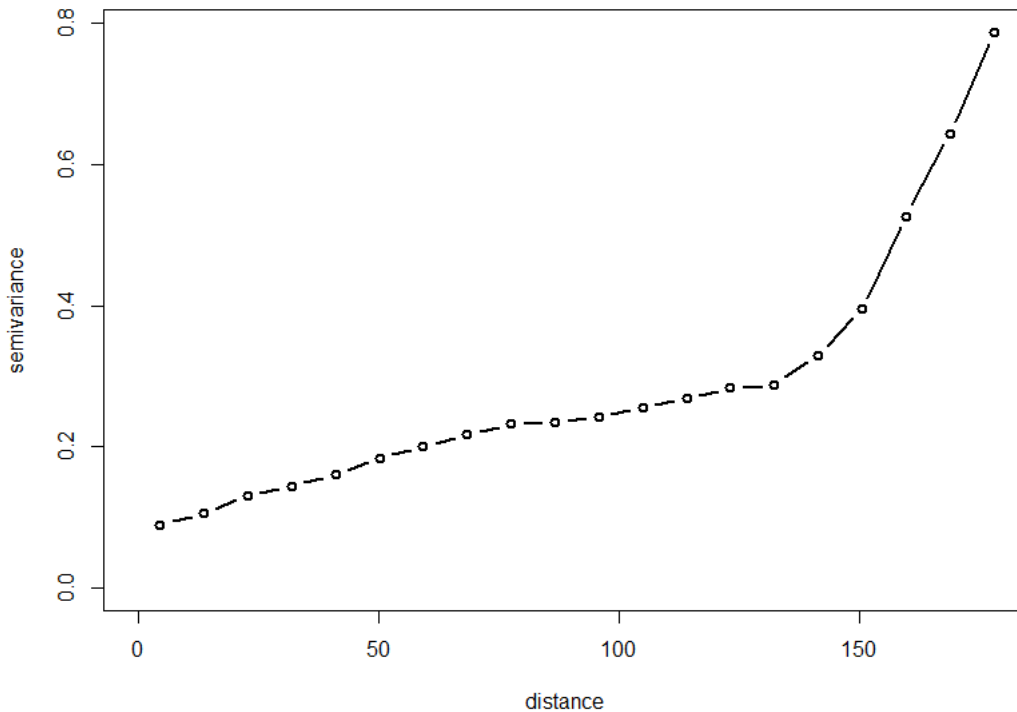
Si riportino il valore del range, della soglia e del nugget stimati per i due modelli. Si calcoli il nugget relativo per il modello sferico interpretando sinteticamente il valore ottenuto.

3. Si riporti il grafico dei due modelli stimati e del variogramma empirico aggiungendo una legenda esplicativa di quanto rappresentato.
4. Si stabilisca quale dei due modelli presenta un fit migliore motivando la risposta fornita.

Si riporti nello spazio preposto il codice utilizzato per produrre la stima del variogramma empirico e il fit wls nei due casi.

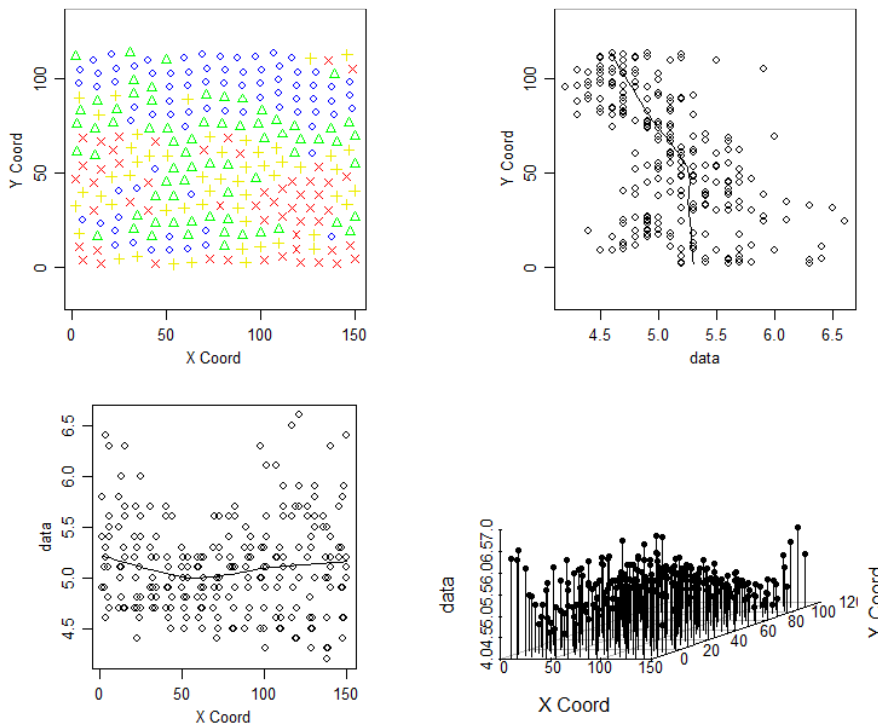
Svolgimento (aumentare lo spazio dedicato alla risposta dove necessario)

Punto 1 Riportare di seguito il grafico del variogramma empirico



Riportare di seguito il commento sul grafico precedente

L'andamento del variogramma evidenzia una componente di nugget non trascurabile e quindi suggerisce la presenza di una componente di variabilità spazialmente non strutturata nella variabile considerata. Il variogramma suggerisce la presenza di un possibile trend a cui potrebbe essere attribuito il deciso aumento dei valori del variogramma empirico che si registrano a distanze elevate. Il trend agisce principalmente nella direzione sud-nord come si evince anche dal seguente grafico. Si osservi comunque che la stima di tali valori di "coda" del variogramma risulta essere non particolarmente attendibile dato il ridotto numero di coppie di punti presenti a tali lag (35 e 7 per gli ultimi due bin del grafico precedente) .



Riportare di seguito il valore del variogramma empirico stimato e le relative distanza

Riportare di seguito i valori di $2\hat{\gamma}(h)$

0.089 0.105 0.131 0.144 0.160 0.184 0.200 0.217 0.232 0.235 0.242 0.255 0.268
0.283 0.288 0.330 0.395 0.527 0.645 0.788

Riportare di seguito i valori di h

4.566 13.697 22.829 31.960 41.091 50.223 59.354 68.486 77.617 86.748
95.880 105.011 114.143 123.274 132.405 141.537 150.668 159.800 168.931 178.062

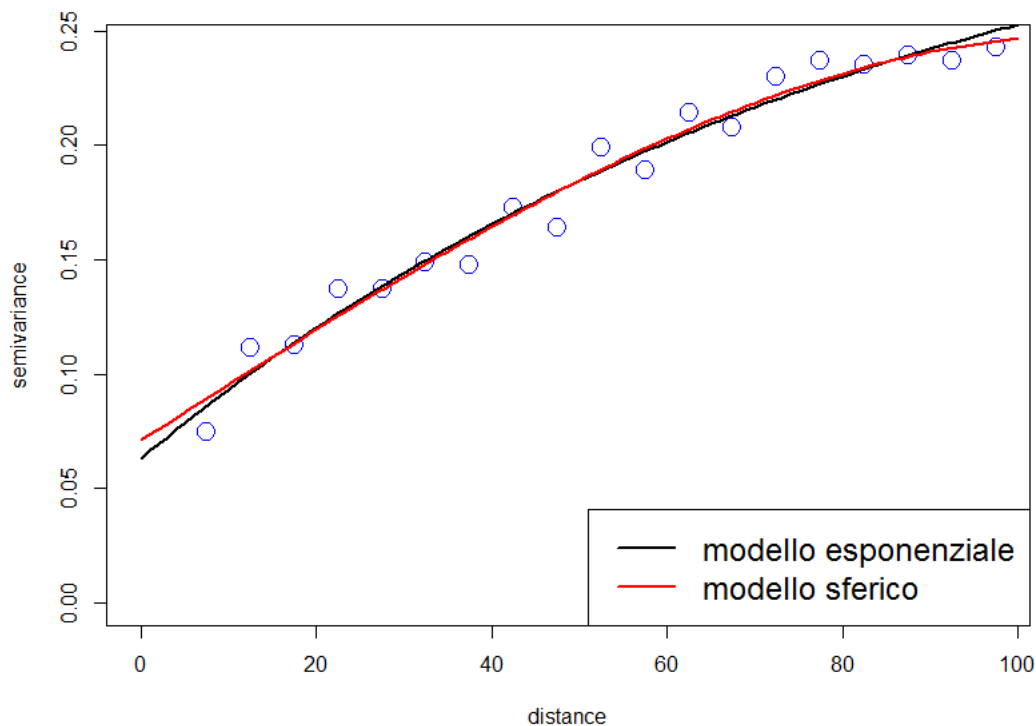
Punto 2 Riportare di seguito le stime dei parametri dei due modelli

	Nugget	Soglia parziale	Parametro di range	Range effettivo
Modello sferico	0.07	0.18	108.85	108.85
Modello esponenziale	0.07	0.30	100.00	299.57

La soglia si ottiene sommando soglia parziale e nugget ...

Il modello sferico presenta un nugget relativo pari a $0.07/(0.07 + 0.18) = 0.28$. Quindi il 28% circa della variabilità del PH non è attribuibile alla struttura spaziale ma ad effetti locali spazialmente non strutturati.

Punto 3 Riportare di seguito il grafico dei variogramma stimati



Punto 4 Riportare di seguito le considerazioni sul fit dei due modelli

I due modelli presentano un fit sostanzialmente equivalente. Il modello sferico può essere preferito in quanto caratterizzato da un miglior adattamento come misurato dalla somma dei quadrati (tabella seguente)

	Somma dei quadrati
Modello sferico	66.88
Modello esponenziale	67.85

Riportare di seguito il codice richiesto

```
require(geoR)
obj <- soja98 [,c("X", "Y", "PH")]
dgeo <- as.geodata(obj, coords.col=1:2, data.col=3)
varioTr <- variog(dgeo, estimator.type="classical",
                 uvec=20, max.dist=100) # variogramma empirico
plot(varioTr, main="", col="blue", cex=2)
varioexp <- variofit(varioTr, ini.cov.pars=c(0.08, 100), weights="cressie",
                   cov.model="exponential", fix.nugget=FALSE, nugget=0.07)
lines(varioexp, lwd=2)
variosph <- variofit(varioTr, ini.cov.pars=c(0.08, 100), weights="cressie",
                   cov.model="spherical", fix.nugget=FALSE, nugget=0.07)
```

Esercizio 2

Il file DiseaseMN.csv riporta dati relativi ad alcuni allevamenti della provincia di Mantova sottoposti ad un test sullo stato di salute del bestiame allevato. Il file contiene tre colonne:

E_2013: esito del test - Positivo, Negativo, non pervenuto

EST: longitudine dell'allevamento

NORD: latitudine dell'allevamento

I file MN.shp, MN.dbf, MN.shx riportano i confini amministrativi della provincia.

1. Si rappresentino su una carta gli allevamenti in cui si è manifestato l'evento come un processo di punto marcato con finestra data dallo shape file e marker identificati dal fatto che l'allevamento sia risultato positivo, negativo o non abbia fornito l'esito del test. Si riporti la mappa nello spazio indicato e si dica quale sistema di coordinate è stato impiegato per la geocodifica dei siti specificando se si tratta di coordinate geografiche o cartografiche e l'eventuale proiezione utilizzata.

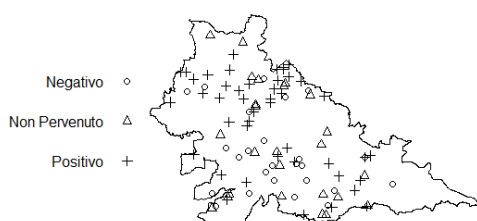
2. Si consideri solamente gli allevamenti con esito "Positivo".

Considerando questo sottoinsieme di eventi si stimi l'intensità del processo rappresentante l'incidenza della malattia sul territorio supponendo che il processo di punto sia un processo di Poisson omogeneo e si interpreti il valore ottenuto.

3. Usando il sottoinsieme dei dati indicato al punto precedente si esegua un test per la CSR basato sulla distanza dal vicino più vicino. Si riportino su un grafico le curve rilevanti per l'analisi, gli involuipi Montecarlo e un'opportuna legenda esplicativa. Si riporti sull'asse di ascissa la variabile in esso rappresentata tramite un'opportuna etichetta. Per l'analisi si eseguano 50 repliche Montecarlo fissando il seme della generazione casuale tramite il valore 1024. Si riporti il codice necessario all'analisi eseguita nei punti 2. e 3. e una breve interpretazione del risultato ottenuto.

Svolgimento (aumentare lo spazio dedicato alla risposta se necessario)

Riportare la mappa di seguito



Come si evince dai valori delle colonne EST e NORD, il sistema di coordinate utilizzato è di tipo cartografico con proiezione UTM.

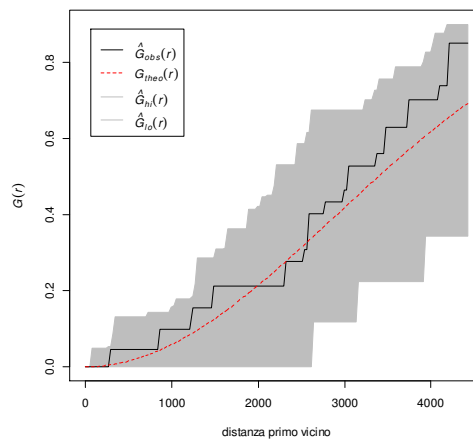
Riportare di seguito il valore dell'intensità stimata e l'interpretazione di tale valore

Poiché l'area del poligono rappresentante la provincia è pari a 2342558779 unità di misura quadrate e ci sono un totale di 46 eventi, allora l'intensità è pari a $46/2342558779$

Il valore dell'intensità è pari a $1.963665e-08$.

Questo valore rappresenta la stima del numero atteso di casi positivi per unità di superficie.

Riportare di seguito il test grafico per la CSR



Riportare di seguito il commento sul grafico precedente

Alla luce del grafico precedente l'ipotesi di csr appare ragionevolmente supportata dai dati in quanto.....

Riportare di seguito il codice R utilizzato ai punti 2. e 3.

```
require(maptools)
require(spatstat)
MN<-readShapePoly("MN.shp", verbose=TRUE)
d=read.table("diseaseMN13.csv", header=T, sep=";")
ppp<-d[d$E_2013=="Positivo", 2:3]
ppp0=as.ppp(ppp, W=MN)
```

```
GGb<-Gest(ppp0, r=NULL, breaks=NULL,correction="none");  
#  inviluppo Monte Carlo per test su CSR  
set.seed(1024)  
envpp<-envelope(ppp0,fun=Gest,nsim=50,verbose=TRUE)  
plot(envpp,main="",xlab="distanza primo vicino")
```