

ΜΟΝΟΠΑΡΑΜΕΤΡΙΚΗ ΚΑΙ  
ΠΟΛΥΠΑΡΑΜΕΤΡΙΚΗ  
ΣΤΑΤΙΣΤΙΚΗ ΑΝΑΛΥΣΗ  
ΔΕΔΟΜΕΝΩΝ



# ΣΤΑΤΙΣΤΙΚΗ ΕΠΕΞΕΡΓΑΣΙΑ ΓΙΑ ΤΟΝ ΕΝΤΟΠΙΣΜΟ ΓΕΩΧΗΜΙΚΗΣ ΑΝΩΜΑΛΙΑΣ

- Στατιστική ανάλυση του γεωχημικού δείγματος → μας δίνει πληροφορίες για τον γεωχημικό πληθυσμό που μελετάμε.
- Ποσοτικοποίηση Αβεβαιότητας των μετρήσεων
- Πειραματικά δεδομένα → οργανωμένα με τρόπο που επιτρέπει την εύκολη εισαγωγή και διαχείρισή τους στον υπολογιστή.
- Στατιστική εκτίμηση με υπολογισμούς παραμέτρων και εξέταση υποθέσεων.

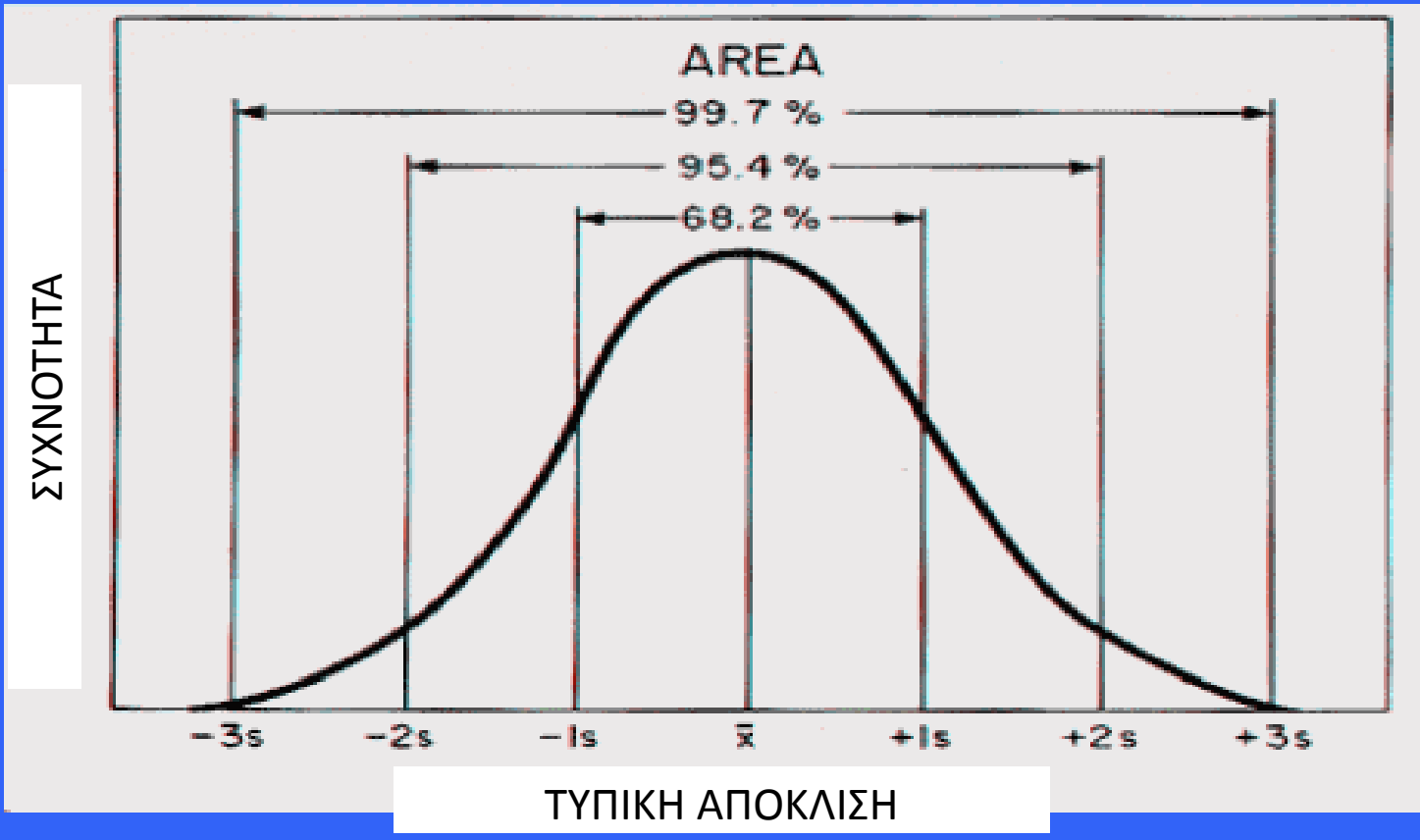
# ΒΗΜΑΤΑ ΣΤΑΤΙΣΤΙΚΗΣ ΑΝΑΛΥΣΗΣ

1. **Μονοπαραγοντική** (univariate) ανάλυση → Απεικόνιση και μελέτη της στατιστικής κατανομής των δεδομένων ανά παράμετρο ενδιαφέροντος
2. Υπολογισμός περιγραφικών στατιστικών παραμέτρων.
3. Αναγνώριση απόμακρων τιμών.
4. Προσδιορισμός τιμών «γεωχημικού πλαισίου».
5. Αναγνώριση «γεωχημικής ανωμαλίας»
6. **Ανάλυση με 2 μεταβλητές** (bivariate stats) → σχέση μεταξύ μεταβλητών  
Συντελεστής συσχέτισης, μέθοδος ελαχίστων τετραγώνων
7. **Πολυμεταβλητή ανάλυση** (multivariate stats) → ομαδοποίηση δεδομένων  
(Factor analysis, Principal Components Analysis)
8. **Γεωστατιστική ανάλυση** → χωρική διάσταση (semivariograms and kriging)

# ΚΑΜΠΥΛΗ ΚΑΝΟΝΙΚΗΣ ΚΑΤΑΝΟΜΗΣ

4

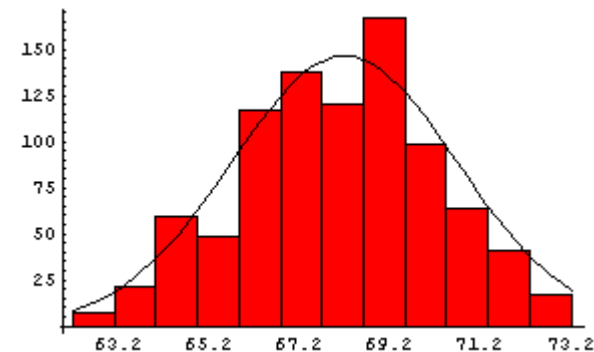
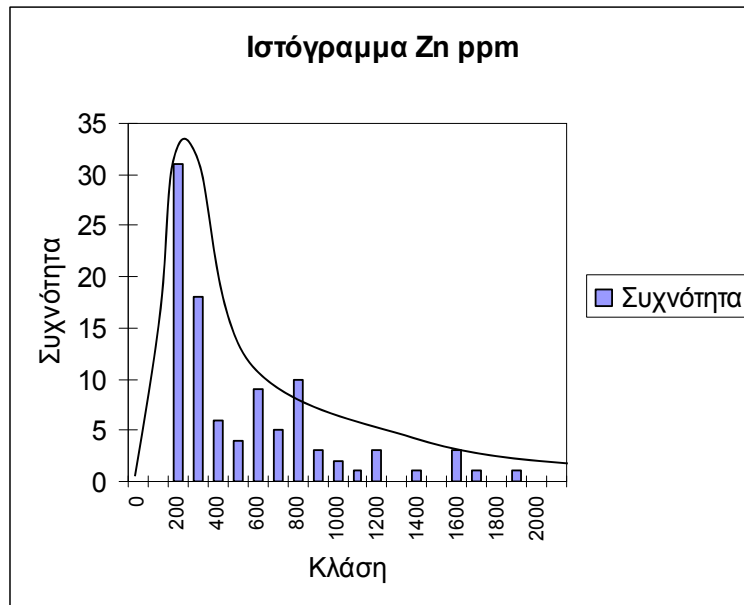
## Gaussian Curve



# ΣΤΑΤΙΣΤΙΚΗ ΚΑΤΑΝΟΜΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ

5

- Έλεγχος με γραφικές μεθόδους (π.χ. ιστόγραμμα) ή με εφαρμογή στατιστικών δοκιμών.
- Ο υπολογισμός πολλών στατιστικών παραμέτρων προϋποθέτει κανονική κατανομή δεδομένων → **πιθανή αναγκαιότητα η μετατροπή** (π.χ. με λογαρίθμιση) **για κανονικοποίηση**.



# ΠΕΡΙΓΡΑΦΙΚΑ ΣΤΑΤΙΣΤΙΚΑ – (Α) ΕΓΓΥΤΗΤΑ ΣΤΗΝ ΚΕΝΤΡΙΚΗ ΤΙΜΗ

6

ο **αριθμητικός μέσος** ή μέση τιμή (arithmetic mean or average)  $n$  μετρήσεων:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

**διάμεσος** (median), δηλαδή η τιμή για την οποία το ήμισυ των δειγμάτων στην κατανομή έχει τιμές μικρότερες και το ήμισυ μεγαλύτερες. Η διάμεση τιμή αποτελεί πιο σταθερή παράμετρο από τη μέση τιμή σε περιπτώσεις λίγων μετρήσεων ( $n < 5$ ).

**τιμή μέγιστης συχνότητας** (mode). Τα δεδομένα υποδιαιρούνται σε ειδικά διαστήματα (κλάσεις) και λαμβάνεται το κέντρο της κλάσης με την υψηλότερη συχνότητα. Πολλές γεωχημικές κατανομές έχουν περισσότερες από μία τιμή μέγιστης συχνότητας.

Ο **γεωμετρικός μέσος** (geometric mean). Η μέση τιμή του δεκαδικού λογαρίθμου των δεδομένων.

## ΠΕΡΙΓΡΑΦΙΚΑ ΣΤΑΤΙΣΤΙΚΑ – (B) ΔΙΑΣΠΟΡΑ

7

**Περιοχή ή εύρος τιμών (range).** Η διαφορά μεταξύ της μέγιστης και της ελάχιστης τιμής των δεδομένων.

**Διακύμανση (variance).** Το τετράγωνο της μέσης διαφοράς μεταξύ τιμών δεδομένων και της μέσης τιμής τους:

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

**Τυπική απόκλιση (standard deviation).** Η τετραγωνική ρίζα της διακύμανσης. Είναι η συνηθέστερη παράμετρος έκφρασης της διασποράς και σε αντιδιαστολή με τη διακύμανση εκφράζεται με τις μονάδες μέτρησης της μέσης τιμής.

**Σχετική τυπική απόκλιση (coefficient of variation).** Ο λόγος:  
τυπική απόκλιση / μέση τιμή

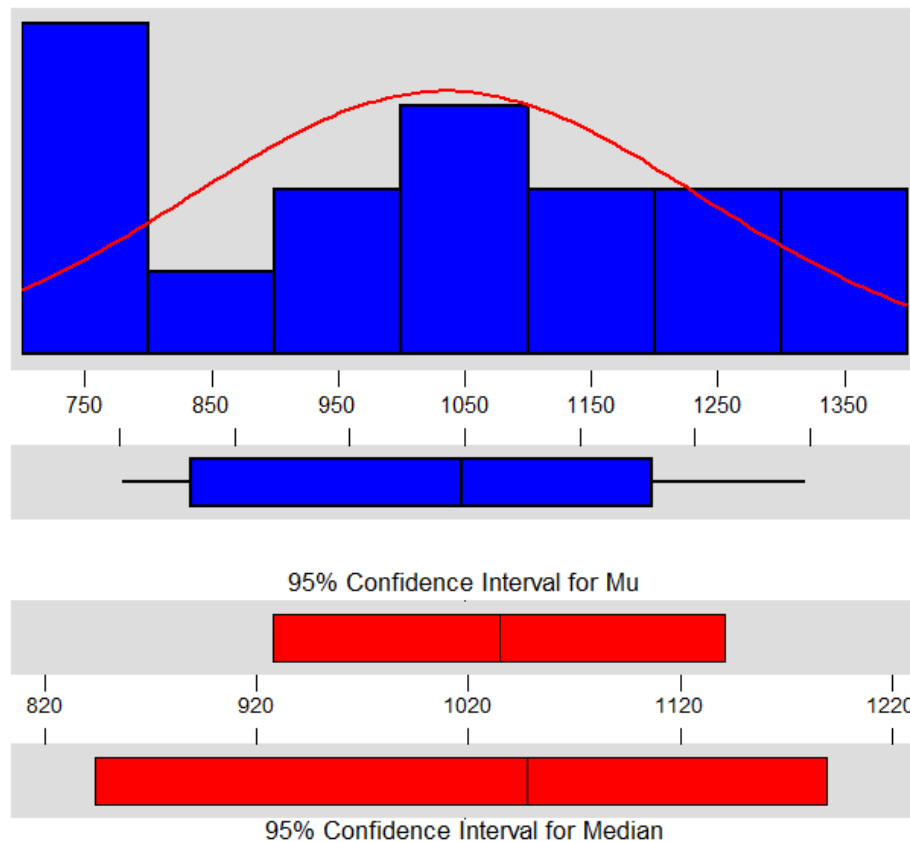
# Μονοπαραγοντική ανάλυση- Κυτιογράφημα απολήξεων

data values or <i>estimated values</i>	boxplot features	mapping symbols
<b>minimum</b> →	* — outliers	
<i>lower outer fence</i> →	⋯ — outliers	○
<i>lower inner fence</i> →	⋯ — outliers	
<b>lower whisker</b> →	—	○
<b>lower hinge (Q1)</b> →	▭ — IQR #	
<b>median (Q2)</b> →	— — hinge width	•
<b>upper hinge (Q3)</b> →	▭ —	+
<b>upper whisker</b> →	—	
<i>upper inner fence</i> →	⋯ — 'mild' outliers	
<i>upper outer fence</i> →	⋯ — 'far' outliers	+
<b>maximum</b> →	* — 'far' outliers	

(Carranza, 2009)



## Descriptive Statistics



### Variable: Pb soil

#### Anderson-Darling Normality Test

A-Squared: 0.355  
P-Value: 0.415

Mean 1034.55  
StDev 200.25  
Variance 40099.1  
Skewness -7.2E-02  
Kurtosis -1.30536  
N 16

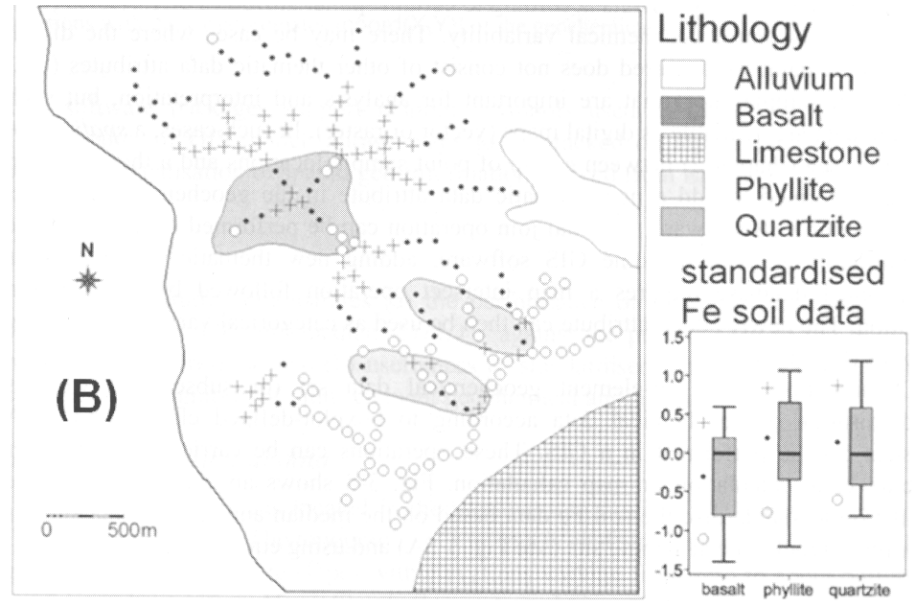
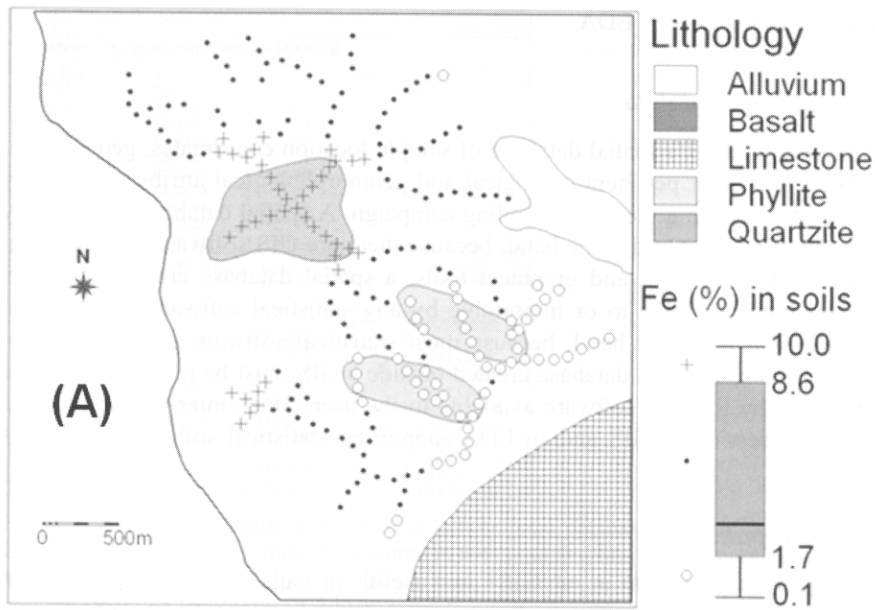
Minimum 752.70  
1st Quartile 812.48  
Median 1047.80  
3rd Quartile 1213.00  
Maximum 1346.20

95% Confidence Interval for Mu  
927.85 1141.25

95% Confidence Interval for Sigma  
147.92 309.92

95% Confidence Interval for Median  
843.39 1189.62

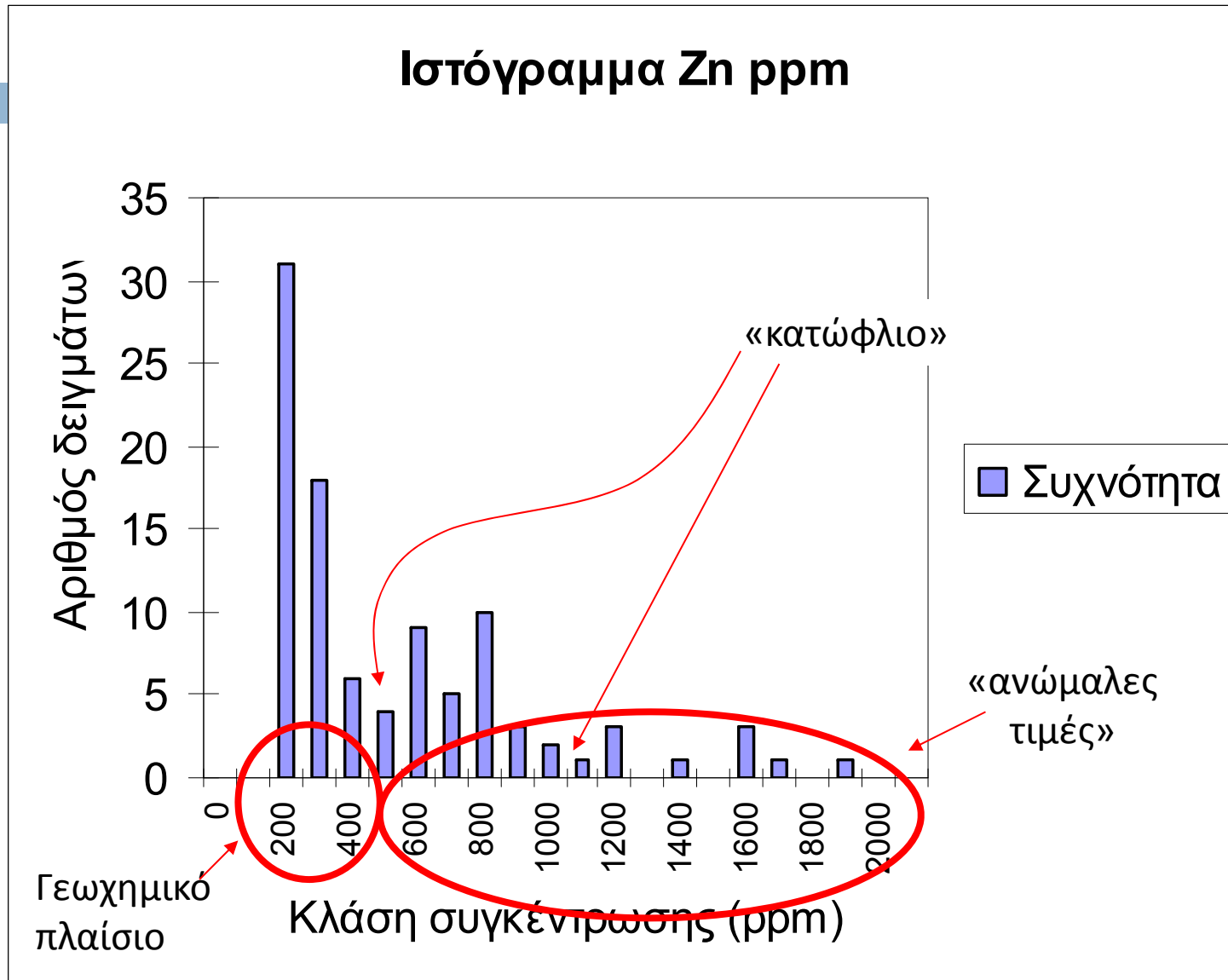
# Χωρική απεικόνιση – ταξινόμηση τιμών με βάση το κυτιογράφημα απολήξεων



(Carranza, 2009)

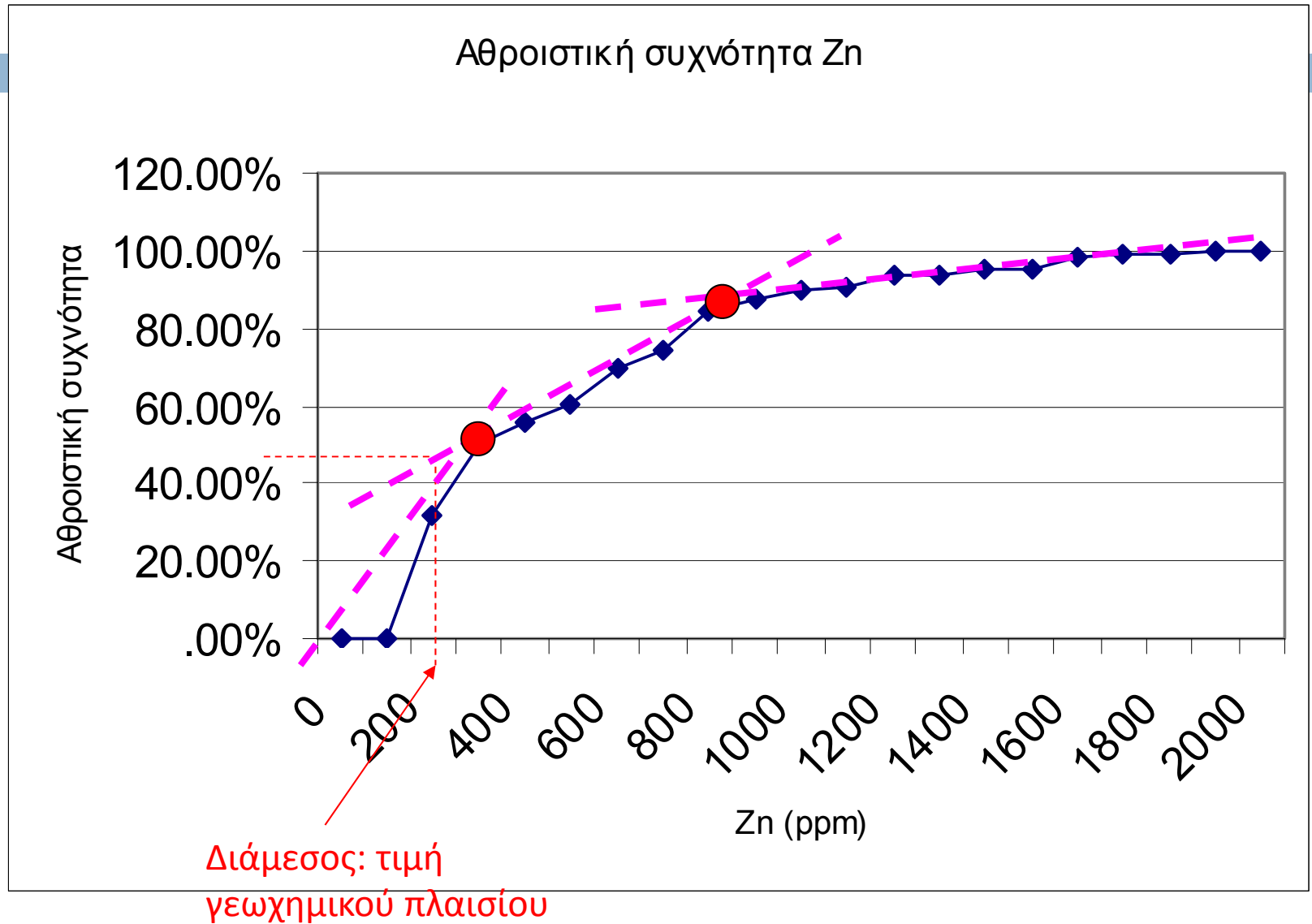
# ΑΠΕΙΚΟΝΙΣΕΙΣ ΔΕΔΟΜΕΝΩΝ ΣΕ ΓΡΑΦΗΜΑΤΑ

11



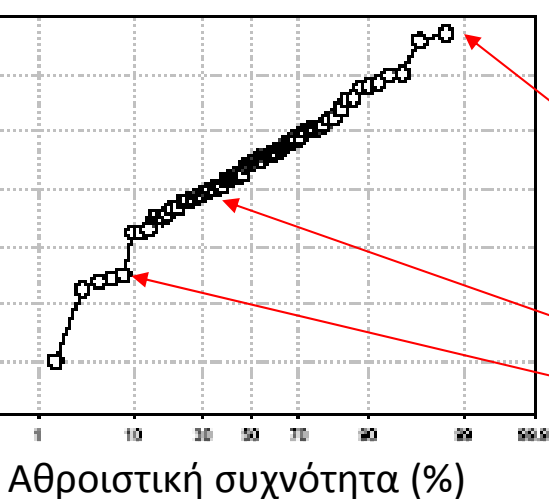
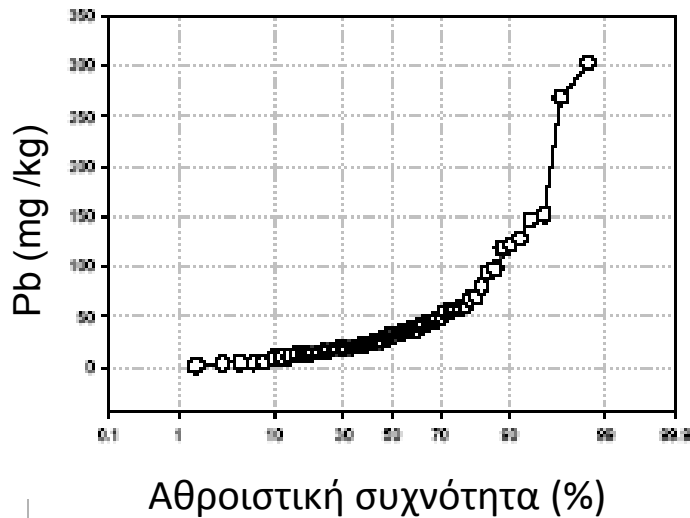
# ΓΡΑΦΗΜΑ ΑΘΡΟΙΣΤΙΚΗΣ ΣΥΧΝΟΤΗΤΑΣ

12



# ΓΡΑΦΗΜΑ ΠΙΘΑΝΟΤΗΤΩΝ

13



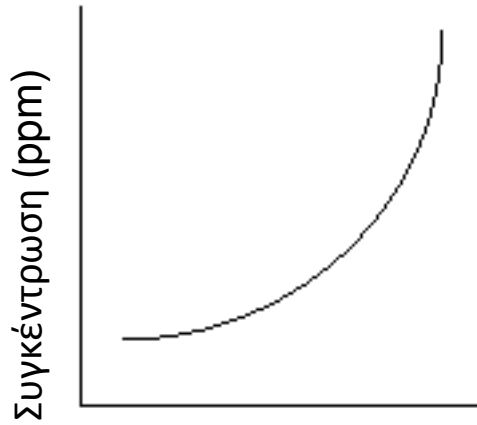
Προβολή αθροιστικής  
συχνότητας σε χαρτί  
αριθμητικών πιθανοτήτων.

Επιτρέπει την αναγνώριση :

1. είδους κατανομής των δεδομένων (π.χ. κανονική, λογαριθμική).
2. Απόμακρων τιμών.
3. Πολλαπλών πληθυσμών στα δεδομένα.

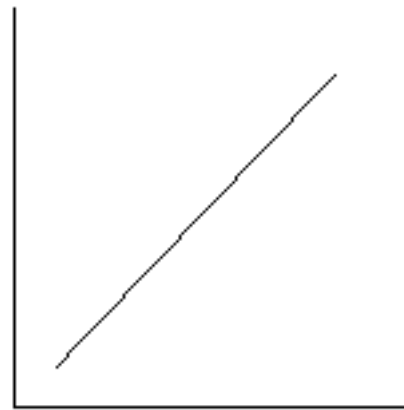
# ΓΡΑΦΗΜΑ ΠΙΘΑΝΟΤΗΤΩΝ – ΕΙΔΟΣ ΚΑΤΑΝΟΜΗΣ

14

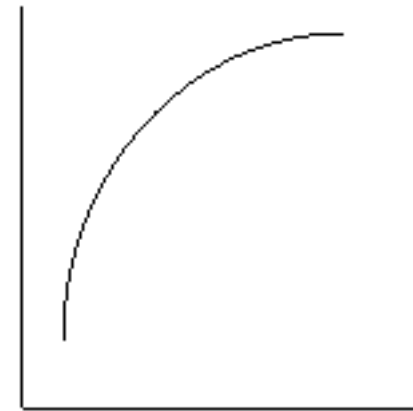


Αθροιστική  
συχνότητα (%)

Λοξή κατανομή με  
«ουρά» προς τα  
δεξιά



Κανονική  
κατανομή

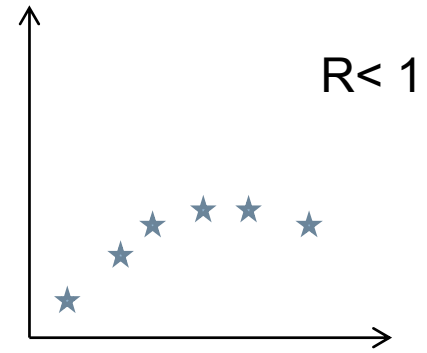
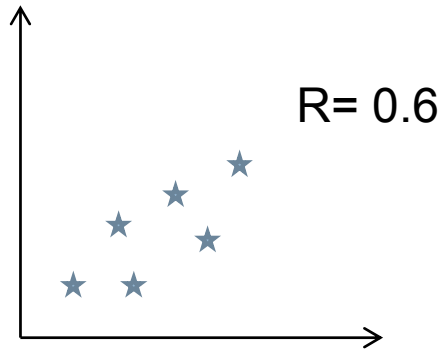
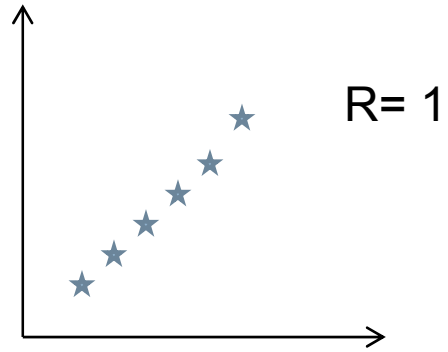
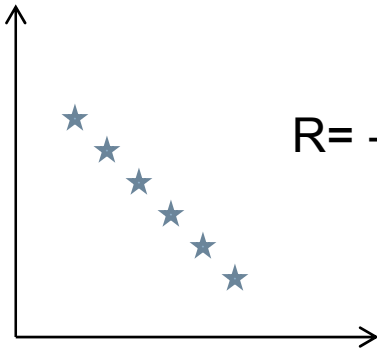


Λοξή κατανομή με  
«ουρά» προς τα  
αριστερά

Συντελεστής συσχέτισης (correlation coefficient)  $r$ .

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}}$$

Ο συντελεστής αυτός παίρνει τιμές από 1 έως -1 με την τιμή 1 να σημαίνει τέλεια θετική συσχέτιση, -1 τέλεια αρνητική συσχέτιση και 0 καμία συσχέτιση. Γενικά τιμές  $r >$  από 0.5 σημαίνουν καλή στατιστική συσχέτιση μεταξύ παραμέτρων.



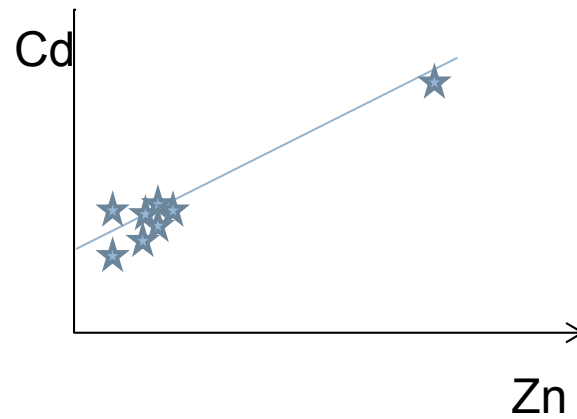
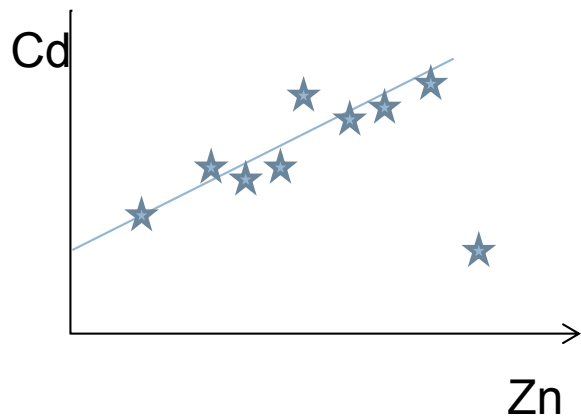
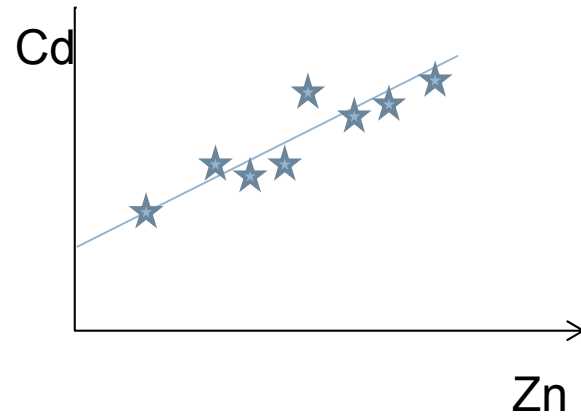
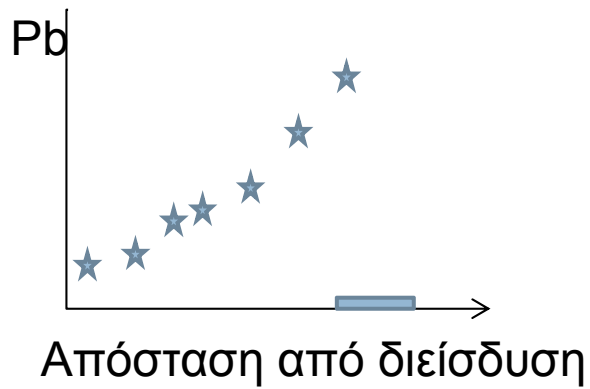


	<i>Pb</i>	<i>Zn</i>	<i>Cu</i>	<i>As</i>	<i>Cd</i>	<i>Mn</i>	<i>Fe</i>	<i>Ni</i>	<i>Co</i>
<i>Zn</i>	0.77 0.000								
<i>Cu</i>	0.33 0.019	0.50 0.000							
<i>As</i>	0.87 0.000	0.77 0.000	0.38 0.000						
<i>Cd</i>	0.66 0.000	0.86 0.000	0.54 0.000	0.71 0.000					
<i>Mn</i>	0.51 0.000	0.79 0.000	0.32 0.023	0.58 0.000	0.75 0.000				
<i>Fe</i>	0.08 0.592	0.23 0.118	0.22 0.135	0.21 0.142	0.29 0.044	0.36 0.010			
<i>Ni</i>	-0.10 0.480	-0.16 0.279	-0.28 0.055	-0.14 0.355	-0.04 0.781	-0.04 0.799	0.22 0.123		
<i>Co</i>	-0.33 0.020	-0.22 0.122	-0.23 0.112	-0.29 0.045	-0.10 0.538	-0.10 0.481	0.46 0.001	0.80 0.000	
<i>Cr</i>	-0.11 0.471	-0.22 0.137	-0.26 0.072	-0.15 0.308	-0.10 0.591	-0.14 0.339	0.24 0.096	0.97 0.000	0.75 0.000

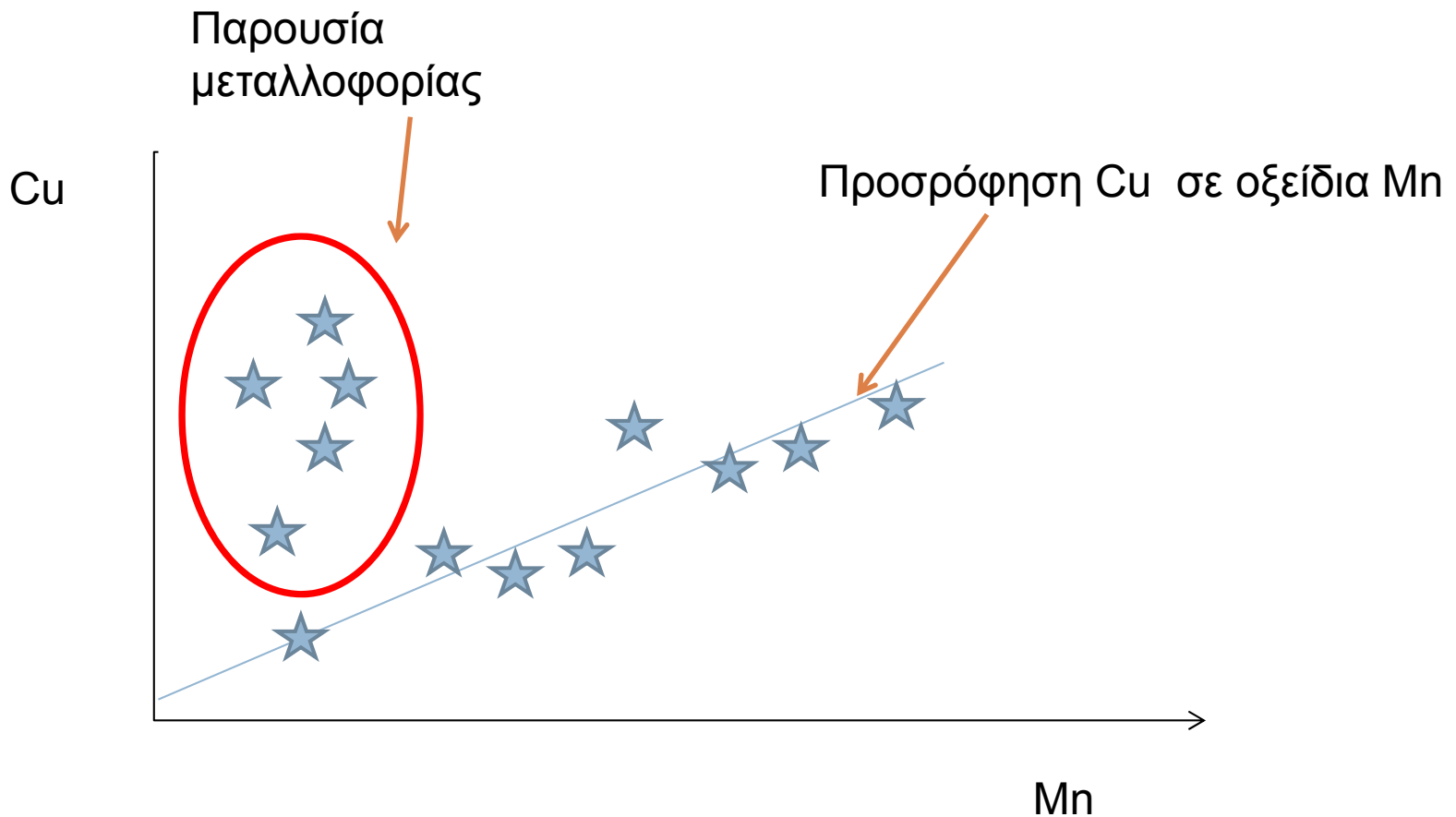
ΠΙΝΑΚΑΣ ΣΥΣΧΕΤΙΣΗΣ

# ΓΡΑΦΗΜΑ ΣΥΣΧΕΤΙΣΗΣ- Διαγράμματα διασποράς

18



# Διάκριση διεργασιών



# Ανάλυση παλινδρόμησης (regression)

- Μη αναστρέψιμη – διάκριση ανεξάρτητης μεταβλητής (predictor) και απόκρισης (response)
- Παλινδρόμηση του  $y$  στο  $x$  ελαχιστοποιεί την ποσότητα  $(y_i - y)^2$
- Υπολογισμός μοντέλου συσχέτισης λόγω γεωχημικών διεργασιών → γραμμική εξίσωση  $y = a x + b$
- Κλίση:

$$a = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

## Regression

The regression equation is  
 $y = 191 + 1.45 x$

Predictor	Coef	StDev	T	P
Constant	191.4	989.4	0.19	0.848
x	1.4451	0.8708	1.66	0.111

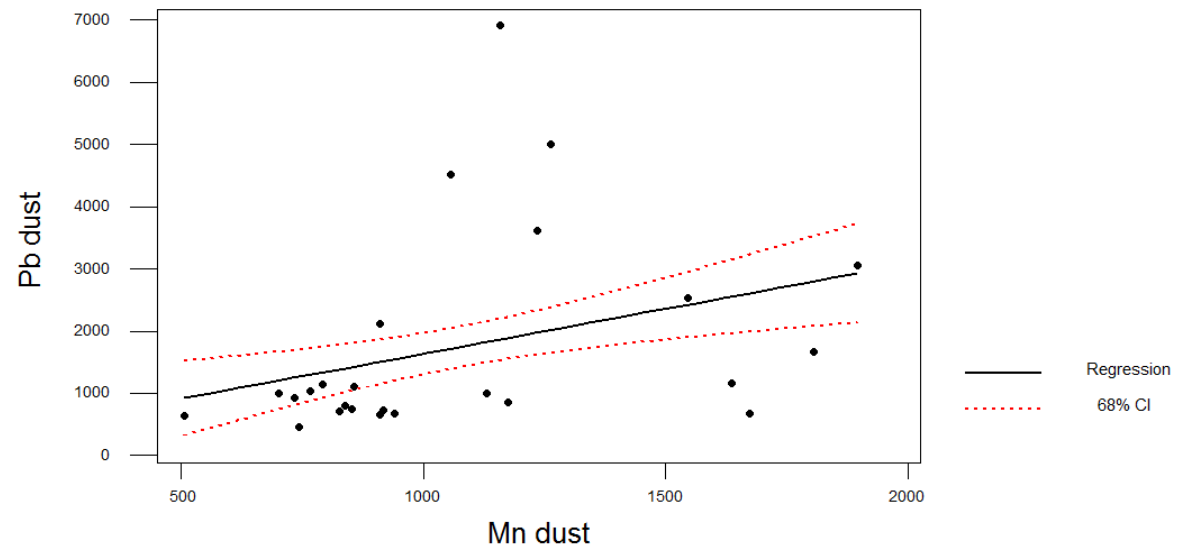
S = 1599      R-Sq = 10.7%      R-Sq(adj) = 6.8%

### Analysis of Variance

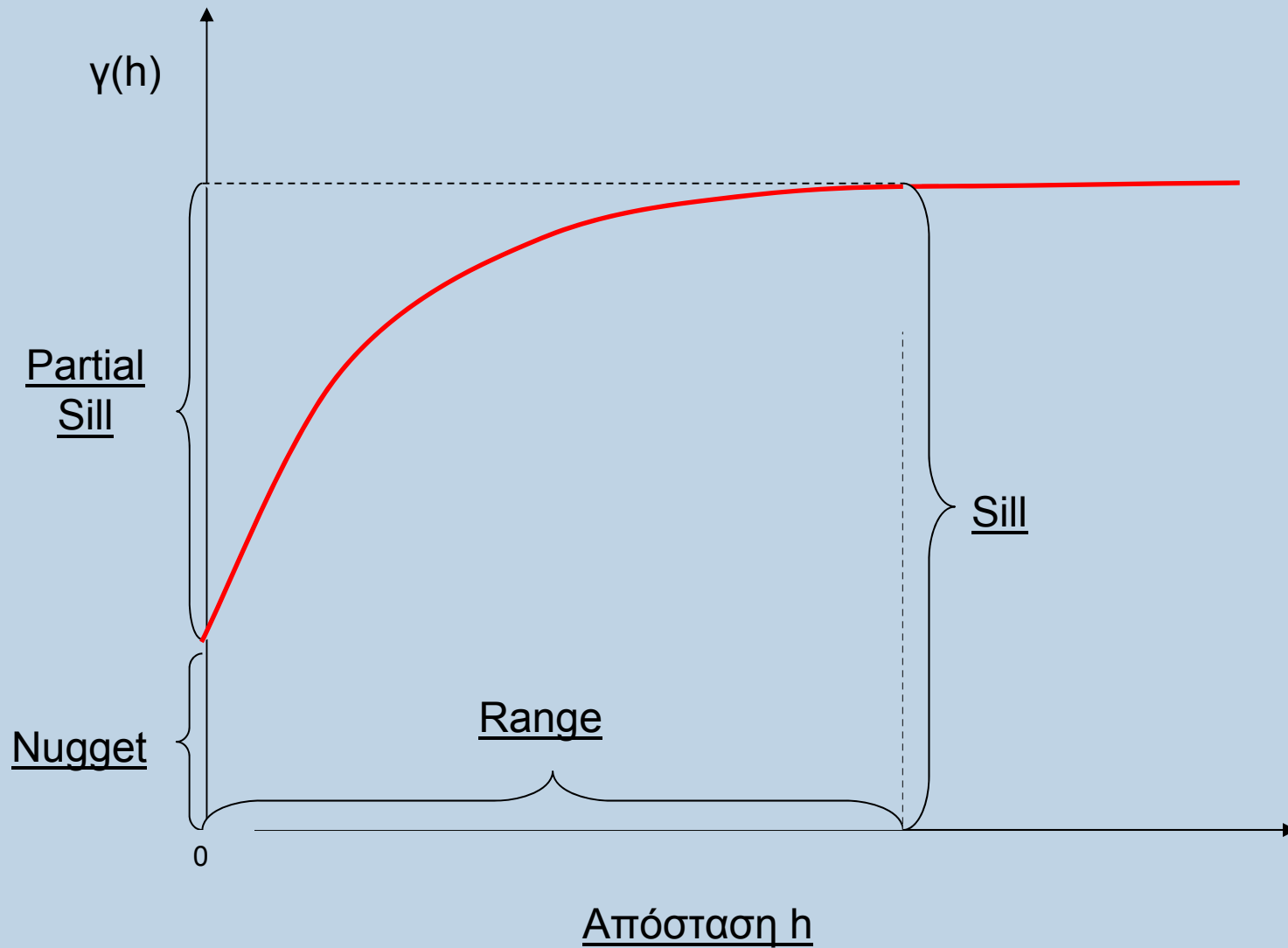
Source	DF	SS	MS	F	P
Regression	1	7044104	7044104	2.75	0.111
Residual Error	23	58826076	2557655		
Total	24	65870180			

### Regression Plot

$Y = 191.370 + 1.44512X$   
R-Sq = 10.7%



# Ημιβαριόγραμμα



# Μέθοδος παρεμβολής - kriging

Με βάση το ημιβαριόγραμμα υπολογίζονται τα βάρη των τιμών κάθε σημείου στο χώρο με βάση μετρήσεις στα σημεία παρατήρησης

$$\hat{f}(x^*) = \sum_{i=1}^n \lambda_i(x^*) f(x_i)$$