

# Introduction\_to\_spatial\_statistics

April 26, 2018

## 0.1 Lezione :: Analisi statistica spaziale di alcune proprietà dei suoli

### 0.1.1 Laurea magistrale in scienze forestali ed ambientali. Corso di Geografia e Valutazione del suolo

Giuliano Langella glangella@unina.it

Tobler's Law of Geography (1970): "Everything is related to everything else, but near things are more related than distant things".

---

### 0.1.2 Step #1: Come si costruisce un variogramma sperimentale

**Nota:** Il kernel utilizzato per Jupyter è MatLab per cui il codice è in tale linguaggio. Ai fini dello studio degli argomenti trattati è possibile ignorare i comandi MatLab e concentrarsi solo sulla procedura.

---

#### Directory di lavoro:

```
In [24]: cd ~/git/seminars/GeogrValutSuolo/
```

**Rappresentazione cartografica dei punti di campionamento:** crea 2 punti,  $P_1(1,1)$  e  $P_2(5,4)$ :

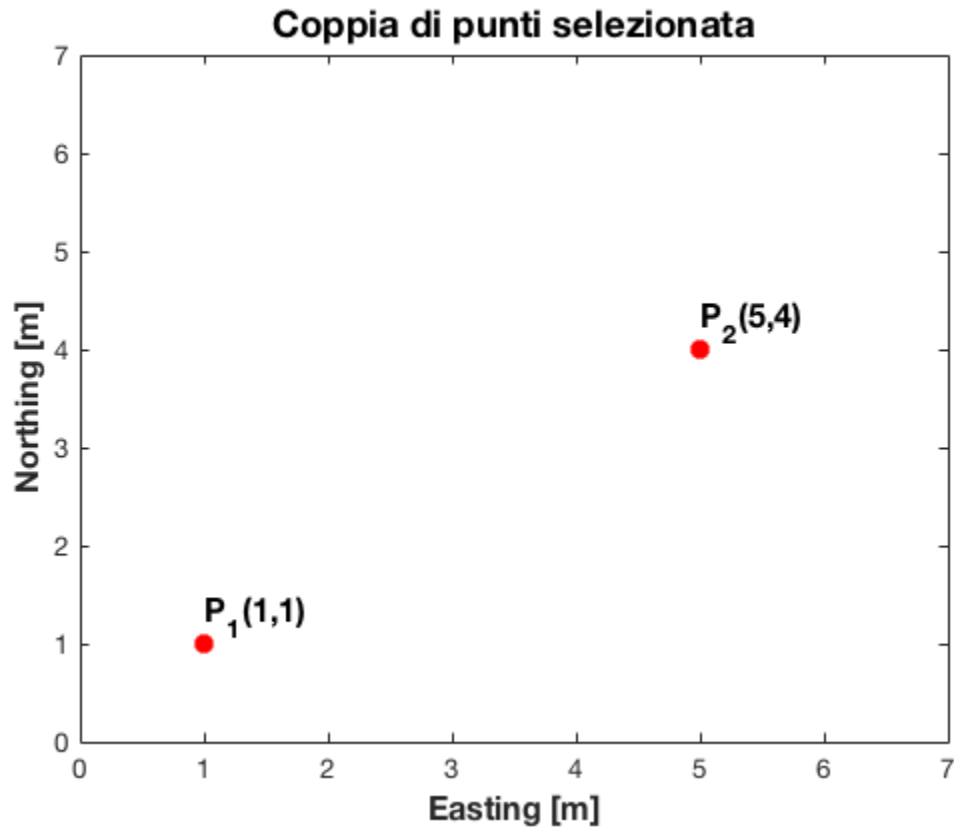
```
In [25]: x=[1,5];
         y=[1,4];
         % P1(1,1)
         P1 = ['P_1(', num2str(x(1)), ',', num2str(y(1)), ')'];
         % P2(5,4)
         P2 = ['P_2(', num2str(x(2)), ',', num2str(y(2)), ')'];

In [26]: % FIGURE:
         figure
         scatter(x,y,450, 'r', 'r'), axis([0,7,0,7])
         xlabel('Easting [m]', 'FontWeight', 'b', 'FontSize', 12)
```

```

ylabel('Northing [m]', 'FontWeight', 'b', 'FontSize', 12)
title('Coppia di punti selezionata', 'FontWeight', 'b', 'FontSize', 14)
text(x(1), y(1), P1, 'HorizontalAlignment', 'left', 'VerticalAlignment', 'bottom', 'FontWeight', 'b', 'FontSize', 12)
text(x(2), y(2), P2, 'HorizontalAlignment', 'left', 'VerticalAlignment', 'bottom', 'FontWeight', 'b', 'FontSize', 12)
box on

```



Crea 100 punti geospaziali di coordinate  $P_i(X_i, Y_i)$  casuali. (funzione MatLab RANDI) Crea una vettore di 100 punti (ossia matrice di dimensioni 100x1) con coordinate casuali (randi) ed intere (randi):

X nel range [1,550]

Y nel range [1,700]

La divisione per 100 obbliga i valori tra 0 e 5.50 / 7.00

```

In [27]: X=randi(550,100,1)/100;% con intervallo da 0 a 5.50
Y=randi(700,100,1)/100;% con intervallo da 0 a 7.00
fprintf('%14s, %14s\n','Easting [m]','Northing [m]')
for ii=1:5 fprintf('%14.3f, %14.3f\n',X(ii),Y(ii)); end
%[X(1:5),Y(1:5)] % osserviamo i primi 5 valori dei vettori coordinate X ed Y:

```

Easting [m],	Northing [m]
0.330,	2.970
3.760,	0.660

0.240,	4.190
0.400,	3.300
2.870,	4.880

Crea una vettore di 100 punti (ossia matrice di dimensioni 100x1)

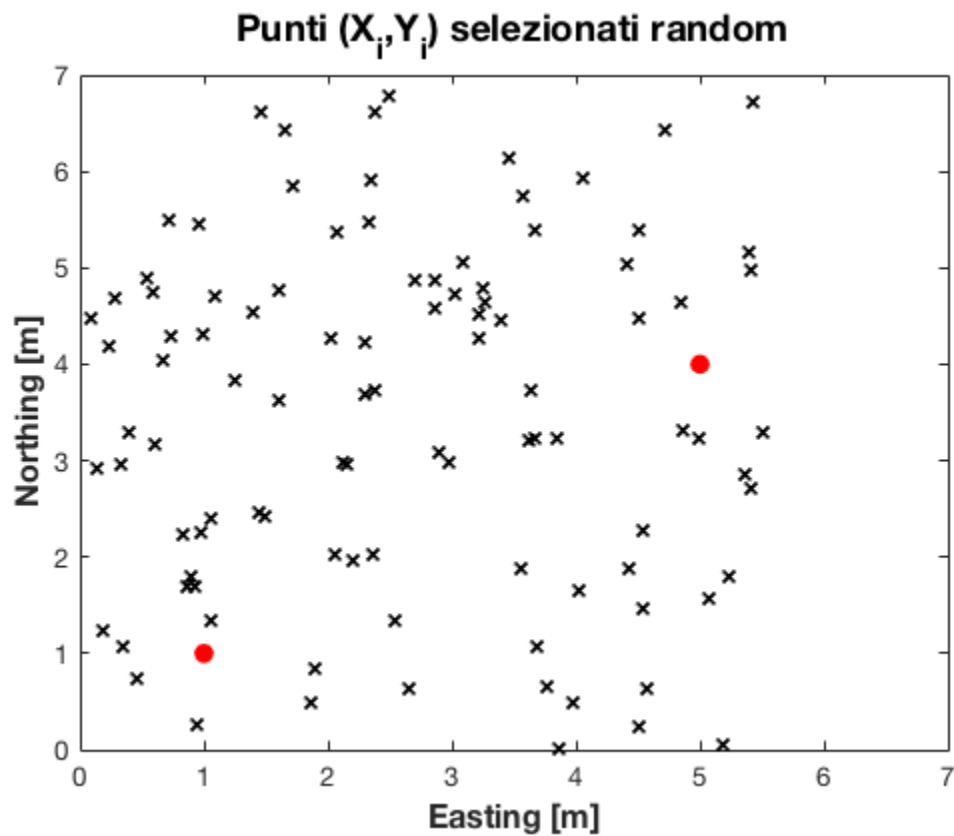
con valori casuali di Carbonio Organico  $\left[\frac{g}{kg}\right]$

nell'orizzonte superficiale (es.  $A_p$  o profondità fissa 0-40 cm)

La divisione per 1000 obbliga i valori tra 0.000 e 2.000  $\frac{g}{kg}$

In [28]: `Z=randi(2000,100,1)/1000;% con intervallo da 1 a 2.000 g/kg`

```
In [29]: figure,hold on
scatter(X,Y,07,'x','k'),axis([0,7,0,7])
scatter(x,y,450,'.','r'),axis([0,7,0,7])
hold off
xlabel('Easting [m]','FontWeight','b','FontSize',12)
ylabel('Northing [m]','FontWeight','b','FontSize',12)
title('Punti (Xi,Yi) selezionati random','FontWeight','b','FontSize',14)
box on
```



**Note sul campionamento:** Area/Scala di indagine;  
Localizzazione e densità dei punti di campionamento;  
Tipologie di campionamento (profili, pit, trivellate, osservazioni, ...);  
Scopo del campionamento: osservare l'ubicazione ed i tipi di suoli nell'area di indagine;  
Raccolta di campioni di suolo, per ciascun orizzonte, per le analisi fisico-chimiche;  
C'è un lavoro preliminare per ottenere dati sui suoli utili alle indagini (geo-)statistiche;

**Richiamo delle nozioni cartografiche:** Meridiani (Greenwich) e Paralleli (Equatore);  
Geoide vs Ellissoide (WGS84);  
Sistema di coordinate (Geocentrico, Geografico, Piano);  
Trasformazioni tra sistemi;  
Datum (WGS84, Roma40, ...):

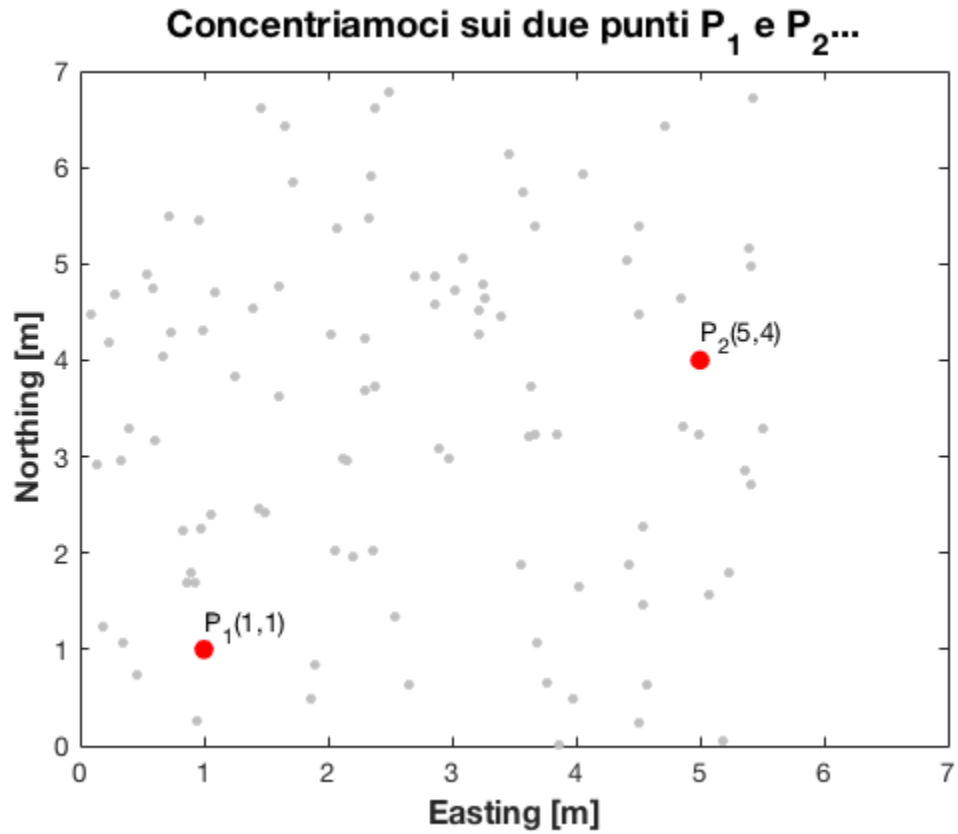
```
<li> scegliere lellissoide, </li>  
<li> eseguirne una materializzazione, </li>  
<li> orientarlo in un punto (verticale corrispondente alla normale), </li>
```

Le analisi statistiche spaziali richiedono sistemi di riferimento piani (proiezioni)!

**Sistemi di riferimento:** geografico, Globale [EPSG:4326]  
piano, Italia Fuso Ovest [EPSG:32632]  
piano, Italia Fuso Est [EPSG:32633]

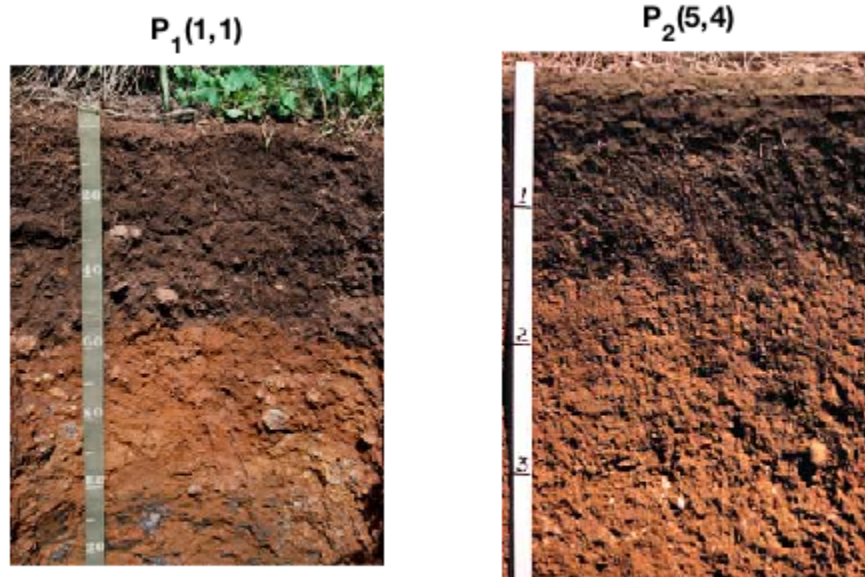
**Riprendiamo i due punti a scelti a caso di cui prima,  $P_1$  e  $P_2$ ...**

```
In [30]: hold on  
         scatter(x,y,450,'.','r'),axis([0,7,0,7]),xlabel('Easting [m]'),ylabel('Northing [m]')  
         scatter(X,Y,16,[.75 .75 .75],'filled'),axis([0,7,0,7])  
         hold off  
         xlabel('Easting [m]', 'FontWeight','b','FontSize',12)  
         ylabel('Northing [m]', 'FontWeight','b','FontSize',12)  
         title('Concentriamoci sui due punti P_1 e P_2...', 'FontWeight','b','FontSize',14)  
         text(x(1),y(1),P1, 'HorizontalAlignment','left', 'VerticalAlignment','bottom')  
         text(x(2),y(2),P2, 'HorizontalAlignment','left', 'VerticalAlignment','bottom')  
         box on
```

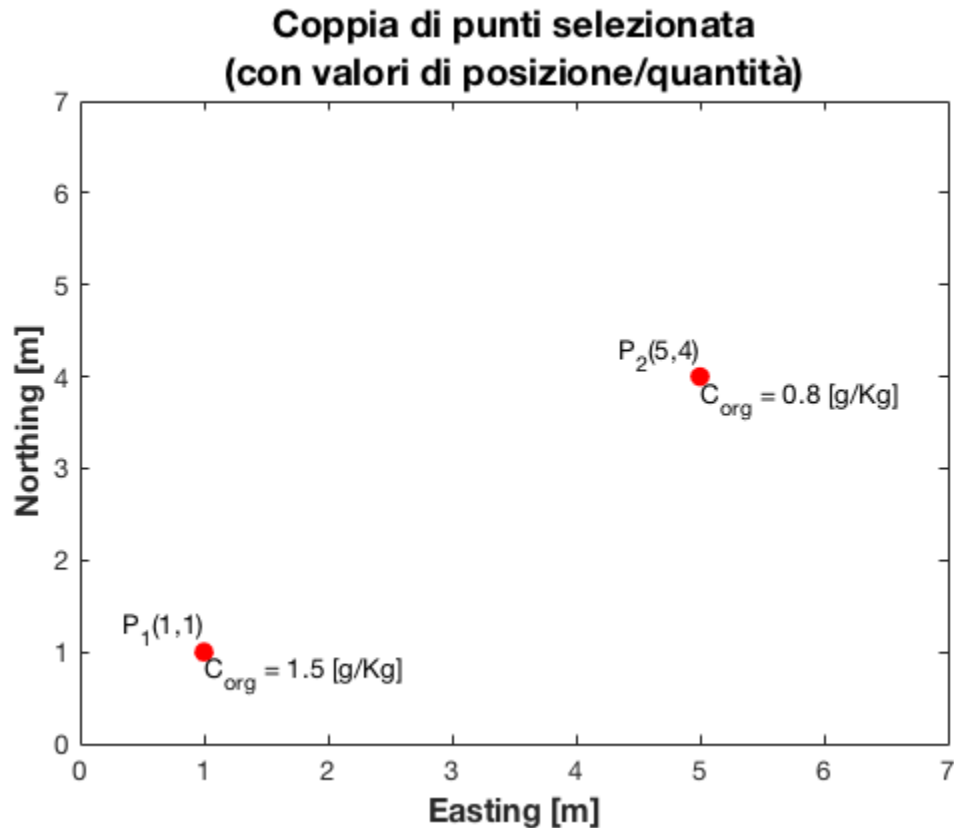


...ciascuno di essi avrà un valore specifico di una proprietà del suolo: ad esempio per il Carbonio Organico (orizz.  $A_{(p)}$ ):

```
In [31]: subplot(121)
         imshow('artwork/P1.jpg'),title(P1)
         subplot(122)
         imshow('artwork/P2.jpg'),title(P2)
```



```
In [32]: z1 = 1.5;% [g/Kg]
         z2 = 0.8;% [g/Kg]
         figure
         scatter(x,y,450,'.','r'),axis([0,7,0,7])
         xlabel('Easting [m'],'FontWeight','b','FontSize',12)
         ylabel('Northing [m'],'FontWeight','b','FontSize',12)
         title({'Coppia di punti selezionata';'(con valori di posizione/quantità)'},'FontWeight'
         text(x(1),y(1),P1,'HorizontalAlignment','right','VerticalAlignment','bottom')
         text(x(2),y(2),P2,'HorizontalAlignment','right','VerticalAlignment','bottom')
         text(x(1),y(1),['C_{org} = ',num2str(z1),' [g/Kg]'],'HorizontalAlignment','left','Verti
         text(x(2),y(2),['C_{org} = ',num2str(z2),' [g/Kg]'],'HorizontalAlignment','left','Verti
         box on
```



**0.1.3 Obiettivo** —> calcolare la distanza tra i due punti  $P_1$  e  $P_2$

POSIZIONE :: Distanza geospaziale tra i due punti

QUANTITA' :: Distanza relativa alla variabile (proprietà del suolo | C.O.) misurata nei due punti)

**1. Calcoliamo la distanza euclidea tra i due punti geospaziali  $P_1$  e  $P_2$**

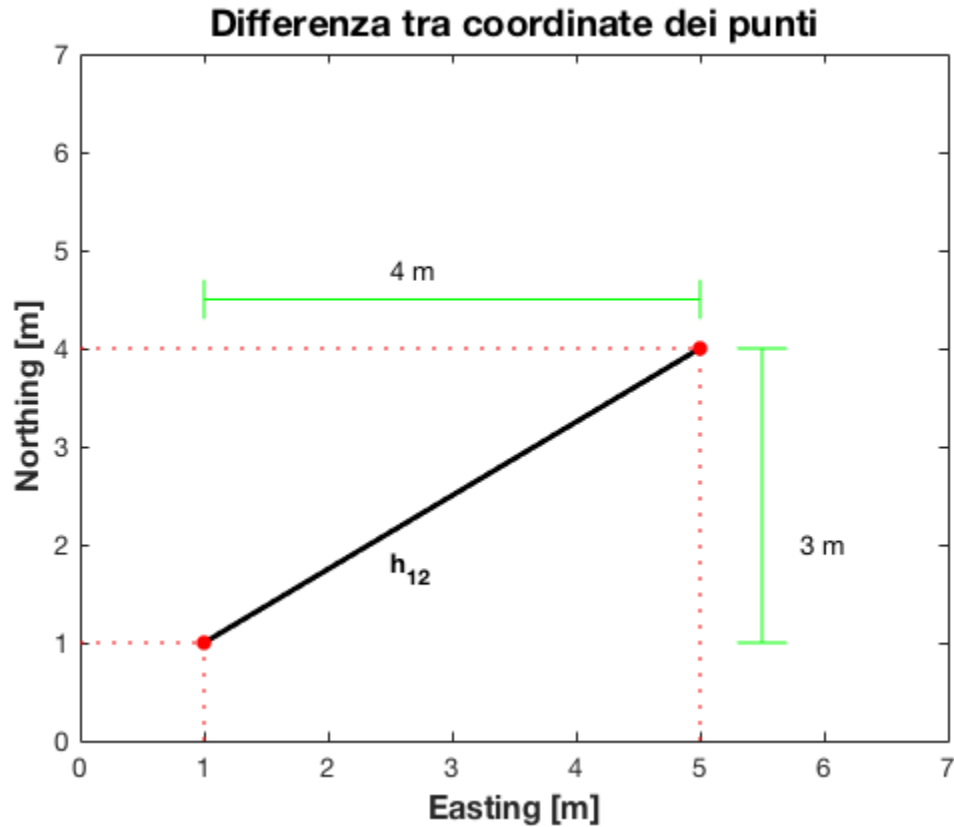
In [47]: hold on

```

plot(x,y,'-k','LineWidth',2)
scatter(x,y,255,'.','r'),axis([0,7,0,7]),xlabel('Easting [m]'),ylabel('Northing [m]')
plot([x(1),x(1)],[0,y(1)],'r:'),plot([x(2),x(2)],[0,y(2)],'r:'),plot([0,x(1)],[y(1),y(1)],'r:'),
plot([x(1),x(2)],[y(2)+0.5,y(2)+0.5],'g'),plot([1,1],[y(2)+0.5-0.2,y(2)+0.5+0.2],'g'),
plot([x(2)+0.5,x(2)+0.5],[y(1),y(2)],'g'),plot([x(2)+0.5-0.2,x(2)+0.5+0.2],[y(1),y(1)],'g'),
hold off
text(x(2)/2,y(2)+0.8,[num2str(x(2)-x(1)),' m'])
text(x(2)+0.8,y(2)/2,[num2str(y(2)-y(1)),' m'])
text(x(2)/2,y(2)/2,['h_{12}'],'HorizontalAlignment','left','VerticalAlignment','top','FontWeight','b','FontSize',12)
xlabel('Easting [m]','FontWeight','b','FontSize',12)
ylabel('Northing [m]','FontWeight','b','FontSize',12)

```

```
title('Differenza tra coordinate dei punti','FontWeight','b','FontSize',14)
box on
```



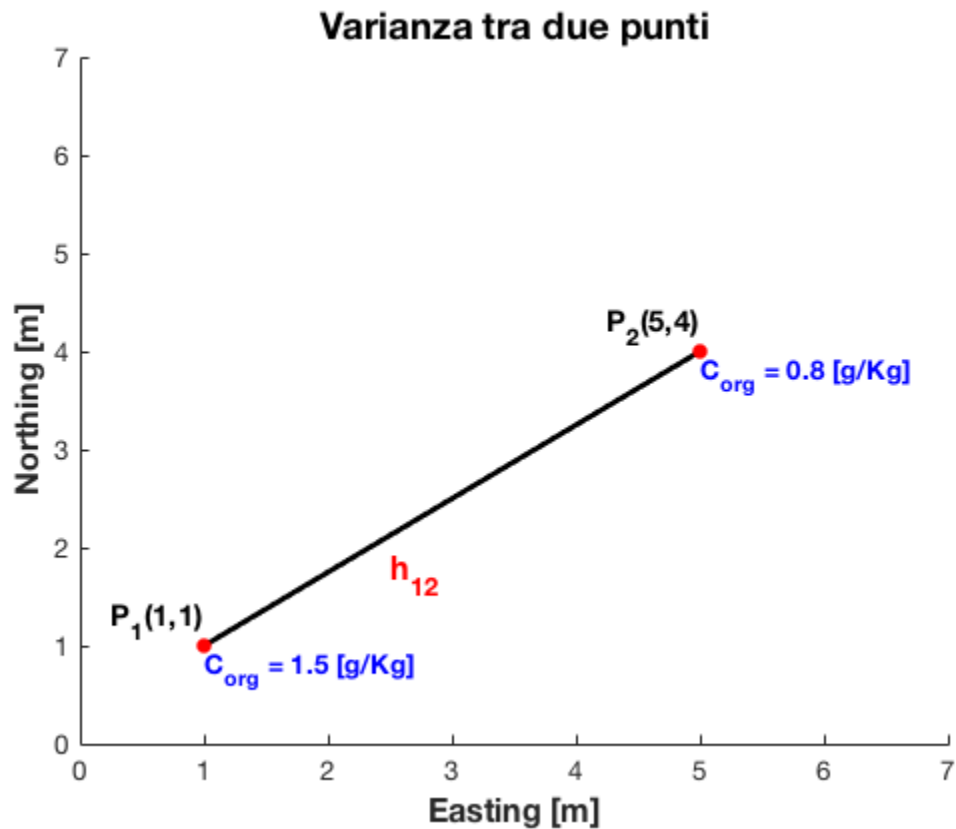
**Applichiamo il teorema di Pitagora:**  $h_{12} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} = \sqrt{4^2 + 3^2} = \sqrt{16 + 9} = \sqrt{25} = 5 \text{ [m]}$

## 2. Varianza: Distanza in C.O. [g/Kg] misurato nei due punti **P<sub>1</sub>** e **P<sub>2</sub>**

```
In [34]: hold on
plot(x,y,'-k','LineWidth',2)
scatter(x,y,255,'.','r'),axis([0,7,0,7]),xlabel('Easting [m]'),ylabel('Northing [m]')
hold off
text(x(2)/2,y(2)/2,['\color{red}h_{12}'],'HorizontalAlignment','left',...
     'VerticalAlignment','top','FontWeight','b','FontSize',12)
text(x(1),y(1),P1,'HorizontalAlignment','right','VerticalAlignment','bottom','FontWeight','b','FontSize',12)
text(x(2),y(2),P2,'HorizontalAlignment','right','VerticalAlignment','bottom','FontWeight','b','FontSize',12)
text(x(1),y(1),['C_{org} = ',num2str(z1),' [g/Kg]'],'color','blue', ...
     'HorizontalAlignment','left','VerticalAlignment','top','FontWeight','b','FontSize',12)
text(x(2),y(2),['C_{org} = ',num2str(z2),' [g/Kg]'],'color','blue', ...
     'HorizontalAlignment','left','VerticalAlignment','top','FontWeight','b','FontSize',12)
xlabel('Easting [m]','FontWeight','b','FontSize',12)
```



```
ylabel('Northing [m]', 'FontWeight', 'b', 'FontSize', 12)
title('Varianza tra due punti', 'FontWeight', 'b', 'FontSize', 14)
```



$$CO_{P_1} = 1.5 \text{ [g/Kg]} \quad \rightarrow \quad z(P_1) = z(x_i) \quad CO_{P_2} = 0.8 \text{ [g/Kg]} \quad \rightarrow \quad z(P_2) = z(x_i + h)$$

Utilizziamo  $z$  per indicare il contenuto in C.O. degli orizzonti  $A_p$ , ma può essere utilizzata per indicare una qualsiasi proprietà del suolo (e non).

$$\gamma(h_{12}) = (z(x_1 + h_{12}) - z(x_1))^2 = (0.8 - 1.5)^2 = -0.7^2 = 0.49$$

```
In [35]: gamma = (z2 - z1)^2;
          fprintf('gamma = %.2f', gamma)
```

```
gamma = 0.49
```

#### 0.1.4 Estendiamo il calcolo a più punti contigui, fissando $P_1$

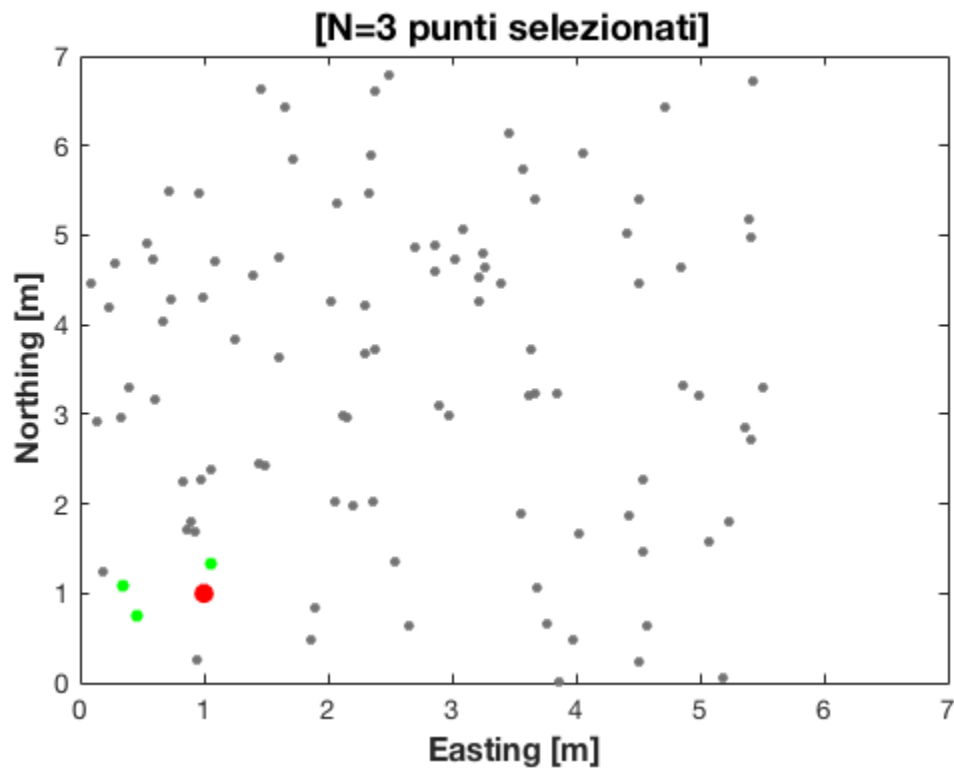
```
In [36]: N = 3;
          D = sqrt((X - x(1)).^2 + (Y - y(1)).^2);
          [~, iD] = sort(D);
          x1_n = X(iD(1:N));
          y1_n = Y(iD(1:N));
          d1_n = round(D(iD(1:N)), 3);
          z1_n = [z1-0.1:-0.1:z1-0.1*N]'; %contenuto in C.O. [g/Kg] nei punti contigui
```

```

hold on
scatter(x(1),y(1),450,'.','r')
scatter(X,Y,16,[.45 .45 .45],'filled')
scatter(x1_n,y1_n,25,'green','filled')
hold off
axis([0,7,0,7])
box on
xlabel('Easting [m]','FontWeight','b','FontSize',12)
ylabel('Northing [m]','FontWeight','b','FontSize',12)
title({'Fissiamo il punto P_1 e calcoliamo distanza & varianza'; ...
      ['[N=',num2str(N),' punti selezionati]']},'FontWeight','b','FontSize',14)

```

## Fissiamo il punto $P_1$ e calcoliamo distanza & varianza



### Zoom in $P_1$

```

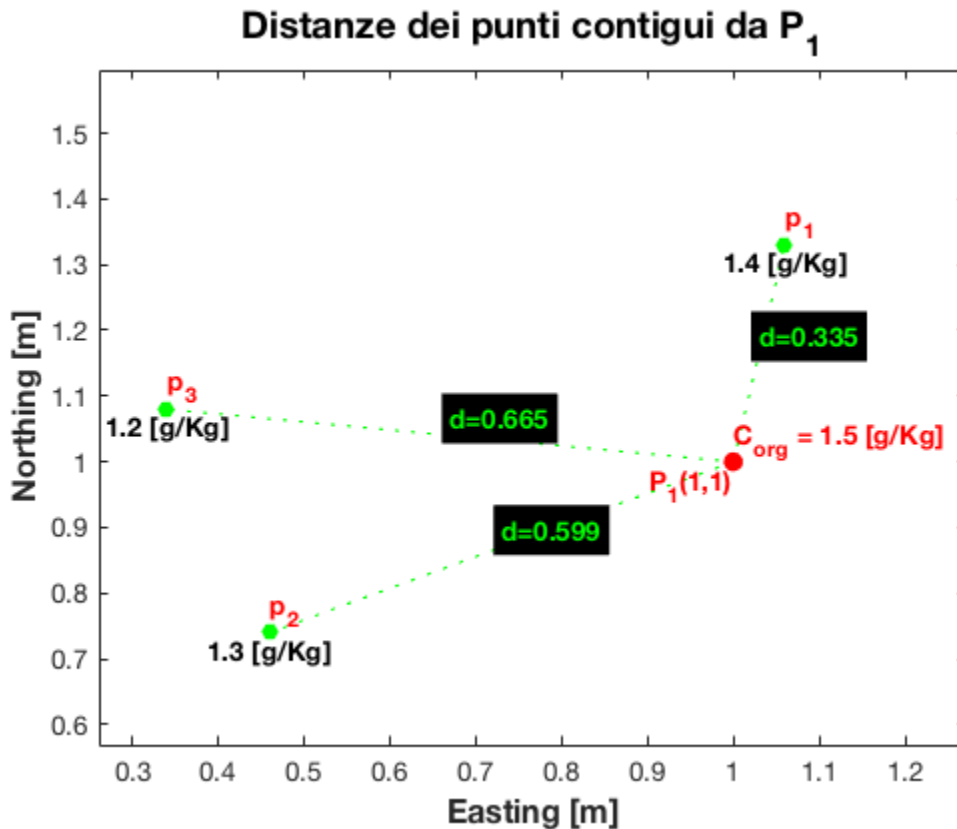
In [60]: figure
hold on
% plot of N neighbour locations to P1:
scatter(x1_n,y1_n,41,'green','filled'),axis([0,7,0,7]),xlabel('Easting [m]'),ylabel('Northing [m]')
for ii = 1:N
    %pi = ['P_',num2str(ii),'(',num2str(x1_n(ii))',' ',num2str(y1_n(ii)),')'];
    pi = ['p_',num2str(ii)];
end

```

```

plot([x(1),x1_n(ii)],[y(1),y1_n(ii)],'g:')
text(abs(x(1)+x1_n(ii))/2,abs(y(1)+y1_n(ii))/2,['d=',num2str(d1_n(ii))],...
     'color','g','background','black','VerticalAlignment','bottom','FontWeight','b',
text(x1_n(ii),y1_n(ii),[num2str(z1_n(ii)),' [g/Kg]'], ...
     'HorizontalAlignment','center','VerticalAlignment','top',...
     'FontWeight','b','FontSize',10)
text(x1_n(ii),y1_n(ii),pi,'HorizontalAlignment','left',...
     'VerticalAlignment','bottom','FontWeight','b','color','red')
end
% plot P1:
scatter(x(1),y(1),450,'.','r'),axis([0,7,0,7]),xlabel('Easting [m]'),ylabel('Northing [m]')
text(x(1),y(1),['C_{org} = ',num2str(z1),' [g/Kg]'],'HorizontalAlignment','left',...
     'VerticalAlignment','bottom','FontWeight','b','color','red')
text(x(1),y(1),P1,'HorizontalAlignment','right','VerticalAlignment','top','FontWeight','b',...
     'color','red')
hold off
axis([min(x1_n)/1.3,max(x1_n)*1.2,min(y1_n)/1.3,max(y1_n)*1.2])
box on
xlabel('Easting [m]','FontWeight','b','FontSize',12)
ylabel('Northing [m]','FontWeight','b','FontSize',12)
title('Distanze dei punti contigui da P_1','FontWeight','b','FontSize',14)

```

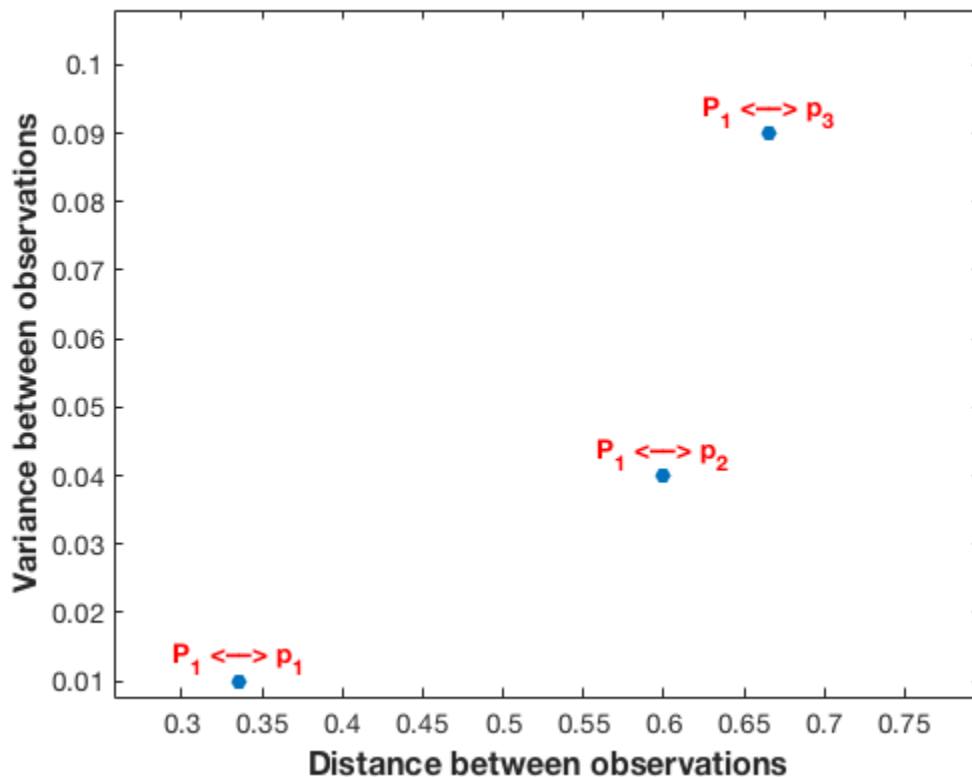


Calcoliamo le distanze geospaziali vs varianze per le N coppie di punti rappresentate nel grafico sopra:  $h_{1i} = \sqrt{(x_i - x_1)^2 + (y_i - y_1)^2}$   $\gamma(h_{1i}) = (z(x_1 + h_{1i}) - z(x_1))^2$

```
In [38]: gamma = (z1_n - z1).^2;
fprintf('%8s, %8s\n', 'Distanza', 'Varianza')
for ii=1:numel(gamma) fprintf('%8.3f, %8.3f\n', d1_n(ii), gamma(ii)); end
```

```
Distanza, Varianza
0.335,    0.010
0.599,    0.040
0.665,    0.090
```

```
In [64]: scatter(d1_n, gamma, 'filled')
xlabel('Distance between observations', 'FontWeight', 'b', 'FontSize', 12)
ylabel('Variance between observations', 'FontWeight', 'b', 'FontSize', 12)
%axis([0, 1, 0, 0.125])
box on
for ii = 1:N
text(d1_n(ii), gamma(ii), ['P_1 <=> p_', num2str(ii)], 'HorizontalAlignment', 'center', ...
    'VerticalAlignment', 'bottom', 'FontWeight', 'b', 'color', 'red')
end
axis([min(d1_n)/1.3, max(d1_n)*1.2, min(gamma)/1.3, max(gamma)*1.2])
```



### 0.1.5 Estendiamo il calcolo a tutti i punti di campionamento

(cioè collegando ciascun punto con tutti gli altri...)

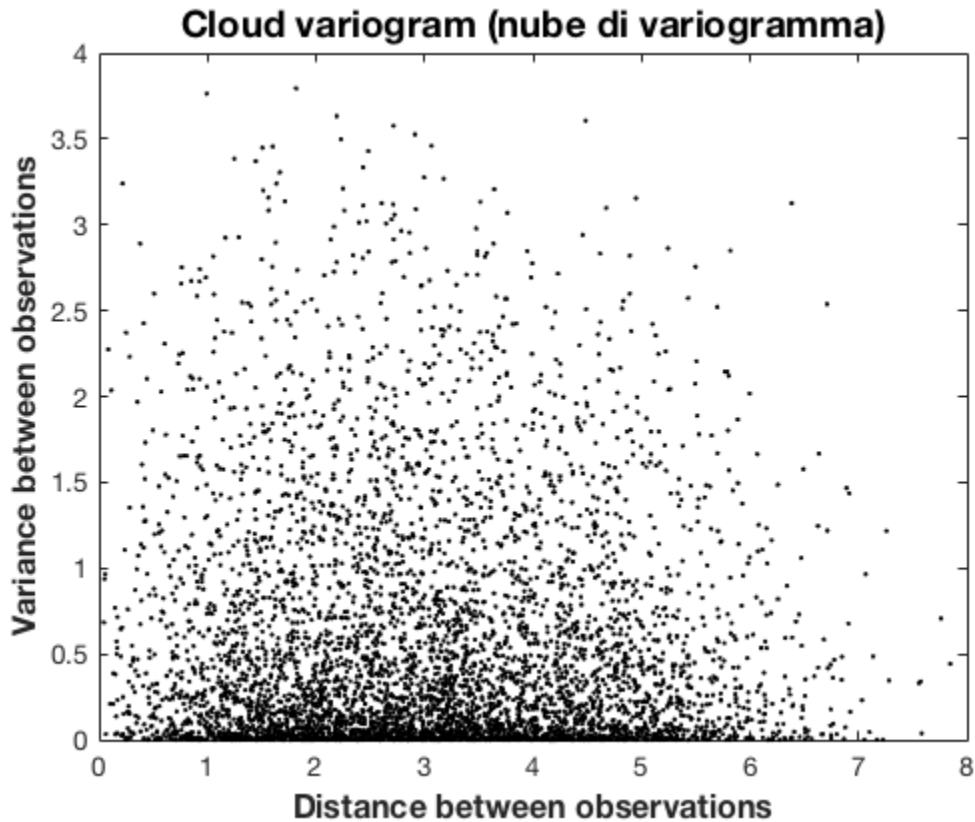
```
In [40]: % DISTANCES
[X1,X2] = meshgrid(X);
[Y1,Y2] = meshgrid(Y);
D = sqrt((X1 - X2).^2 + (Y1 - Y2).^2);

% SEMIVARIANCE
[Z1,Z2] = meshgrid(Z);
Gamma = (Z1 - Z2).^2;

% LOWER-LEFT TRIANGLE
indx = 1 : length(Z);
[C,R] = meshgrid(indx);
I = R > C;
% we only take I, because values are duplicated over the two triangles:
r=23;c=2;
fprintf( 'D(%d,%d) = %7.4f\nD(%d,%d) = %7.4f\n', r,c,D(r,c), c,r,D(c,r) )

D(23,2) = 0.8202
D(2,23) = 0.8202

In [41]: plot(D(I),Gamma(I),'Marker','.','LineStyle','none',...
    'MarkerFaceColor',[.6 .6 .6],'MarkerEdgeColor','k','MarkerSize',1)
xlabel('Distance between observations','FontWeight','b','FontSize',12)
ylabel('Variance between observations','FontWeight','b','FontSize',12)
title('Cloud variogram (nube di variogramma)','FontWeight','b','FontSize',14)
```



### 0.1.6 Costruzione del V A R I O G R A M M A sperimentale

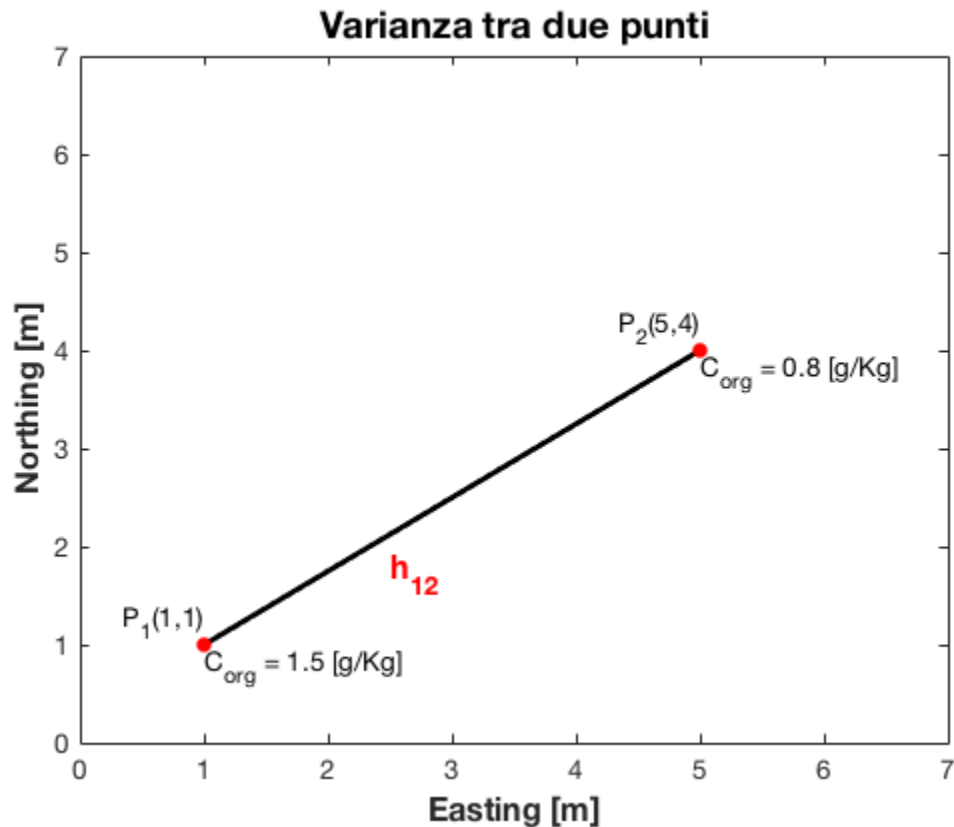
$$\gamma(h) = \sum_{i=1}^{n(h)} \frac{(z(x_i + h) - z(x_i))^2}{n(h)}$$

Per costruire il variogramma sperimentale calcoliamo due parametri di base seguendo regole empiriche (rule of thumb).

```
In [66]: hold on
         plot(x,y,'-k','LineWidth',2)
         scatter(x,y,255,'.','r'),axis([0,7,0,7])
         hold off
         xlabel('Easting [m'],'FontWeight','b','FontSize',12)
         ylabel('Northing [m'],'FontWeight','b','FontSize',12)
         title('Varianza tra due punti','FontWeight','b','FontSize',14)

         text(x(2)/2,y(2)/2,['\color{red}h_{12}'],'HorizontalAlignment','left',...
              'VerticalAlignment','top','FontWeight','b','FontSize',12)
         text(x(1),y(1),P1,'HorizontalAlignment','right','VerticalAlignment','bottom')
         text(x(2),y(2),P2,'HorizontalAlignment','right','VerticalAlignment','bottom')
         text(x(1),y(1),['C_{org} = ',num2str(z1),' [g/Kg]'],'HorizontalAlignment','left','Verti
```

```
text(x(2),y(2),['C_{org} = ',num2str(z2),' [g/Kg]'],'HorizontalAlignment','left','VerticalAlignment','top');
box on
```



massima distanza ( $h_{max}$ ) tra i 2 punti della coppia separata da  $h \rightarrow \max(\frac{D}{2})$   
dimensione del lag (\*DEFINIRE IL LAG!)  $\rightarrow \text{mean}(\min(D))$

```
In [43]: hmax = max(D(:))/2; % rule of thumb #1
D2 = D.*(diag(X*NaN)+1);
lag = mean(min(D2)); % rule of thumb #2
max_lags = floor(hmax/lag);

LAGS = ceil(D/lag);
for i = 1 : max_lags
    SEL = (LAGS == i);
    PN(i) = sum(sum(SEL == 1))/2; % numero di coppie di punti in quel LAG
    DE(i) = mean(mean(D(SEL))); % distanza media a quel LAG
    GE(i) = mean(mean(Gamma(SEL))); % varianza media a quel LAG
end
```

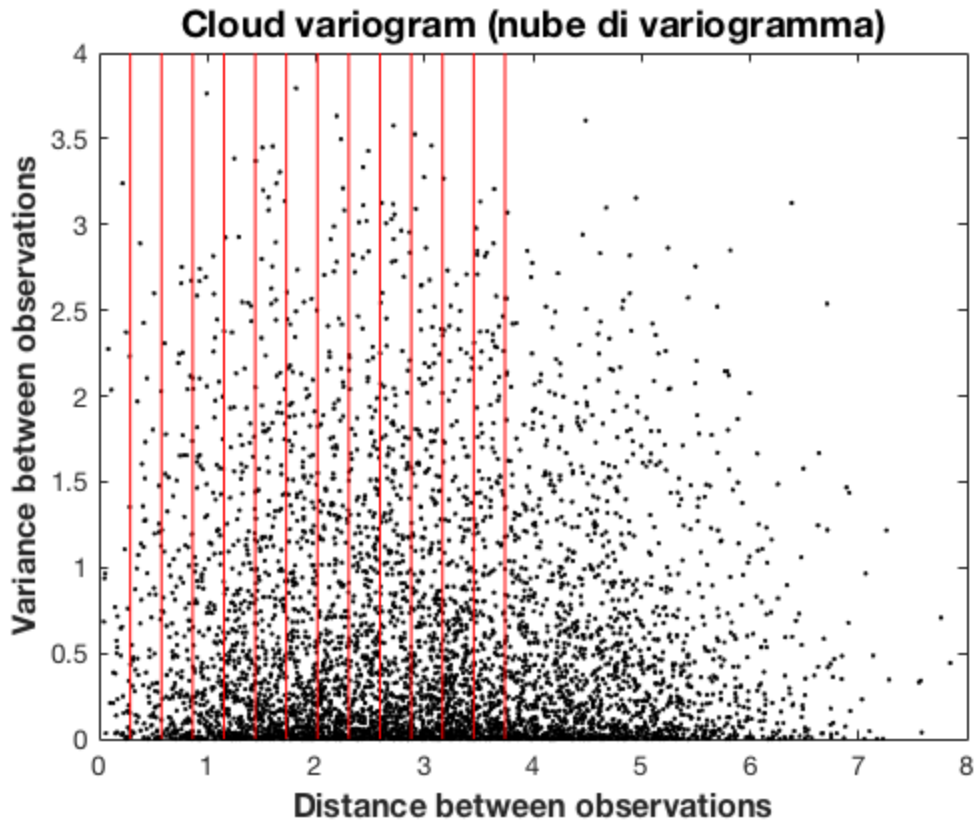
Vediamo graficamente la nube di variogramma sperimentale con la separazione in lag:

```
In [44]: hold on
plot(D(I),Gamma(I),'Marker','.', 'LineStyle','none',...
```

```

    'MarkerFaceColor',[.6 .6 .6],'MarkerEdgeColor','k','MarkerSize',1)
xlabel('Distance between observations','FontWeight','b','FontSize',12)
ylabel('Variance between observations','FontWeight','b','FontSize',12)
title('Cloud variogram (nube di variogramma)','FontWeight','b','FontSize',14)
for ii = 1:max_lags
    clag = lag*ii;
    plot([clag,clag],[0,4],'-r')% max(Gamma(:))
end
hold off, box on

```



Vediamo graficamente il variogramma sperimentale:

```

In [45]: plot(DE,GE,'Marker','o','LineStyle','none',...
    'MarkerFaceColor','green','MarkerEdgeColor','r','MarkerSize',8)
var_z = var(Z);
b = [0 max(DE)];
c = [var_z var_z];
hold on
%plot(b,c,'-k')
yl = 1.1 * max(GE);
ylim([0 yl])

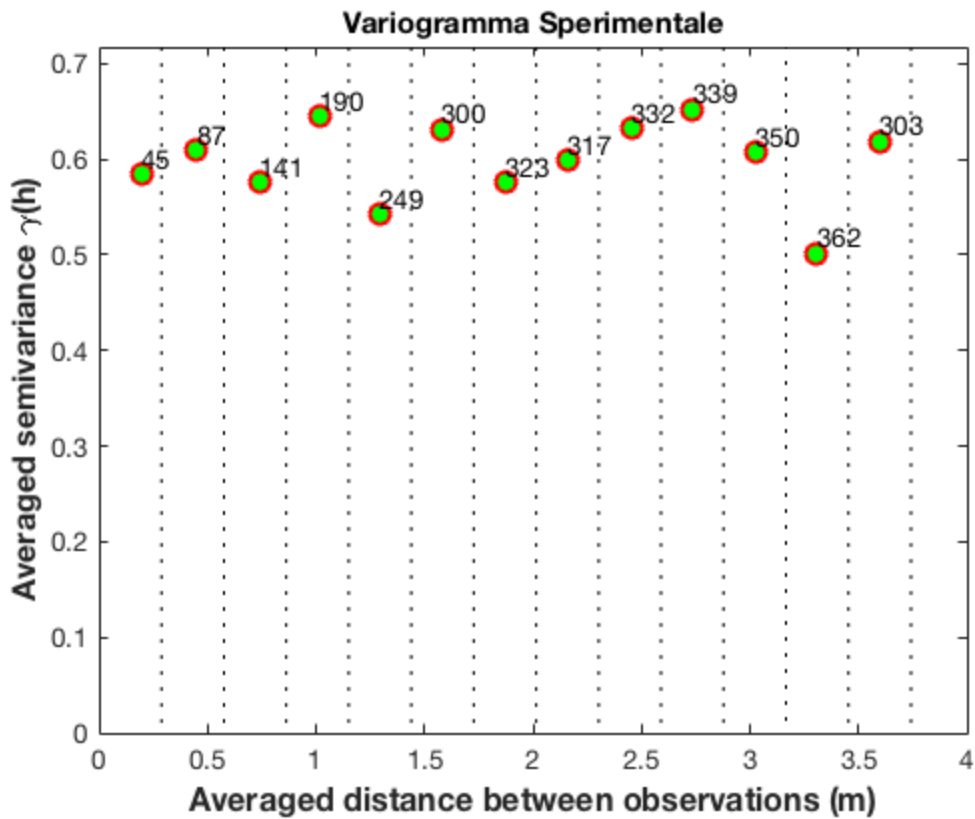
```



```

xlabel('Averaged distance between observations (m)','FontWeight','b','FontSize',12)
ylabel('Averaged semivariance \gamma(h)','FontWeight','b','FontSize',12)
for ii = 1:max_lags
    clag = lag*ii;
    plot([clag,clag],[0,max(Gamma(:))],':k')
end
hold off
for ii=1:numel(PN)
    text(DE(ii),GE(ii),num2str(PN(ii)),'Verticalalignment','bottom','HorizontalAlignment','right')
end
title('Variogramma Sperimentale')

```



Arriviamo alla costruzione della formula del variogramma sperimentale: (devo dividere in blocchi e spiegarli uno per uno)  $\gamma(h) = \sum_{i=1}^{n(h)} \frac{(z(x_i + h) - z(x_i))^2}{n(h)}$

```

In [46]: fprintf( '%8s, %8s, %8s, %8s\n', 'LAG', 'DIST', 'GAMMA', 'Ncoppie' )
for ii=1:max_lags
    fprintf( '%8.2f, %8.2f, %8.2f, %8d\n', lag*ii, DE(ii), GE(ii), PN(ii) )
end

```

LAG,	DIST,	GAMMA,	Ncoppie
0.29,	0.20,	0.58,	45

0.58,	0.45,	0.61,	87
0.86,	0.74,	0.58,	141
1.15,	1.02,	0.64,	190
1.44,	1.29,	0.54,	249
1.73,	1.58,	0.63,	300
2.01,	1.87,	0.58,	323
2.30,	2.16,	0.60,	317
2.59,	2.45,	0.63,	332
2.88,	2.73,	0.65,	339
3.17,	3.03,	0.61,	350
3.45,	3.31,	0.50,	362
3.74,	3.60,	0.62,	303

## 0.2 Facciamo alcuni esempi

Illustrare la procedura in Excel, e consegnare agli studenti (assegnare esercizio a casa utilizzando i dati estratti dal DEM per la costruzione del variogramma)

Entriamo in R ed installiamo il pacchetto gstat Entriamo in GIS (SAGA, QGIS, ...), importiamo il DEM ed astraiano i punti di quota in corrispondenza dei punti di campionamento pedologico. Importare in R i dati di quota estratti, costruire un variogramma sperimentale, eseguire il fitting mediante un variogramma teorico autorizzato, utilizzare diversi modelli di variogramma per il kriging, confrontare le mappe (diversi modelli di variogramma, mappa di kriging vs DEM originale).