Μέτρα Διασποράς

Γιατί μετράμε την διασπορά;

Παράδειγμα

Δίνεται το ετήσιο ποσοστό κέρδους δύο επιχειρήσεων για 6 χρόνια. Αν έπρεπε να επιλέξετε την μετοχή μιας εκ των 2 με κριτήριο το ποσοστό κέρδους αυτά τα 6 χρόνια. Ποια θα ήταν η επιλογή;

Έτος	1η Επιχείρηση	2η Επιχείρηση
1990	5,2	7,9
1991	4,5	7,0
1992	3,9	-5,3
1993	4,8	14,2
1994	5,0	-11,0
1995	5,4	16,0

Ο Αριθμητικός Μέσος και στις δύο περιπτώσεις είναι 4,8

ΜΕΤΡΑ ΔΙΑΣΠΟΡΑΣ

- Εύρος
- Τεταρτημοριακή Απόκλιση ή Ημι-ενδοτεταρτημοριακό Εύρος
- Μέση Απόκλιση
- Διακύμανση και Τυπική Απόκλιση
- Συντελεστής Μεταβλητότητας
- Συντελεστής Gini

Εύρος

Η διαφορά μεγαλύτερης και μικρότερης τιμής

$$X_{\max} - X_{\min}$$

- **Πλεονέκτημα**: Εύκολος τρόπος υπολογισμού
- **Μειονέκτημα**: Εξαρτάται από τις ακραίες τιμές
- Δίνει μια πρώτη γρήγορη εικόνα της διασποράς
- Συνήθεις χρήσεις: Θερμοκρασίες Τιμές χρηματιστηρίου

  Σε στατιστικές που αφορούν ποιοτικούς ελέγχους στην βιομηχανία

Τεταρτημοριακή Απόκλιση

$$\frac{Q_3 - Q_1}{2}$$

- Λαμβάνει υπόψη το 50% των δεδομένων
- **Πλεονέκτημα**: Δεν επηρεάζεται από τις ακραίες τιμές
- Εξαρτάται αποκλειστικά από δύο τιμές

Μέση Απόκλιση

Αταξινόμητα στοιχεία

$$M.A = \frac{\sum_{i=1}^{n} \left| X_i - \overline{X} \right|}{n}$$

Ταξινομημένα στοιχεία

$$M.A = \frac{\sum_{i=1}^{k} f_i \left| X_i - \overline{X} \right|}{n}$$

- **Βασικό πλεονέκτημα:**
  εξαρτάται από όλες τις τιμές της μεταβλητής

- **Βασικό μειονέκτημα:**
  Περιορισμένη δυνατότητα αλγεβρικών υπολογισμών

Διακύμανση και Τυπική Απόκλιση

Το σημαντικότερο μέτρο διασποράς λόγω των σημαντικών μαθηματικών ιδιοτήτων που έχει

	Διακύμανση	Τυπική Απόκλιση
Πληθυσμός ⇒	$\sigma^2 = \frac{\sum_{i=1}^{N}(X_i - \mu)^2}{N}$	$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(X_i - \mu)^2}{N}}$
Δείγμα ⇒	$S^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n}$	$S = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n}}$
Αμερόληπτη Εκτίμηση του σ^2 ⇒	$S^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}$	$S = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}}$

Διακύμανση   Τυπική Απόκλιση

Αμερόληπτη Εκτίμηση του σ^2 για **Ταξινομημένα στοιχεία**

$$S^2 = \frac{\sum_{i=1}^{k} f_i (X_i - \overline{X})^2}{n-1}$$

$$S = \sqrt{\frac{\sum_{i=1}^{\kappa} f_i (X_i - \overline{X})^2}{n-1}}$$

- Μεγαλύτερες αποκλίσεις έχουν μεγαλύτερο βάρος από τις μικρότερες.

  Είναι προτιμότερες πολλές και μικρές αποκλίσεις από λίγες και μεγάλες.

- Η τυπική απόκλιση χρησιμοποιείται γιατί μετράται σε μονάδες μέτρησης της μεταβλητής

Παράδειγμα

X_i	$X_i - \overline{X}$	$(X_i - \overline{X})^2$
2.0	-1.9	3.5
2.1	-1.8	3.2
2.3	-1.6	2.5
2.3	-1.6	2.5
3.2	-0.7	0.5
3.2	-0.7	0.5
3.3	-0.6	0.3
3.3	-0.6	0.3
3.5	-0.4	0.1
3.5	-0.4	0.1
3.5	-0.4	0.1
3.7	-0.2	0.0
3.7	-0.2	0.0
3.7	-0.2	0.0
3.9	0.0	0.0
3.9	0.0	0.0
3.9	0.0	0.0
3.9	0.0	0.0
4.1	0.2	0.0
4.1	0.2	0.0
4.3	0.4	0.2
4.3	0.4	0.2
4.3	0.4	0.2
4.4	0.5	0.3
4.4	0.5	0.3
4.5	0.6	0.4
5.7	1.8	3.3
5.7	1.8	3.3
5.7	1.8	3.3
6.0	2.1	4.5
116.4	**0.0**	**29.8**

$$\overline{X} = \frac{116{,}4}{30} = 3{,}88$$

$$S_X^2 = \frac{29{,}8}{29} = 1{,}027$$

$$S_X = 1{,}013$$

Θεώρημα Chebyshev

Ανεξάρτητα από την μορφή της κατανομής μέσα στο διάστημα $\overline{X} \pm kS_X$ βρίσκεται τουλάχιστον το $1-\frac{1}{k^2}$ μέρος των τιμών της X

$$\Pr\left(\left|X_i - \overline{X}\right| \geq kS_X\right) \leq \frac{1}{k^2}$$

<u>Παράδειγμα</u>

στο διάστημα $\overline{X} \pm 2S_X$ βρίσκεται τουλάχιστον το 75%

στο διάστημα $\overline{X} \pm 3S_X$ βρίσκεται τουλάχιστον το 88%

Συντελεστής Μεταβλητότητας ή Σχετική Διασπορά

Αποτελεί σχετική μέτρηση της διασποράς και εκφράζεται ως ποσοστό επι τοις εκατό %

Πληθυσμός

$$CV = \frac{\sigma}{\mu}$$

Δείγμα

$$CV = \frac{S_X}{\overline{\overline{X}}}$$

Δεν εξαρτάται από τη μονάδα μέτρησης της μεταβλητής και επιτρέπει την σύγκριση μεταβλητών ή ομάδων δεδομένων που εκφράζονται σε διαφορετικές μονάδες μέτρησης

Σχετική Θέση τιμών

$$Z \frac{x - \overline{X}}{s}$$

Η στατιστική Ζ μας πληροφορεί πόσες τυπικές αποκλίσεις μία τιμή του δείγματος είναι πάνω ή κάτω από τον μέσο όρο του δείγματος.

Μέτρα Ανισοκατανομής (Συγκέντρωσης)

Η καμπύλη Lorenz και ο συντελεστής Gini

Μέτρο ανισοκατανομής ή συγκέντρωσης: Ελέγχουν πόσο έντονα ανισοκατανέμεται η συνολική τιμή μιας μεταβλητής (π.χ. Γεωργικό οικογενειακό εισόδημα, Γεωργική Γη) στα μέλη ενός πληθυσμού (π.χ. στους νομούς της Ελλάδος).

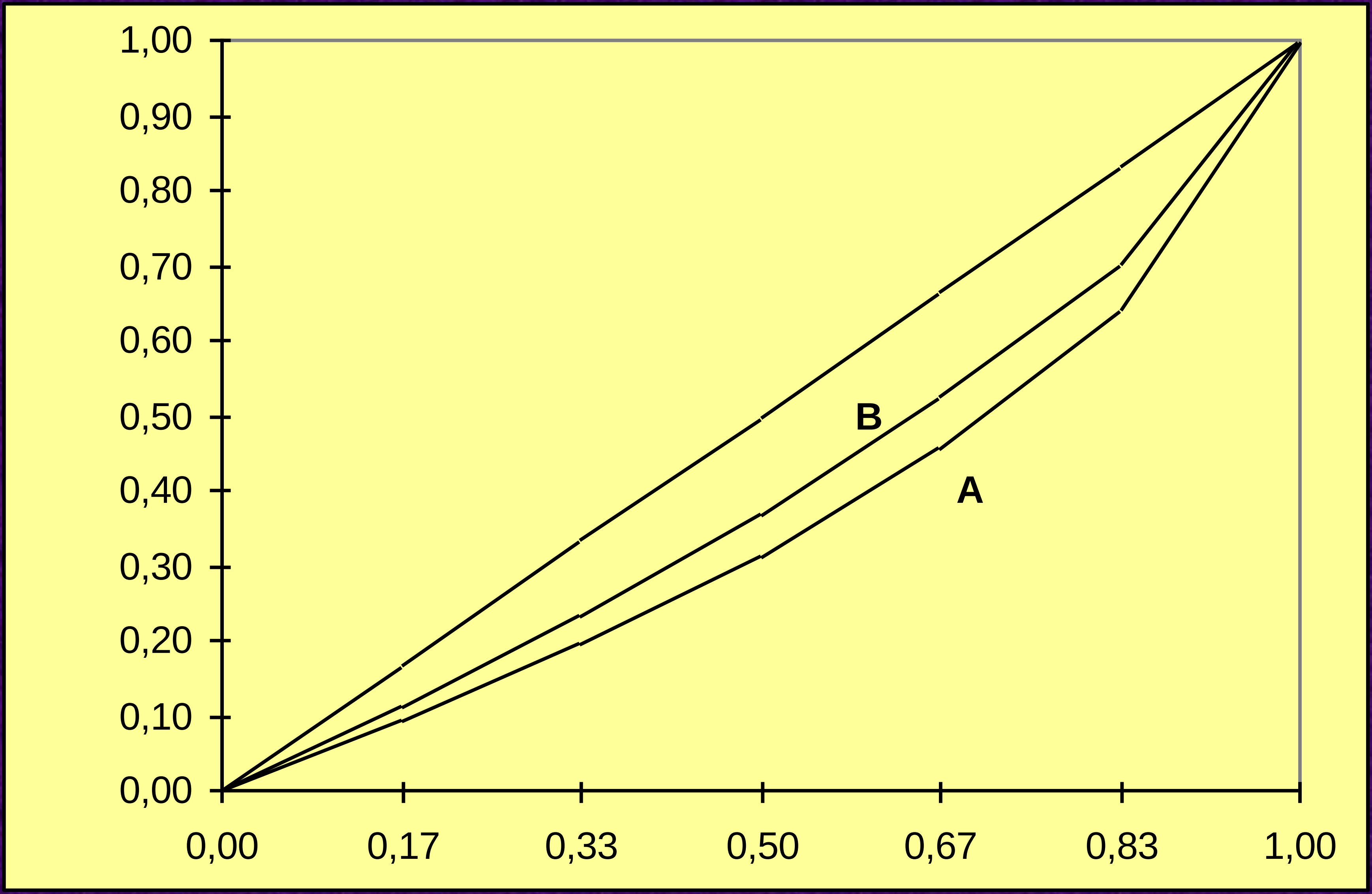
Η ένταση της ανισοκατανομής ελέγχεται διαγραμματικά με την καμπύλη Lorenz είτε υπολογιστικά με τον συντελεστή Gini

Η καμπύλη Lorenz είναι μια μέθοδος γραφικής απεικόνισης της ανισότητας.
Βασίζεται στην σχέση της ποσοστιαίας αθροιστικής κατανομής ενός μεγέθους με την κατανομή των μονάδων που αντιστοιχούν σ΄ αυτή.

Παράδειγμα

Τα μηνιαία εισοδήματα 6 εργαζομένων είναι: 190, 165, 300, 230, 150 και 580 χιλ. δρχ. Να κατασκευαστεί η καμπύλη Lorenz. Να συγκριθεί με την καμπύλη που προκύπτει από τα εισοδήματα 160 , 170, 190, 220, 250 και 420 χιλ. δρχ.

Y_i	f_i	F_i	$F_i\%$	Φ_i	$\Phi_i\%$
150	1	1	(1/6) = 0.1666	150	(150/1615) =0.0929
165	1	2	(2/6) = 0.3333	315	(315/1615) =0.1950
190	1	3	(3/6) = 0.5000	505	(505/1615) = 0.3127
230	1	4	(4/6) = 0.6666	735	(735/1615) = 0.4551
300	1	5	(5/6) = 0.8333	1035	(1035/1615)=0.6409
580	1	6	(6/6) =1.0000	1615	(1615/1615)=1.0000
1615	6				



1,00
0,90
0,80
0,70
0,60
0,50
0,40
0,30
0,20
0,10
0,00
B
A
0,00
0,17
0,33
0,50
0,67
0,83
1,00

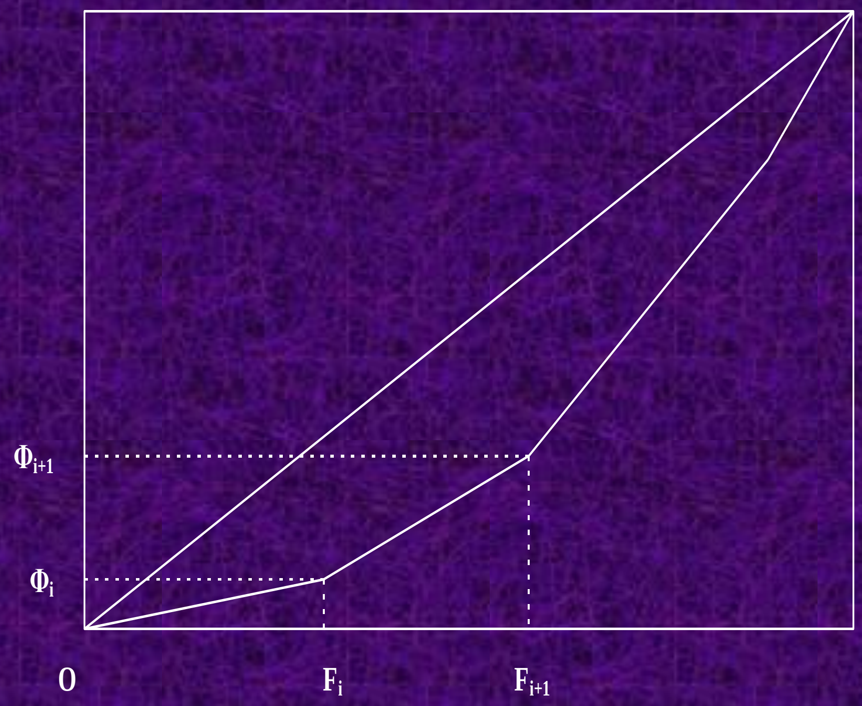
Συντελεστής Gini

- **Δείκτης ανισότητας** που βασίζεται στην καμπύλη Lorenz. Δίνει το λόγο του εμβαδού ανισοκατανομής που είναι με μεγάλη προσέγγιση το εμβαδόν του κάτω δεξιά τριγώνου της καμπύλης Lorenz.

- Όσο μεγαλύτερη είναι η ανισότητα τόσο μεγαλύτερο είναι και το εμβαδόν μεταξύ της καμπύλης Lorenz και της διαγωνίου.

- Ο συντελεστής Gini αντιπροσωπεύει το μερίδιο του εμβαδού μεταξύ της καμπύλης Lorenz και της διαγωνίου στο συνολικό εμβαδόν του τριγώνου που ορίζεται από την διαγώνιο και τους δύο άξονες.

1η μέθοδος (Γεωμετρική)



Εμβαδόν κάτω από την καμπύλη Lorenz και μεταξύ των σημείων F_i και F_{i+1}

$$\left(F_{i+1} - F_i\right)\Phi_i + \frac{1}{2}\left(F_{i+1} - F_i\right)\left(\Phi_{i+1} - \Phi_i\right) = \frac{1}{2}\left(F_{i+1} - F_i\right)\left(\Phi_{i+1} + \Phi_i\right)$$

Εμβαδόν κάτω από την καμπύλη Lorenz

$$\frac{1}{2}\sum_{i=0}^{n-1}\left(F_{i+1}-F_i\right)\left(\Phi_{i+1}+\Phi_i\right)$$

Το εμβαδόν μεταξύ καμπύλης Lorenz και διαγωνίου

$$\frac{1}{2}-\frac{1}{2}\sum_{i=0}^{n-1}\left(F_{i+1}-F_i\right)\left(\Phi_{i+1}+\Phi_i\right)$$

Συντελεστής Gini

$$G_1=1-\sum_{i=0}^{n-1}\left(F_{i+1}-F_i\right)\left(\Phi_{i+1}+\Phi_i\right)$$

Παράδειγμα

Y_i	f_i	F_i	Φ_i	$(F_{i+1}-F_i)\ (\Phi_{i+1}+\Phi_i)$
150	1	0.1666	0.0929	0.0154
165	1	0.3333	0.1950	0.0479
190	1	0.5000	0.3127	0.0846
230	1	0.6666	0.4551	0.1280
300	1	0.8333	0.6409	0.1827
580	1	1.0000	1.0000	0.2734
1615	6			0.7322

$$G_1 = 1 - \sum_{i=0}^{n-1} \left(F_{i+1} - F_i\right)\left(\Phi_{i+1} + \Phi_i\right) = 1 - 0.7322 = 0.2678$$

Παράδειγμα: Άμεσων ενισχύσεων

(€*1000)	EL	ES	IT	FR	UK	EU25
0 - 1250 €	217,751	198,268	377,987	37,391	14,604	1,595,300
1250 € - 2000 €	134,564	137,591	182,105	38,526	16,369	762,004
2000 € - 5000 €	397,189	493,268	537,666	190,930	95,798	2,643,490
5000 € - 10000 €	349,804	620,605	552,098	468,064	196,284	3,669,472
10000 € - 20000 €	238,447	878,768	600,671	1,239,434	372,773	5,452,290
20000 € - 50000 €	97,006	1,132,224	611,125	3,223,712	870,141	8,808,451
50000€ - 100000 €	17,237	503,844	351,516	1,842,704	867,617	4,912,290
100000 € - 200000 €	4,625	287,169	225,889	416,712	650,571	2,338,809
200000 € -300000 €	1,195	83,466	77,317	35,786	199,706	733,157
300000-500000 €	1,180	52,891	68,086	8,709	122,875	657,373
>500000 €	1,184	88,514	107,913	12,940	71,655	932,280
Total	1,460,182	4,476,608	3,692,373	7,514,908	3,478,393	32,504,916

(number*1000)	EL	ES	IT	FR	UK	EU25
0 - 1250 €	552.96	465.8	984.15	71.3	24.65	4359.24
1250 € - 2000 €	84.83	86.29	115.03	24.03	10.15	480.24
2000 € - 5000 €	125.32	152.91	169.38	56.96	28.44	815.46
5000 € - 10000 €	50.98	88.15	79.15	63.63	27.15	518
10000 € - 20000 €	17.87	62.01	43.43	84.81	25.98	384.69
20000 € - 50000 €	3.52	38.38	20.48	102.69	27.32	288.49
50000€ - 100000 €	0.27	7.47	5.14	27.98	12.53	73.22
100000 € - 200000 €	0.04	2.15	1.68	3.36	4.84	17.64
200000 € -300000 €	0.01	0.35	0.32	0.16	0.84	3.07
300000-500000 €	0	0.14	0.18	0.03	0.33	1.73
>500000 €	0	0.08	0.11	0.01	0.09	1.06
Total	835.8	903.73	1419.05	434.96	162.32	6942.84

Αθροιστικές Ποσοστιαίες Συχνότητες

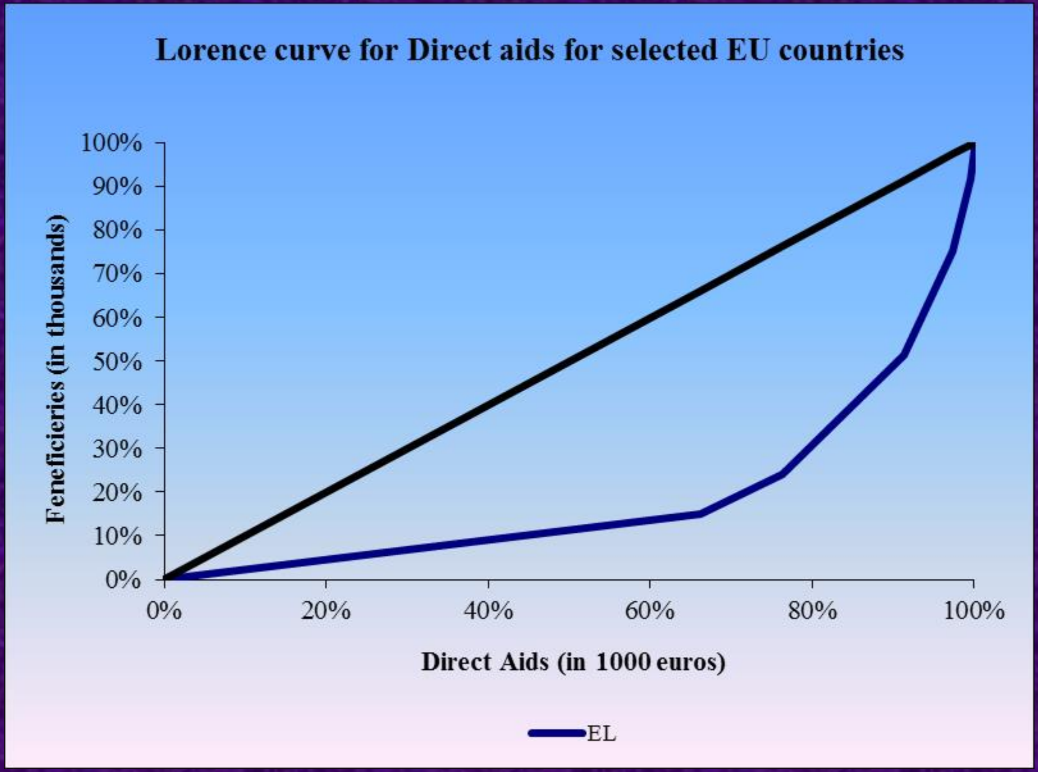
% cumulative

EL	ES	IT	FR	UK	EU25
14.9%	4.4%	10.2%	0.5%	0.4%	4.9%
24.1%	7.5%	15.2%	1.0%	0.9%	7.3%
51.3%	18.5%	29.7%	3.6%	3.6%	15.4%
75.3%	32.4%	44.7%	9.8%	9.3%	26.7%
91.6%	52.0%	61.0%	26.3%	20.0%	43.4%
98.3%	77.3%	77.5%	69.2%	45.0%	70.5%
99.4%	88.6%	87.0%	93.7%	70.0%	85.7%
99.8%	95.0%	93.1%	99.2%	88.7%	92.9%
99.8%	96.8%	95.2%	99.7%	94.4%	95.1%
99.9%	98.0%	97.1%	99.8%	97.9%	97.1%
100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

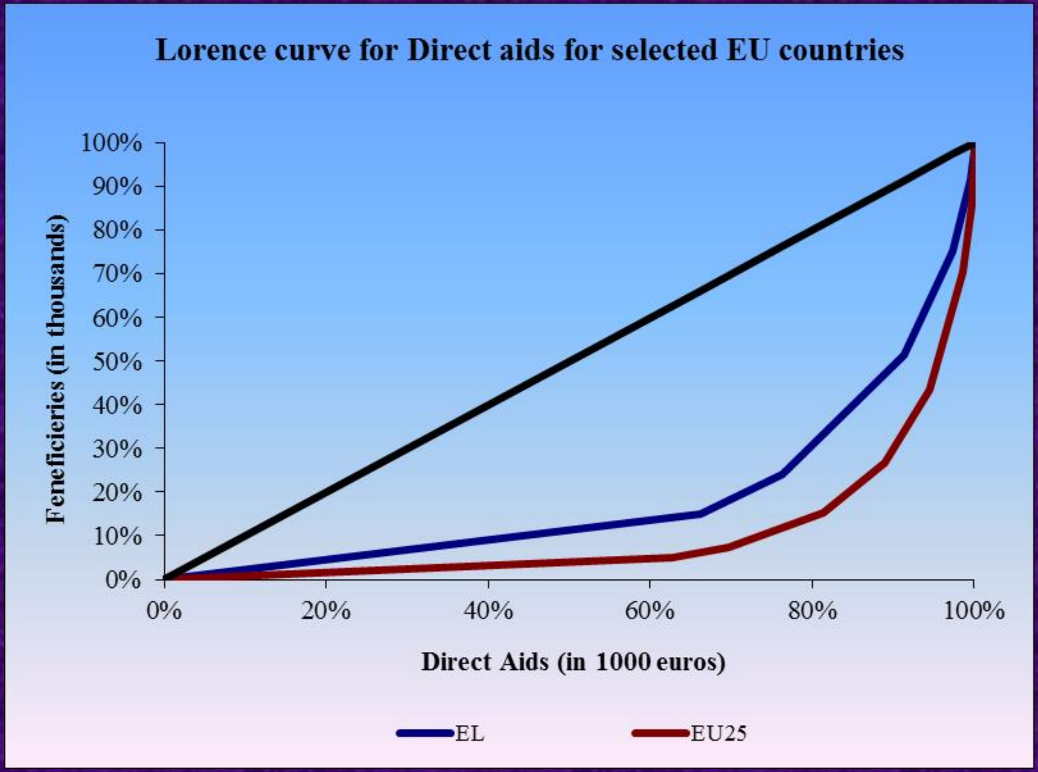
% cumulative

EL	ES	IT	FR	UK	EU25
66.16%	51.54%	69.35%	16.39%	15.19%	62.79%
76.31%	61.09%	77.46%	21.92%	21.44%	69.70%
91.30%	78.01%	89.40%	35.01%	38.96%	81.45%
97.40%	87.76%	94.97%	49.64%	55.69%	88.91%
99.54%	94.63%	98.03%	69.14%	71.69%	94.45%
99.96%	98.87%	99.48%	92.75%	88.52%	98.61%
99.99%	99.70%	99.84%	99.18%	96.24%	99.66%
100.00%	99.94%	99.96%	99.95%	99.22%	99.92%
100.00%	99.98%	99.98%	99.99%	99.74%	99.96%
100.00%	99.99%	99.99%	100.00%	99.94%	99.98%
100.00%	100.00%	100.00%	100.00%	100.00%	100.00%

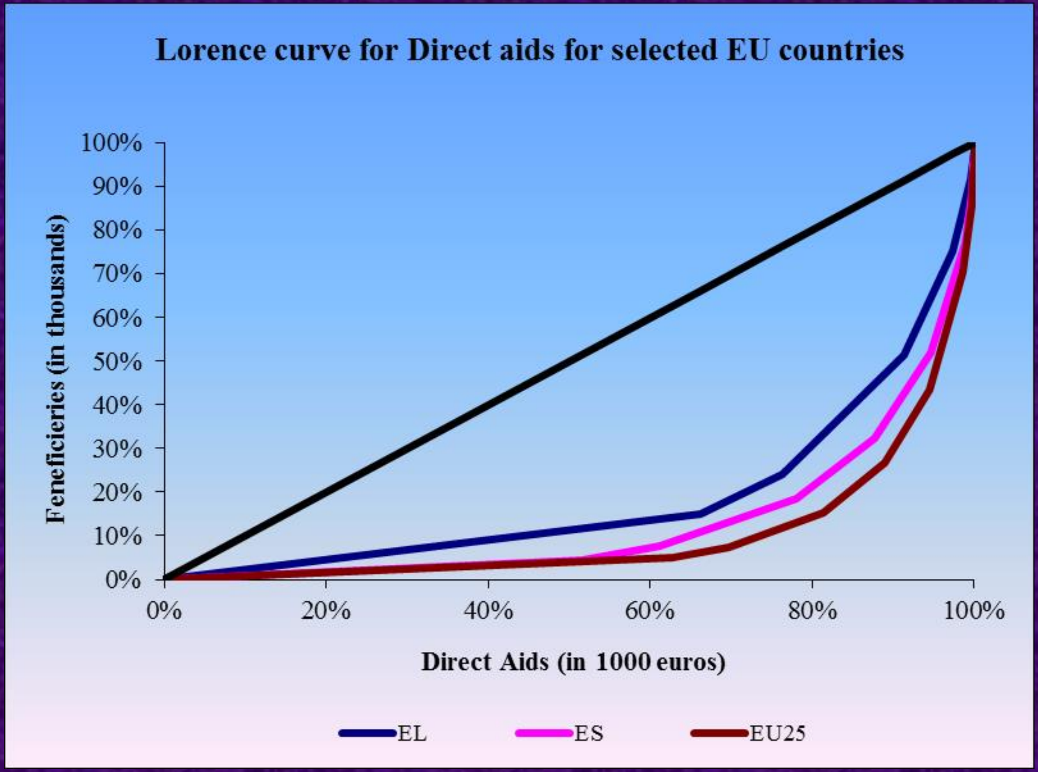
Καμπύλη Lorenze



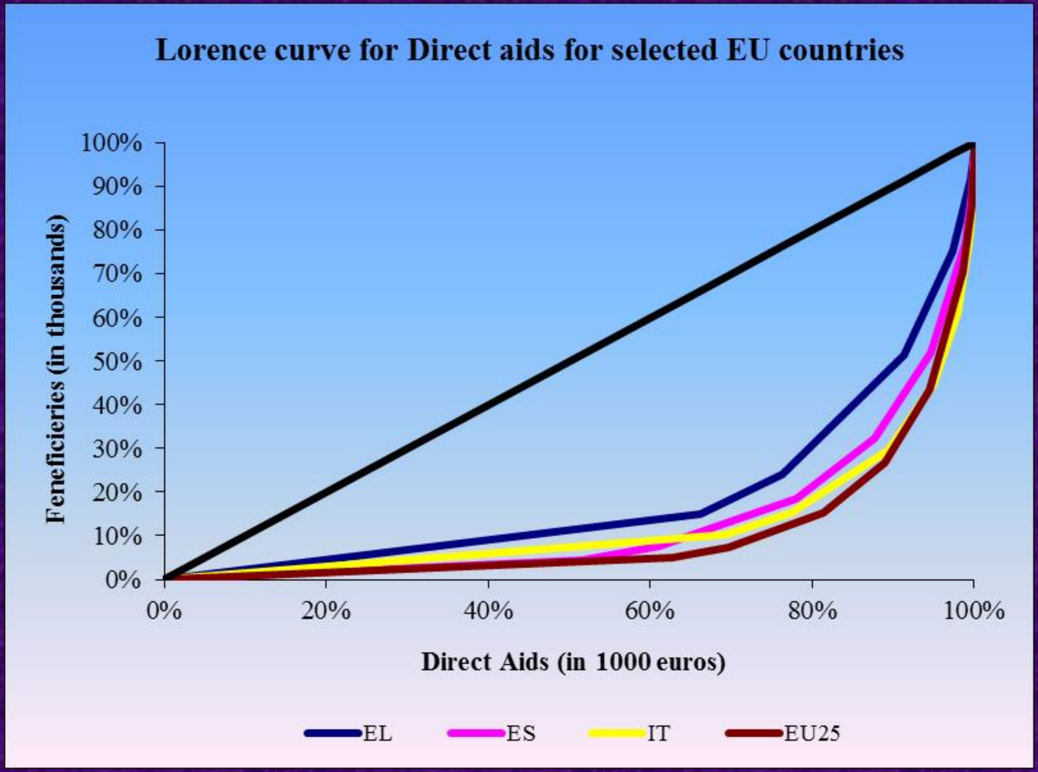
Καμπύλη Lorenze



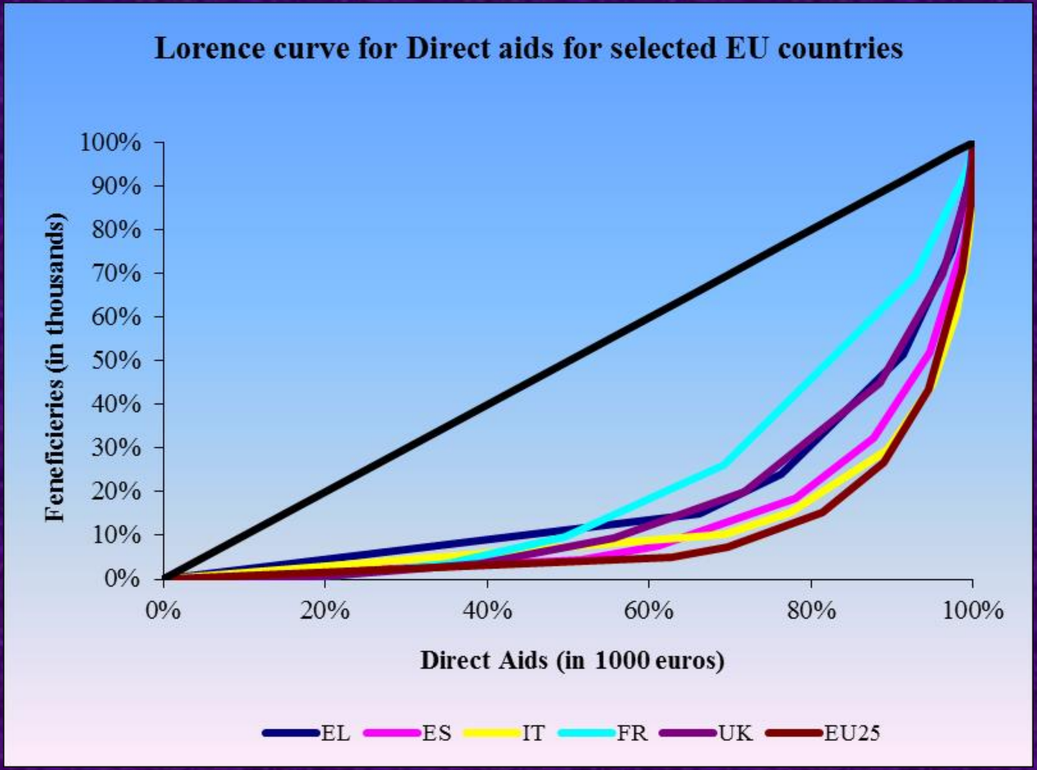
Καμπύλη Lorenze



Καμπύλη Lorenze

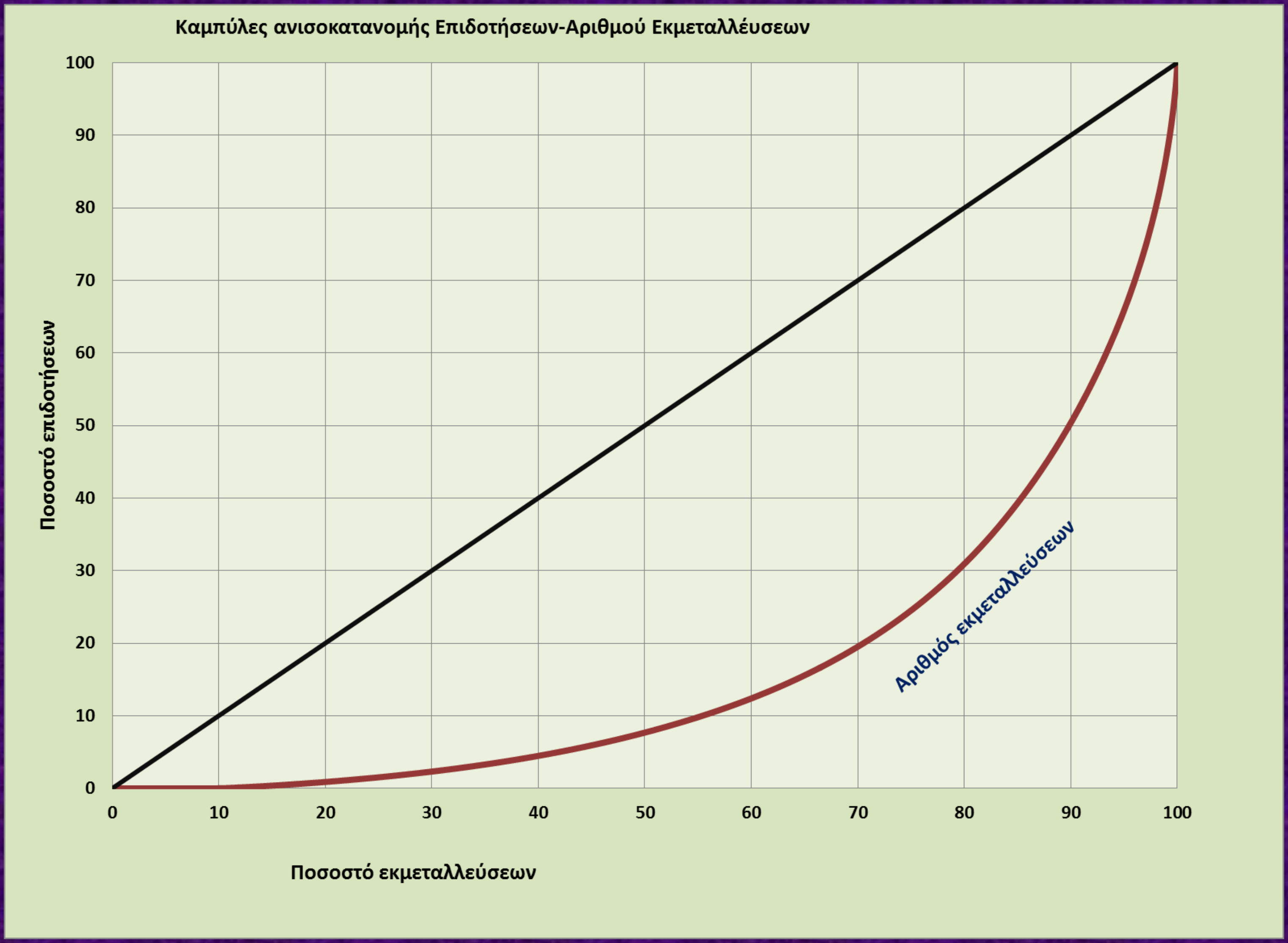


Καμπύλη Lorenze

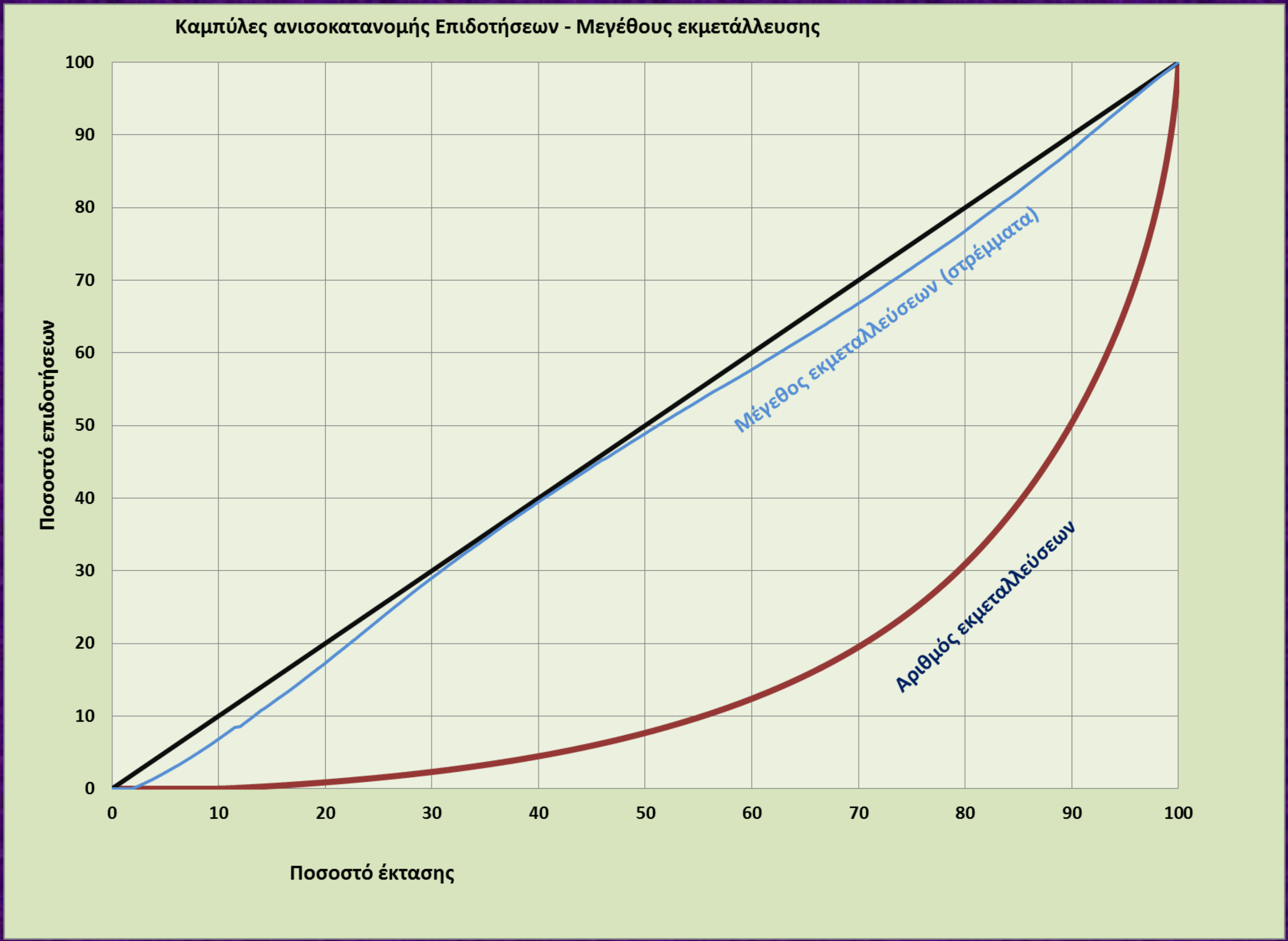


Συντελεστής Gini	EL	ES	IT	FR	UK	EU25
	0.373	0.260	0.248	0.443	0.339	0.206

Ανισοκατανομή των ενισχύσεων;



Ανισοκατανομή των ενισχύσεων;



2η μέθοδος (Kendall and Stuart)

Υπολογίζουμε πρώτα την μέση απόλυτη διαφορά όλων των τιμών της μεταβλητής

$$d = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} \left| Y_i - Y_j \right|}{n^2}$$

$$G_2 = \frac{d}{2\mu}$$

Παράδειγμα:

$\lvert Y_i - Y_j \rvert$	**150**	**165**	**190**	**230**	**300**	**580**
150	0	15	40	80	150	430
165	15	0	25	65	135	415
190	40	25	0	40	110	390
230	80	65	40	0	70	350
300	150	135	110	70	0	280
580	430	415	390	350	280	0

$$d = \frac{5190}{36} = 144{,}2$$

$$\mu = 269{,}1$$

$$G_2 = \frac{144{,}2}{2 * 269{,}1} = 0{,}2678$$

Μέτρηση Ασυμμετρίας

$\bar{X} - M > 0$ Θετική ασυμμετρία

$\bar{X} - M = 0$ Μηδενική ασυμμετρία (συμμετρία)

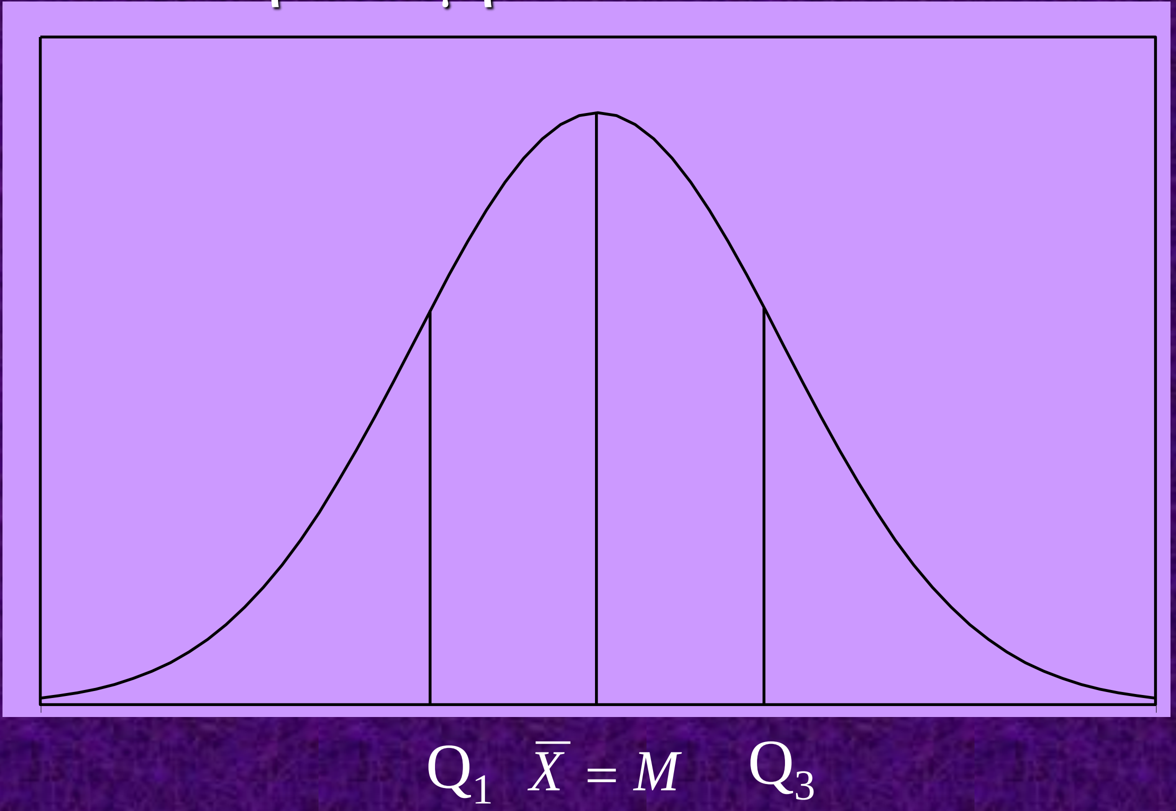
$\bar{X} - M < 0$ Αρνητική ασυμμετρία

Δείκτης Ασυμμετρίας S_k (κατά Bowley)

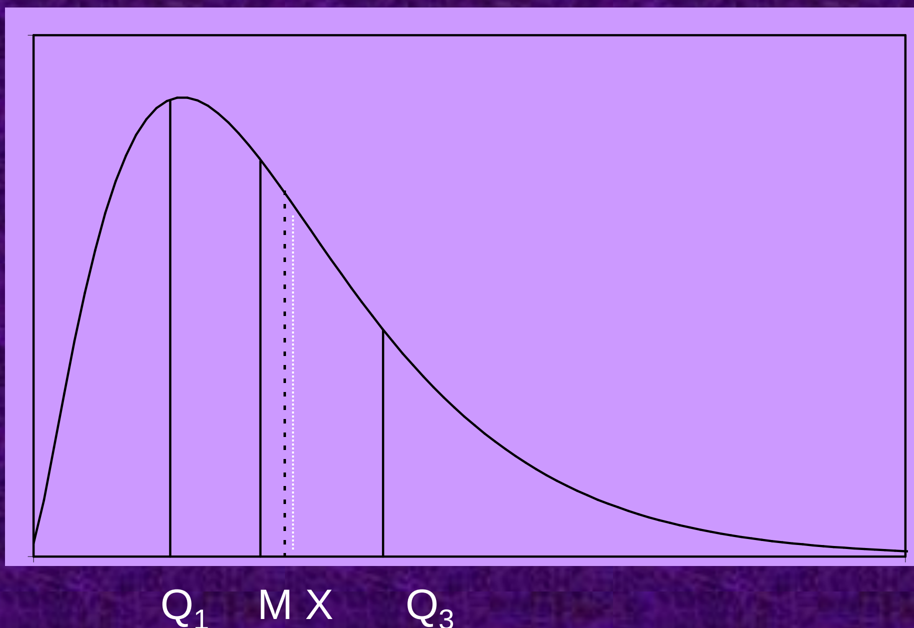
$$\mathbf{S_k = \frac{(Q_3 - M) - (M - Q_1)}{(Q_3 - M) + (M - Q_1)} = \frac{Q_3 + Q_1 - 2M}{Q_3 - Q_1}}$$

Είδη Συμμετρικών Κατανομών

Κανονική κατανομή



Θετική Ασυμμετρία



Αρνητική Ασυμμετρία