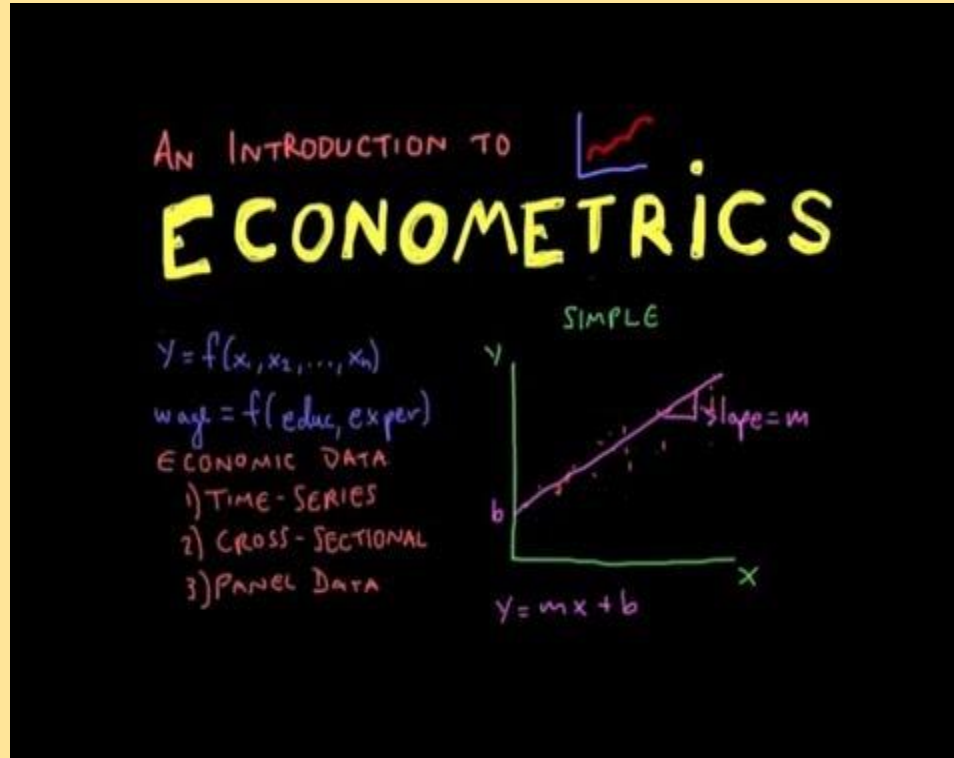


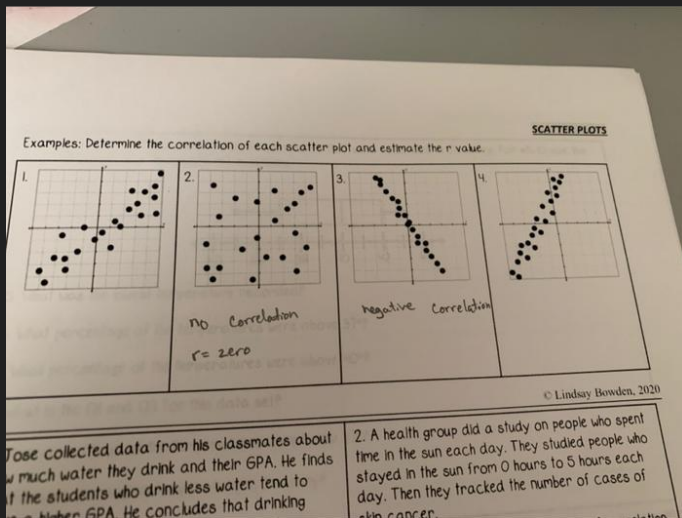
Οικονομικό Πανεπιστήμιο Αθηνών

Μάθημα: "Εφαρμογές Στατιστικών Μεθόδων σε Επιχειρησιακά Προβλήματα"

Συνάντηση: 2η



Δρ. Τρύφωνας Λεμοντζόγλου



ΕΦΑΡΜΟΓΕΣ (χρήση πραγματικών δεδομένων)

Συσχέτιση (Γραμμική) Δύο Μεταβλητών

Μπορείς να κατασκευάσεις μόνη/ος σου ένα
διάγραμμα διασποράς και τον πίνακα με τους
συντελεστές γραμμικής συσχέτισης ;;;



Βήμα 1ο: Επέλεξε μια βάση δευτερογενών
δεδομένων !!!

Προσπάθησε στη βάση δεδομένων της Παγκόσμιας
Τράπεζας ...

<https://data.worldbank.org/indicator>

Συσχέτιση (Γραμμική) Δύο Μεταβλητών

Βήμα 2ο: Επέλεξε την **εξαρτημένη** (Ψ) και την **ανεξάρτητη** (X) μεταβλητή σου!!!
(σκέψου τη διερεύνηση μιας σχέσης που έχει νόημα)



Συσχέτιση (Γραμμική) Δύο Μεταβλητών

Βήμα 3ο: Επέλεξε μια χώρα (ή ένα γκρουπ χωρών)
και όρισε το χρονικό διάστημα των παρατηρήσεων!!!

n=12



Έτη	(Ψ) GDP growth (annual %)	(Χ) international tourism, number of arrivals
1995	2,09	10.712.000
1996	1,38	9.782.000
...	...	...
2005	1,18	15.938.000
2006	6,44	17.284.000



Παράδειγμα **ΘΕΤΙΚΗΣ**
ΣΥΣΧΕΤΙΣΗΣ:
Τουρισμός
και Οικονομική
Ανάπτυξη

Έτη	Ρυθμοί ανάπτυξης (%)	Αφίξεις (αριθμός)
1995	2,099720313	10712000
1996	1,382470172	9782000
1997	3,6028832	10588000
1998	3,436935726	11364000
1999	2,65878589	12606000
2000	4,137827099	13567000
2001	4,650414122	14678000
2002	4,683137359	14918000
2003	5,796823035	14785000
2004	5,377866297	14268000
2005	1,183506102	15938000
2006	6,44339775	17284000
	3,787813922	13374166,67

$(X-X_{\text{μέσο}})$	$(\Psi-\Psi_{\text{μέσο}})$	$(X-X_{\text{μέσο}})*(\Psi-\Psi_{\text{μέσο}})$
-2662166,67	-1,688093609	4493986,541
-3592166,67	-2,40534375	8640395,648
-2786166,67	-0,1849307216	515247,8127
-2010166,67	-0,3508781962	705323,6552
-768166,67	-1,129028032	867281,7035
192833,33	0,3500131775	67494,20655
1303833,33	0,8626002003	1124686,892
1543833,33	0,895323437	1382230,163
1410833,33	2,009009113	2834377,016
893833,33	1,590052375	1421241,809
2563833,33	-2,60430782	-6677011,19
3909833,33	2,655583828	10382890,16
		25758144,42

$(X-X_{\text{μέσο}})^2$	$(\Psi-\Psi_{\text{μέσο}})^2$
7087131378859	2,849660031
12903661385059	5,785678555
7762724713019	0,03419937178
4040770041179	0,1231155086
590080032899	1,274704297
37184693159	0,1225092244
1699981352419	0,7440791056
2383421350819	0,8016040569
1990450685039	4,036117615
798938021819	2,528266556
6573241344019	6,782419221
15286796668379	7,05212547
61154381666667	32,13447901

Βήμα 4ο: Κάνε εισαγωγή των τιμών των μεταβλητών σου!!!

Βήμα 5ο: Υπολόγισε τον συντελεστή συσχέτισης!!!

συντελεστής
συσχέτισης

Έτη	Ρυθμοί ανάπτυξης (%)	Αφίξεις (αριθμός)
1995	2,099720313	10712000
1996	1,382470172	9782000
1997	3,6028832	10588000
1998	3,436935726	11364000
1999	2,65878589	12606000
2000	4,137827099	13567000
2001	4,650414122	14678000
2002	4,683137359	14918000
2003	5,796823035	14785000
2004	5,377866297	14268000
2005	1,183506102	15938000
2006	6,44339775	17284000

0,5810522036

=CORREL(D3:D14;E3:E14)

Μπορείς να δουλέψεις ελεύθερα στο περιβάλλον των φύλλων εργασίας της google.

Κάνε εισαγωγή των τιμών των μεταβλητών σου.

Σε ένα κενό κελί πάτησε το **=** και χρησιμοποίησε την εντολή **correl**.

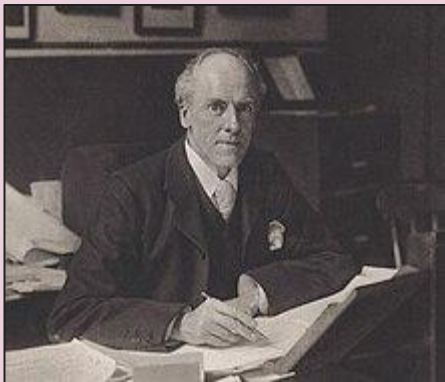
Στην **()** επέλεξε πρώτα τις τιμές της πρώτης μεταβλητής.

Βάλε το σύμβολο **;** και έπειτα επέλεξε τις τιμές της άλλης μεταβλητής.

Εναλλακτικά, επέλεξε **“basic stats”** → **“correlation”**

Βάλε τις τιμές των μεταβλητών σου στις δύο στήλες.

Έπειτα, πάτησε **“calculate”**



Karl Pearson
(1857-1936)



Pearson with Sir Francis Galton

<https://www.statskingdom.com/index.html>

Pearson correlation coefficient (r)	0.5811
r^2	0.3376
P-value	0.04756
Covariance	2341649.4928
Sample size (n)	12
Statistic	2.2577

Το περιβάλλον αυτής της εφαρμογής αποδίδει αυτόματα το διάγραμμα διασποράς όταν υπολογίζει τον συντελεστή συσχέτισης.

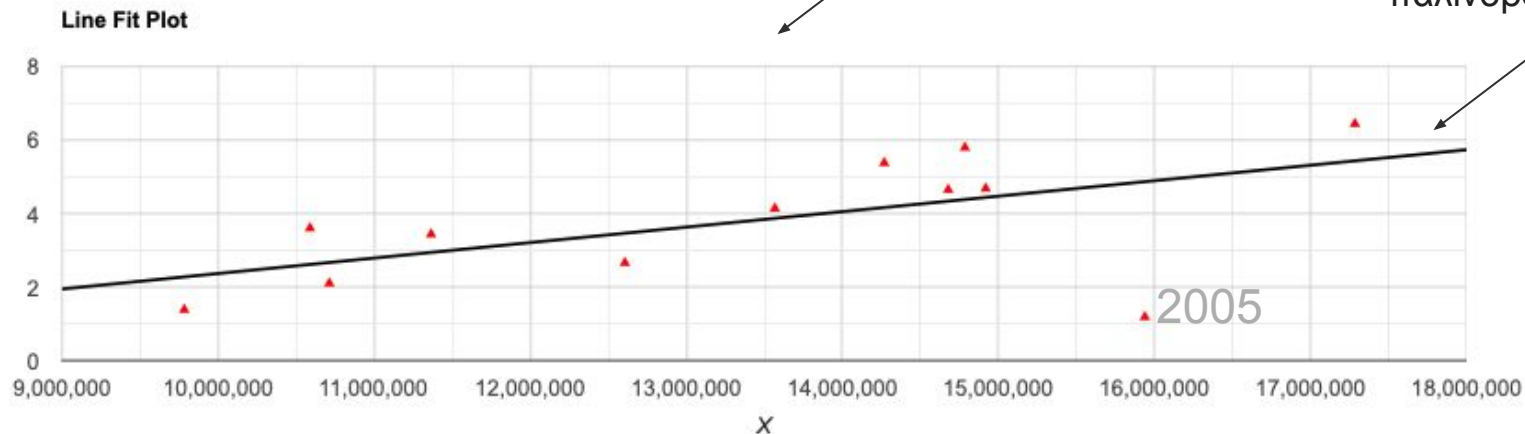


$r = 0.5811$

συντελεστής
συσχέτισης

διάγραμμα
διασποράς

γραμμή
παλινδρόμησης





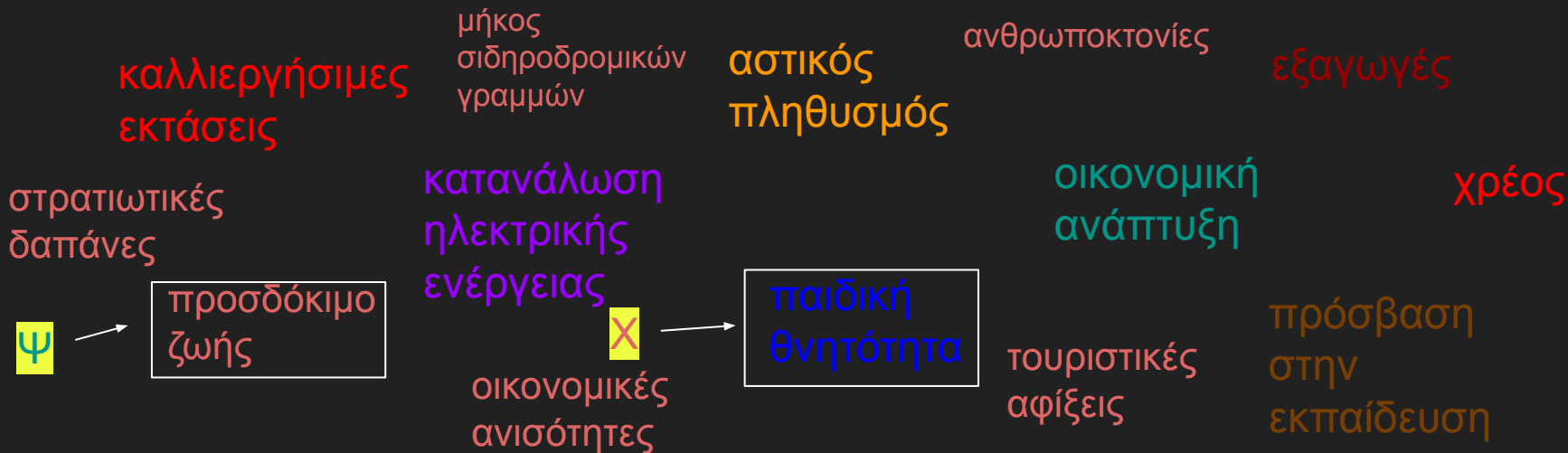
Παράδειγμα **ΑΡΝΗΤΙΚΗΣ**
ΣΥΣΧΕΤΙΣΗΣ: [REDACTED]

Προσδόκιμο ζωής
και [REDACTED]

Παιδική θνητότητα

Συσχέτιση (Γραμμική) Δύο Μεταβλητών

Ας εξετάσουμε τώρα την ύπαρξη μιας ΑΡΝΗΤΙΚΗΣ ΣΥΣΧΕΤΙΣΗΣ



Συσχέτιση (Γραμμική) Δύο Μεταβλητών

		Life expectancy at birth, total (years)	Mortality rate, neonatal (per 1,000 live births)
2020	Albania	76,989	7,2
2020	Argentina	75,892	5,9
2020	Bulgaria	73,65853659	3
2020	Brazil	74,009	8,8
2020	Canada	81,6604878	3,4
2020	Chile	79,377	4,5
2020	China	78,077	3,5
2020	Cyprus	81,391	1,9
2020	Czechia	78,17560976	1,5
2020	Germany	81,04146341	2,3
2020	Spain	82,33170732	1,8
2020	Estonia	78,59512195	0,9
2020	France	82,17560976	2,5

Συσχέτιση (Γραμμική) Δύο Μεταβλητών

(X-X_μέσο)	(Ψ-Ψ_μέσο)	(X-X_μέσο)*(Ψ-Ψ_μέσο)	(X-X_μέσο)^2	(Ψ-Ψ_μέσο)^2
3,569230769	-1,73204128	-6,18205503	12,73940828	2,999966996
2,269230769	-2,82904128	-6,419747519	5,149408283	8,003474564
-0,630769231	-5,062504695	3,193272193	0,3978698228	25,62895378
5,169230769	-4,71204128	-24,35762877	26,72094674	22,20333302
-0,230769231	2,939446525	-0,6783338141	0,05325443798	8,640345873
0,869230769	0,65595872	0,5701795026	0,7555621298	0,4302818423
-0,130769231	-0,64404128	0,08422078292	0,01710059178	0,4147891703
-1,730769231	2,66995872	-4,621082401	2,995562131	7,128679567
-2,130769231	-0,5454315239	1,162188709	4,540177516	0,2974955473
-1,330769231	2,320422135	-3,08794638	1,770946746	5,384358883
-1,830769231	3,610666037	-6,610296284	3,351715977	13,03690923
-2,730769231	-0,1259193288	0,3438566286	7,457100593	0,01585567736
-1,130769231	3,454568476	-3,906319739	1,278639054	11,93404336
-0,0000000023	-0,0000000420	-3,885360932	5,171360947	8,162960578

Συσχέτιση (Γραμμική) Δύο Μεταβλητών

Results

Parameter	Value
Pearson correlation coefficient (r)	-0.598
r^2	0.3576
P-value	0.03087
Covariance	-4.2091
Sample size (n)	13
Statistic	-2.4746

Συσχέτιση (Γραμμική) Δύο Μεταβλητών



↓ $r = -0.598$

