

ΕΡΓΑΣΙΑ SPSS

ΜΑΡΙΟΣ ΤΖΑΜΤΖΗΣ

ΑΕΜ: 10038

Email: tzamtzis@ece.auth.gr

ΧΩΡΑ Α:

ΒΕΛΓΙΟ

Στο Βέλγιο έχουμε κορύφωση της καμπύλης των ημερήσιων κρουσμάτων την 29^η Οκτωβρίου, κατά την οποία καταγράφηκαν 23921 νέα κρούσματα και κορύφωση της καμπύλης των ημερησίων θανάτων την 10^η Νοεμβρίου με 345 νέους θανάτους. Στον παρακάτω πίνακα βρίσκονται τα δεδομένα του Βελγίου για τις επόμενες 20 ημέρες από την ημερομηνία κορύφωσης της καμπύλης των ημερήσιων κρουσμάτων:

Country	Date	New Cases	Date	Deaths	Case Fatality Rate	CFR (%)
Belgium	2020-10-30	20056	2020-11-11	197	0,009822497	0,98%
Belgium	2020-10-31	16915	2020-11-12	133	0,007862844	0,79%
Belgium	2020-11-01	11789	2020-11-13	215	0,01823734	1,82%
Belgium	2020-11-02	6337	2020-11-14	197	0,031087265	3,11%
Belgium	2020-11-03	5186	2020-11-15	118	0,022753567	2,28%
Belgium	2020-11-04	15672	2020-11-16	195	0,012442573	1,24%
Belgium	2020-11-05	11128	2020-11-17	223	0,02003954	2,00%
Belgium	2020-11-06	8703	2020-11-18	186	0,021371941	2,14%
Belgium	2020-11-07	6124	2020-11-19	171	0,027922926	2,79%
Belgium	2020-11-08	6621	2020-11-20	156	0,023561396	2,36%
Belgium	2020-11-09	2393	2020-11-21	170	0,071040535	7,10%
Belgium	2020-11-10	4293	2020-11-22	96	0,022361985	2,24%
Belgium	2020-11-11	7916	2020-11-23	137	0,017306721	1,73%
Belgium	2020-11-12	5002	2020-11-24	183	0,036585366	3,66%
Belgium	2020-11-13	4619	2020-11-25	139	0,030093094	3,01%
Belgium	2020-11-14	6268	2020-11-26	142	0,022654754	2,27%
Belgium	2020-11-15	4659	2020-11-27	120	0,0257566	2,58%
Belgium	2020-11-16	1932	2020-11-28	122	0,063146998	6,31%
Belgium	2020-11-17	2734	2020-11-29	86	0,031455743	3,15%
Belgium	2020-11-18	5182	2020-11-30	98	0,018911617	1,89%

ΧΩΡΑ Β:

ΗΝΩΜΕΝΟ ΒΑΣΙΛΕΙΟ

Για το Ηνωμένο Βασίλειο έχουμε κορύφωση της καμπύλης των ημερήσιων κρουσμάτων την 8η Ιανουαρίου του 2021, κατά την οποία καταγράφηκαν 68192 νέα κρούσματα και κορύφωση των ημερησίων θανάτων την 21η Ιανουαρίου 2021 με 1826 νέους θανάτους. Στον παρακάτω πίνακα βρίσκονται τα δεδομένα του Ηνωμένου Βασιλείου για τις επόμενες 20 ημέρες από την ημερομηνία κορύφωσης της καμπύλης των ημερήσιων κρουσμάτων:

Country	Date	New Cases	Date	Deaths	Case Fatality Rate	CFR(%)
UK	2021-01-09	60098	2021-01-21	1296	0,021564778	2,16%
UK	2021-01-10	55026	2021-01-22	1401	0,025460691	2,55%
UK	2021-01-11	46275	2021-01-23	1352	0,02921664	2,92%
UK	2021-01-12	45648	2021-01-24	611	0,013385033	1,34%
UK	2021-01-13	47662	2021-01-25	594	0,012462759	1,25%
UK	2021-01-14	48804	2021-01-26	1636	0,033521842	3,35%
UK	2021-01-15	55885	2021-01-27	1726	0,030884853	3,09%
UK	2021-01-16	41428	2021-01-28	1239	0,029907309	2,99%
UK	2021-01-17	38670	2021-01-29	1248	0,03227308	3,23%
UK	2021-01-18	37610	2021-01-30	1205	0,032039351	3,20%
UK	2021-01-19	33454	2021-01-31	590	0,017636157	1,76%
UK	2021-01-20	38992	2021-02-01	407	0,010438039	1,04%
UK	2021-01-21	37977	2021-02-02	1451	0,038207336	3,82%
UK	2021-01-22	40321	2021-02-03	1322	0,032786885	3,28%
UK	2021-01-23	33652	2021-02-04	915	0,027190063	2,72%
UK	2021-01-24	30111	2021-02-05	1015	0,033708611	3,37%
UK	2021-01-25	22244	2021-02-06	828	0,037223521	3,72%
UK	2021-01-26	20167	2021-02-07	376	0,01864432	1,86%
UK	2021-01-27	25369	2021-02-08	333	0,013126256	1,31%
UK	2021-01-28	28811	2021-02-09	1052	0,036513832	3,65%

ΜΕΛΕΤΗ Α

Ερώτημα 1

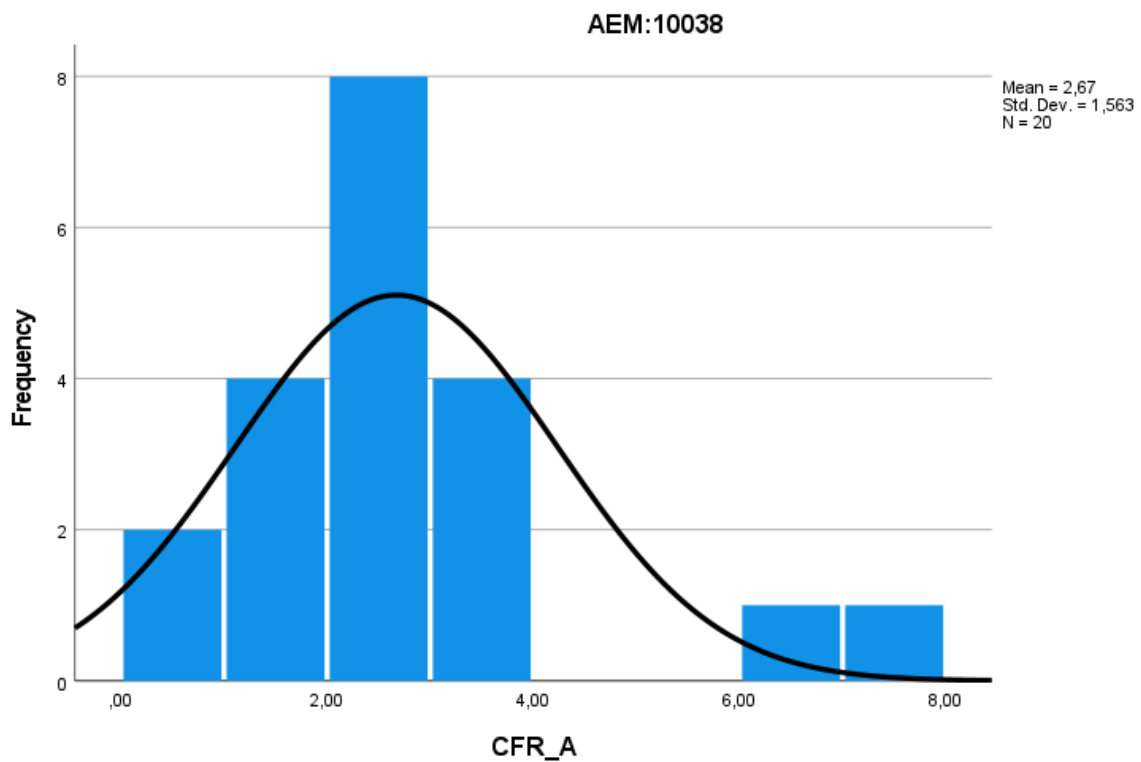
Στον παρακάτω πίνακα παρουσιάζονται τα μέτρα κεντρικής τάσης, δηλαδή μέση τιμή και διάμεσος και τα μέτρα μεταβλητότητας, δηλαδή διασπορά, τυπική απόκλιση, εύρος δεδομένων, πρώτο και τρίτο τεταρτημόριο.

AEM:10038			
		CFR_A	CFR_B
N	Valid	20	20
	Missing	0	0
Mean		2,6725	2,6305
Median		2,2750	2,9550
Std. Deviation		1,56254	,90808
Variance		2,442	,825
Range		6,31	2,78
Percentiles	25	1,8375	1,7850
	75	3,0850	3,3325

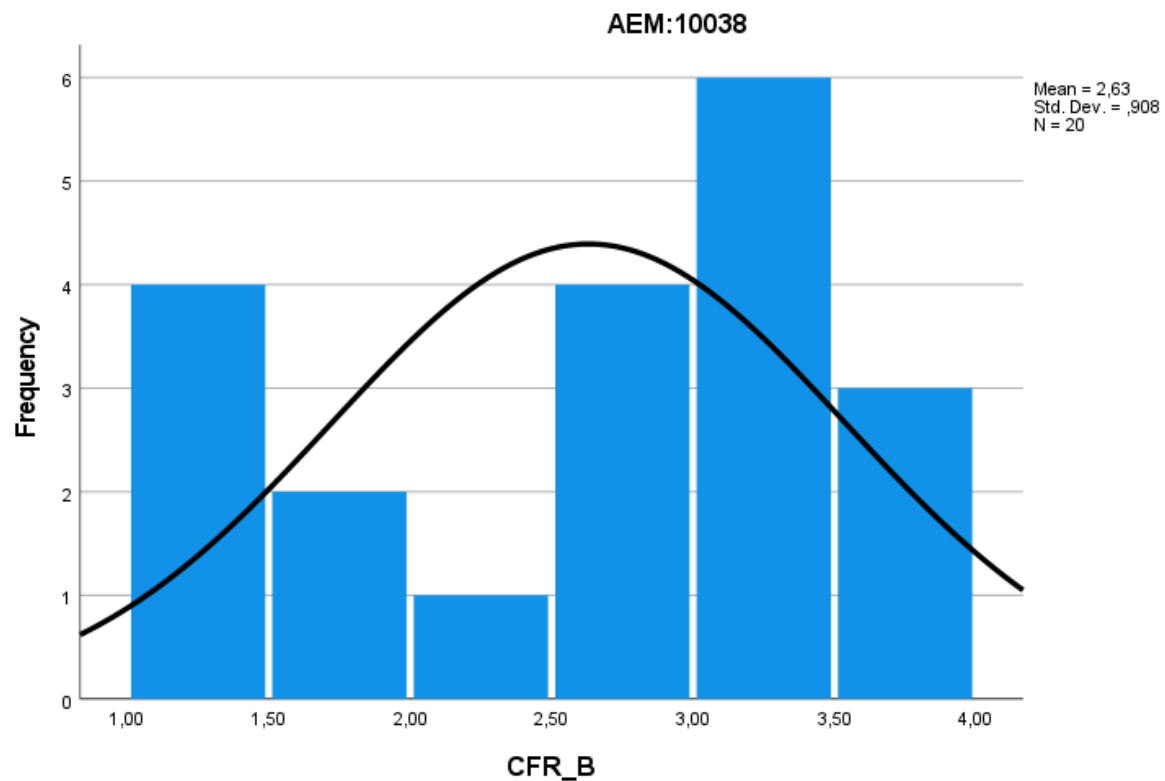
Αυτό που παρατηρούμε από τον πίνακα είναι ότι τα μέτρα μεταβλητότητας του Ηνωμένου Βασιλείου είναι αρκετά μικρότερα από αυτά του Βελγίου. Αυτό σημαίνει ότι η θνητότητα στο Ηνωμένο Βασίλειο μεταβάλλεται λιγότερο συγκριτικά με το Βέλγιο.

Τα μέτρα θέσης (μέση τιμή και διάμεσος) δε διαφέρουν πολύ στις δύο χώρες, οπότε η κεντρική τάση θνητότητας των δύο χωρών είναι περίπου η ίδια.

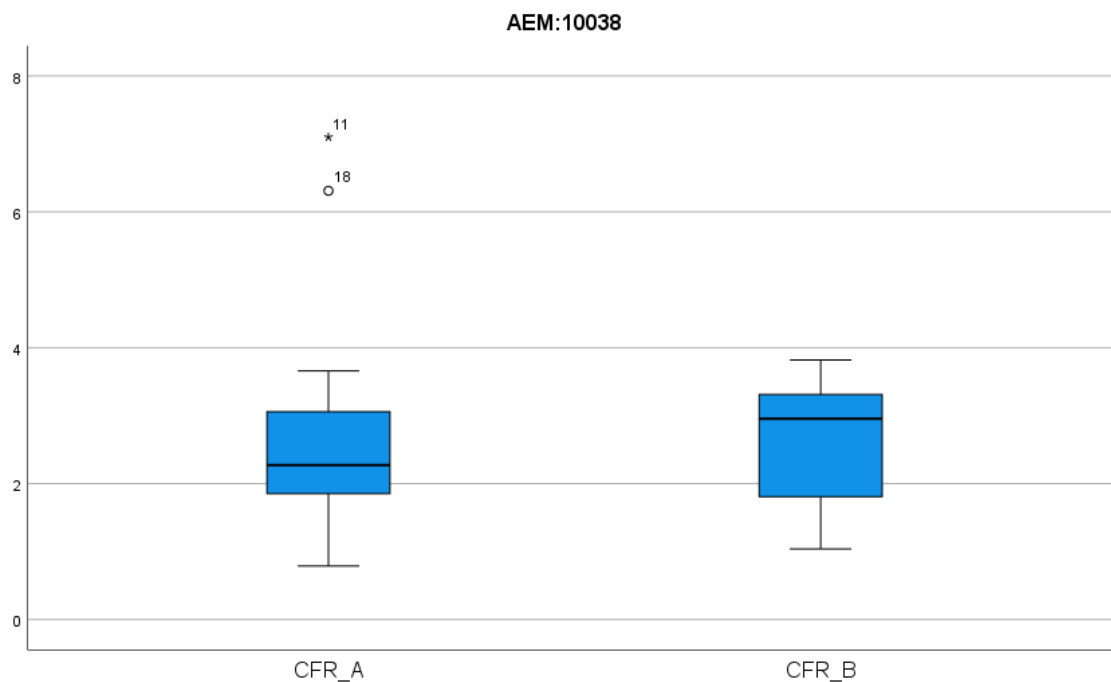
Ιστόγραμμα του Βελγίου: Παρατηρούμε ότι υπάρχει συμμετρία γύρω από την κεντρική τιμή, άρα μπορούμε να θεωρήσουμε ότι προσεγγίζεται από κανονική κατανομή



Ιστόγραμμα Ηνωμένου Βασιλείου: Δεν παρατηρείται κάποια συμμετρία γύρω από την κεντρική τιμή, συνεπώς δε μπορούμε να θεωρήσουμε ότι προσεγγίζεται από κανονική κατανομή.



Τα θηκογράμματα για τα 2 δείγματα:



Συμπέρασμα:

Για το Βέλγιο:

- 1) Η διάμεσος βρίσκεται σχετικά κοντά στο κέντρο του ενδοτεταρτομοριακού εύρους.
- 2) Τα μήκη των δύο μυστάκων είναι περίπου ίσα (δε διαφέρει το ένα σημαντικά από το άλλο).
- 3) Υπάρχουν ακραίες τιμές.

Προσεγγιστικά τόσο από το ιστόγραμμα όσο και από το θηκόγραμμα θα μπορούσαμε να θεωρήσουμε ότι ακολουθούν κανονική κατανομή, με επιφύλαξη, καθώς οι παρατηρήσεις είναι λίγες ($n = 20$).

Για το Ηνωμένο Βασίλειο:

- 1) Η διάμεσος βρίσκεται σχετικά κοντά στο Q3.
- 2) Τα μήκη των δύο μυστάκων είναι περίπου ίσα.
- 3) Δεν υπάρχουν ακραίες τιμές.

Παρατηρώντας τόσο το ιστόγραμμα όσο και το θηκόγραμμα του Ηνωμένου Βασιλείου, δε μπορούμε να θεωρήσουμε ότι ακολουθεί κανονική κατανομή.

Βέβαια, υπάρχει επιφύλαξη, καθώς ο αριθμός των παρατηρήσεων είναι μικρός.

Ερώτημα 2

One-Sample Statistics AEM:10038

	N	Mean	Std. Deviation	Std. Error Mean
CFR_A	20	2,6725	1,56254	,34940
CFR_B	20	2,6305	,90808	,20305

One-Sample Test AEM:10038

Test Value = 0

	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
CFR_A	7,649	19	<,001	2,67250	1,9412	3,4038
CFR_B	12,955	19	<,001	2,63050	2,2055	3,0555

Συμπέρασμα:

1)Για το Βέλγιο το 95% διάστημα εμπιστοσύνης είναι [1.9412 , 3.4038] ενώ για το Ηνωμένο Βασίλειο είναι [2.2055 , 3.0555].

2)Τα διαστήματα εμπιστοσύνης είναι σχετικά μικρά και για τα δύο δείγματα, άρα οι αντίστοιχες δειγματικές μέσες τιμές των δύο χωρών είναι αρκετά ακριβείς.

3)Όσον αφορά την ακρίβεια εκτίμησης της μέσης τιμής του ημερήσιου ποσοστού θανάτων, και για τις δύο χώρες το τυπικό σφάλμα είναι αρκετά μικρό (0.34940 και 0.20305 αντίστοιχα).

4)Η ζητούμενη μέση τιμή (5%) για τον ημερήσιο ποσοστό θνητότητας δεν περιέχεται σε κανένα από τα δύο διαστήματα εμπιστοσύνης άρα είναι αρκετά απίθανο να υπάρξει σε κάποια από τις δύο χώρες.

5)Υπάρχει επιφύλαξη λόγω του μικρού δείγματος (N=20).

Ερώτημα 3

Group Statistics AEM:10038

	COUNTRY	N	Mean	Std. Deviation	Std. Error Mean
CFR	1,00	20	2,6725	1,56254	,34940
	2,00	20	2,6305	,90808	,20305

Independent Samples Test AEM:10038

		Levene's Test for Equality of Variances		t-test for Equality of Means					95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
CFR	Equal variances assumed	,944	,337	,104	38	,918	,04200	,40411	-,77608	,86008
	Equal variances not assumed			,104	30,520	,918	,04200	,40411	-,78272	,86672

Συμπέρασμα:

1)Το 95% διάστημα εμπιστοσύνης της διαφοράς των μέσων τιμών υποθέτοντας ότι οι διασπορές είναι ίσες είναι το [-0.77608 , 0.86008], ενώ υποθέτοντας ότι οι διασπορές δεν είναι ίσες (ισχύει στη συγκεκριμένη περίπτωση) είναι το [-0.78272 , 0.86672].

2)Και στις δύο υποθέσεις το 0 περιέχεται στο διάστημα εμπιστοσύνης, συνεπώς μπορούμε να ισχυριστούμε ότι δεν υπάρχει σημαντική στατιστική διαφορά για το διάστημα εμπιστοσύνης 95%.

3)Υπάρχει επιφύλαξη λόγω του μικρού δείγματος (N = 20).

ΜΕΛΕΤΗ Β

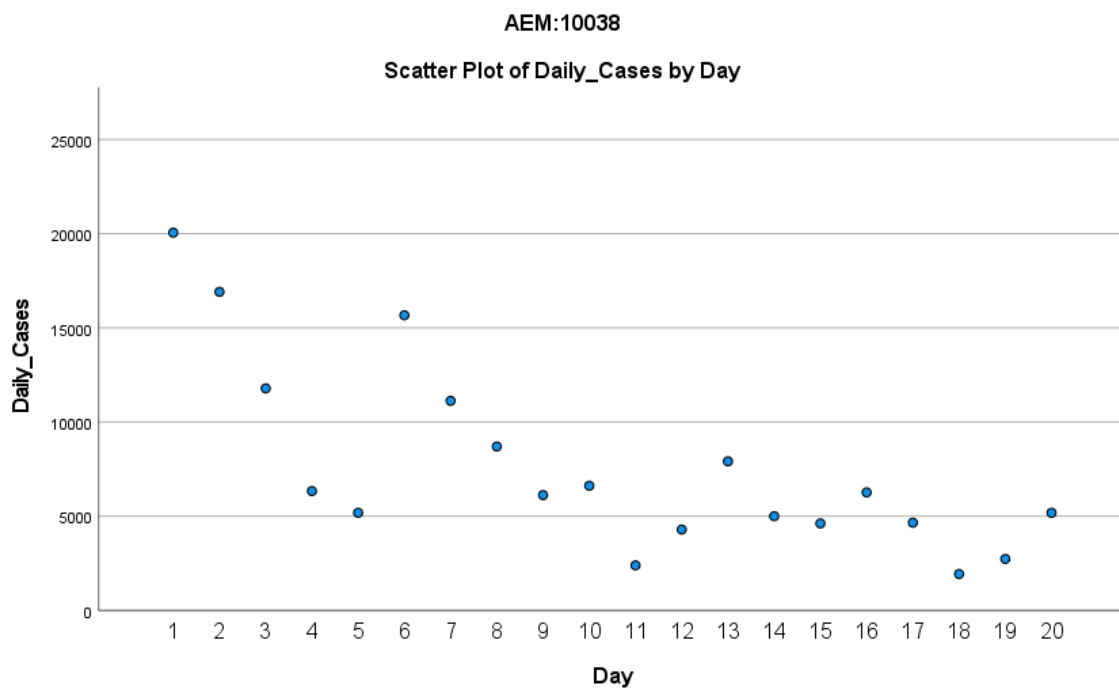
ΕΡΩΤΗΜΑ 4

Διάγραμμα διασποράς για τα ημερήσια κρούσματα του Βελγίου:

Correlations AEM:10038

		Day	Daily_Cases
Day	Pearson Correlation	1	-,745**
	Sig. (2-tailed)		<,001
	N	20	20
Daily_Cases	Pearson Correlation	-,745**	1
	Sig. (2-tailed)	<,001	
	N	20	20

** . Correlation is significant at the 0.01 level (2-tailed).



Συμπέρασμα:

- 1) Καθώς αυξάνει ο αριθμός των ημερών, μειώνεται ως επί το πλείστον ο αριθμός των νέων κρουσμάτων, άρα η εξάρτηση είναι αρνητική.
- 2) Υπάρχει, απλώς, μία μειωτική τάση χωρίς κανένα άλλο γνώρισμα, οπότε είναι γραμμική η εξάρτηση.

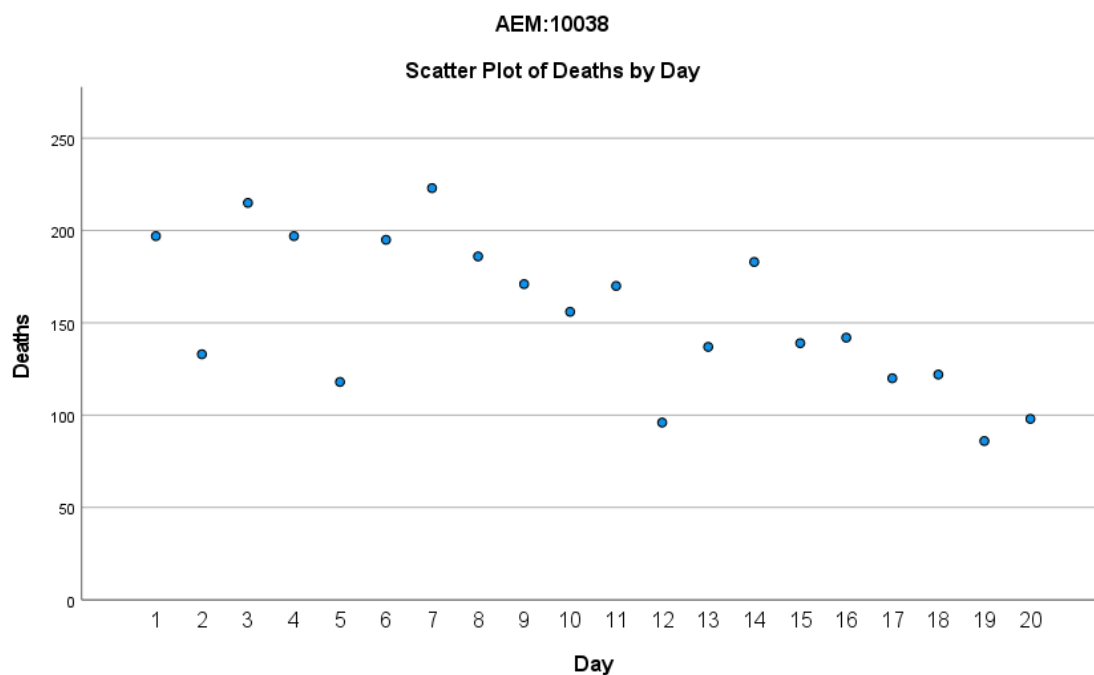
3) Η συσχέτιση είναι μέτρια. Αυτό επιβεβαιώνεται και από τον συντελεστή συσχέτισης Pearson ο οποίος είναι $r = -0,745$.

Διάγραμμα διασποράς για τον ημερήσιο αριθμό θανάτων:

Correlations AEM:10038

		Day	Deaths
Day	Pearson Correlation	1	-,656**
	Sig. (2-tailed)		,002
	N	20	20
Deaths	Pearson Correlation	-,656**	1
	Sig. (2-tailed)	,002	
	N	20	20

** . Correlation is significant at the 0.01 level (2-tailed).



Συμπέρασμα:

1) Καθώς αυξάνει ο αριθμός των ημερών, μειώνεται ως επί το πλείστον ο αριθμός των θανάτων, άρα η εξάρτηση είναι αρνητική.

2) Υπάρχει, απλώς, μία μειωτική τάση χωρίς κανένα άλλο γνώρισμα, οπότε είναι γραμμική.

3) Η συσχέτιση είναι μέτρια. Αυτό επιβεβαιώνεται και από τον συντελεστή συσχέτισης Pearson ο οποίος είναι $r = -0,656$.

ΕΡΩΤΗΜΑ 5

Μοντέλο Γραμμικής Παλινδρόμησης για τα ημερήσια κρούσματα του Βελγίου:

Model Summary AEM:10038

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,745 ^a	,556	,531	3423,182

a. Predictors: (Constant), Day

AEM:10038 Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	14288,016	1590,176		8,985	<,001
	Day	-629,673	132,745	-,745	-4,743	<,001

a. Dependent Variable: Daily_Cases

Έστω ευθεία $y = b + ax$ όπου y ημερήσια νέα κρούσματα, x αύξων αριθμός ημέρας, b κλίση της ευθείας και a σταθερός όρος της ευθείας.

Από τον δεύτερο πίνακα προκύπτει ότι $b=14288.016$ και $a=-629.673$.

Η τυπική απόκλιση σφαλμάτων παλινδρόμησης είναι 3423,182 και το R-square ισούται με 55,6%, συνεπώς το μοντέλο δεν είναι αρκετά κατάλληλο για προβλέψεις.

Μοντέλο Γραμμικής Παλινδρόμησης για τους ημερήσιους θανάτους του Βελγίου:

Model Summary AEM:10038

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,656 ^a	,431	,399	31,801

a. Predictors: (Constant), Day

AEM:10038 Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	201,979	14,773		13,672	<,001
	Day	-4,550	1,233	-,656	-3,690	,002

a. Dependent Variable: Deaths

Έστω ευθεία $y = b + ax$ όπου y ημερήσιοι νέοι θάνατοι, x αύξων αριθμός ημέρας, b κλίση της ευθείας και a σταθερός όρος της ευθείας.

Από τον δεύτερο πίνακα προκύπτει ότι $b=201,979$ και $a=-4,55$.

Η τυπική απόκλιση σφαλμάτων παλινδρόμησης είναι 31,801 και το R-square ισούται με 43,1%, συνεπώς το μοντέλο δεν είναι αρκετά κατάλληλο για προβλέψεις.

Γενικά, η κλίση που αφορά τα κρούσματα είναι πολύ μεγαλύτερη από την κλίση του μοντέλου που αφορά τους θανάτους. Οπότε, ο αριθμός των κρουσμάτων μειώνεται πολύ πιο γρήγορα σε σχέση με τον αριθμό των θανάτων.

Επίσης και τα 2 μοντέλα κρίθηκαν ακατάλληλα.

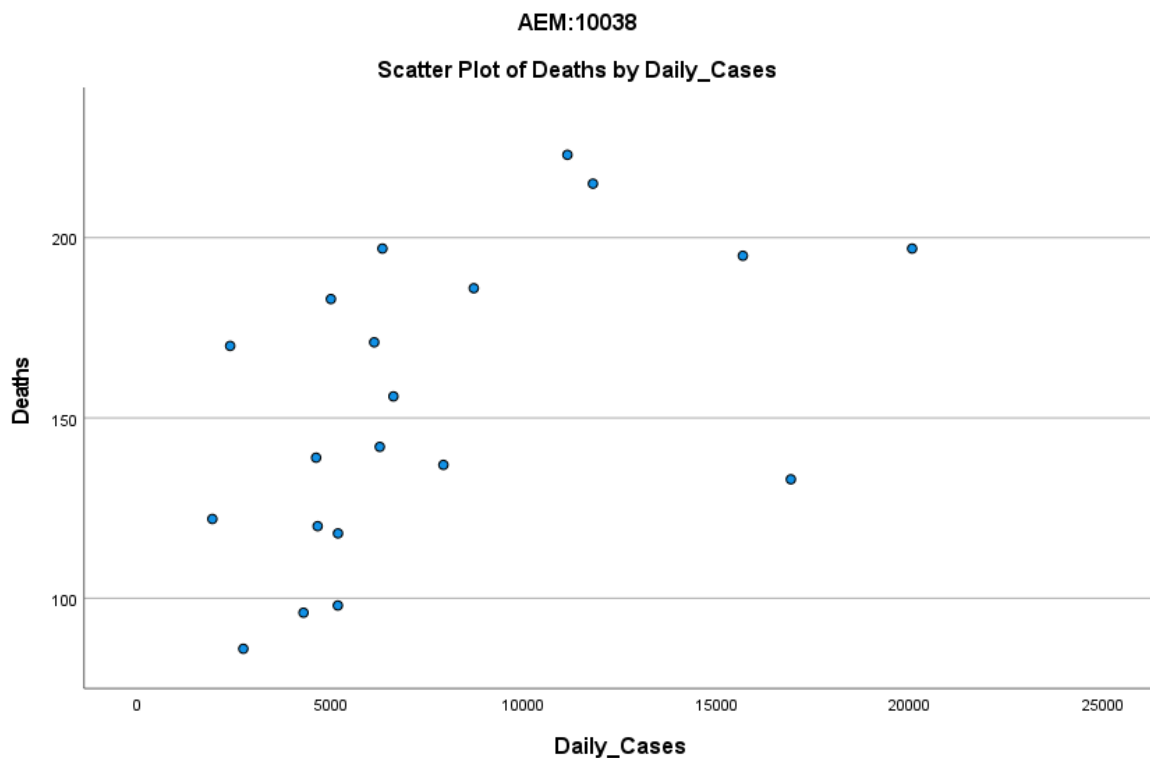
ΕΡΩΤΗΜΑ 6

Η αρχική υστέρηση είναι $\tau=12$. Το αντίστοιχο διάγραμμα διασποράς είναι:

Correlations AEM:10038

		Daily_Cases	Deaths
Daily_Cases	Pearson Correlation	1	,525 [*]
	Sig. (2-tailed)		,017
	N	20	20
Deaths	Pearson Correlation	,525 [*]	1
	Sig. (2-tailed)	,017	
	N	20	20

*. Correlation is significant at the 0.05 level (2-tailed).



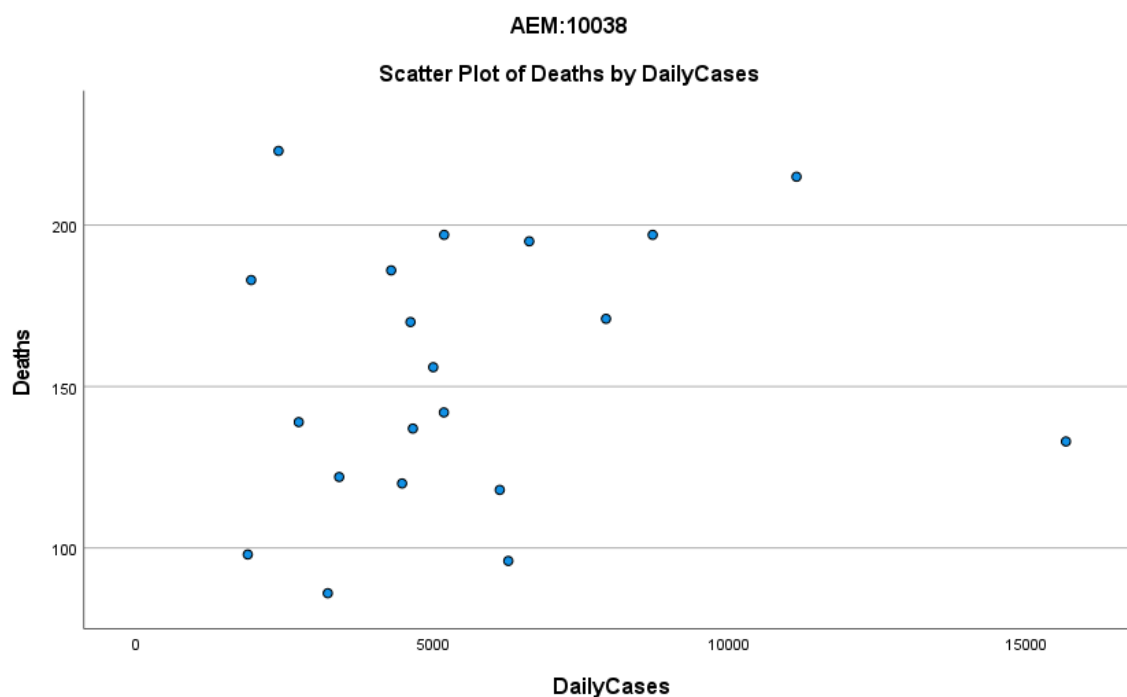
Συμπέρασμα:

- 1) Καθώς αυξάνει ο αριθμός των κρουσμάτων, αυξάνεται ως επί το πλείστον και ο αριθμός των θανάτων, άρα η εξάρτηση είναι θετική.
- 2) Υπάρχει, απλώς, μία αυξητική τάση χωρίς κανένα άλλο γνώρισμα, οπότε θεωρούμε ότι είναι γραμμική η εξάρτηση.
- 3) Η συσχέτιση είναι μέτρια. Αυτό επιβεβαιώνεται και από τον συντελεστή συσχέτισης Pearson ο οποίος είναι $r = 0,525$.

Για υστέρηση 8 ημερών($\tau=4$):

Correlations AEM:10038

		DailyCases	Deaths
DailyCases	Pearson Correlation	1	,162
	Sig. (2-tailed)		,494
	N	20	20
Deaths	Pearson Correlation	,162	1
	Sig. (2-tailed)	,494	
	N	20	20



Συμπέρασμα:

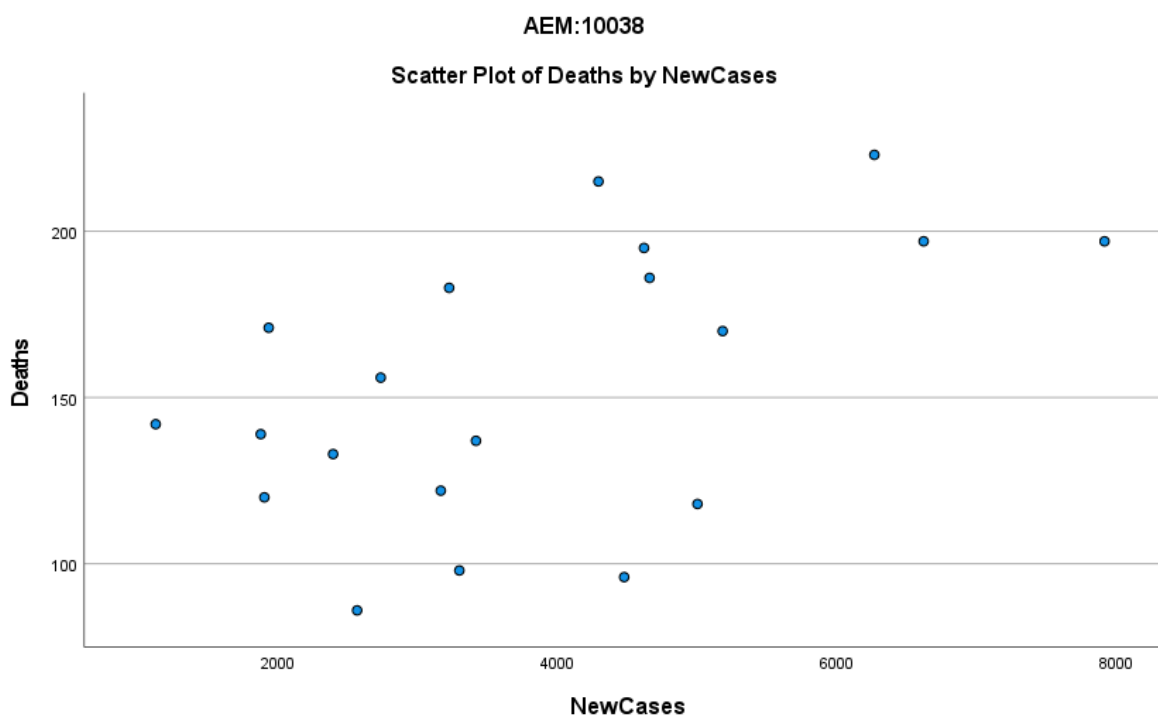
- 1)Καθώς αυξάνει ο αριθμός των κρουσμάτων, αυξάνεται ως επί το πλείστον και ο αριθμός των θανάτων, άρα η εξάρτηση είναι θετική.
- 2)Υπάρχει, απλώς, μία αυξητική τάση χωρίς κανένα άλλο γνώρισμα, οπότε θεωρούμε ότι είναι γραμμική η εξάρτηση.
- 3)Η συσχέτιση είναι ασθενής. Αυτό επιβεβαιώνεται και από τον συντελεστή συσχέτισης Pearson ο οποίος είναι $r = 0,162$.

Για υστέρηση 3 ημερών($\tau=9$):

Correlations AEM:10038

		Deaths	NewCases
Deaths	Pearson Correlation	1	,535 [*]
	Sig. (2-tailed)		,015
	N	20	20
NewCases	Pearson Correlation	,535 [*]	1
	Sig. (2-tailed)	,015	
	N	20	20

*. Correlation is significant at the 0.05 level (2-tailed).



Συμπέρασμα:

- 1) Καθώς αυξάνει ο αριθμός των κρουσμάτων, αυξάνεται ως επί το πλείστον και ο αριθμός των θανάτων, άρα η εξάρτηση είναι θετική.
- 2) Υπάρχει, απλώς, μία αυξητική τάση χωρίς κανένα άλλο γνώρισμα, οπότε θεωρούμε ότι είναι γραμμική η εξάρτηση.
- 3) Η συσχέτιση είναι μέτρια. Αυτό επιβεβαιώνεται και από τον συντελεστή συσχέτισης Pearson ο οποίος είναι $r = 0,535$.

Και για τις τρεις δοκιμές η συσχέτιση του αριθμού ημερήσιων νέων θανάτων και των ημερήσιων νέων κρουσμάτων δεν είναι ισχυρή. Ισχυρότερη συσχέτιση όμως παρουσιάζεται όταν η υστέρηση είναι $\tau = 3$ αφού ο συντελεστής Pearson ισούται με $r = 0.535$.

ΕΡΩΤΗΜΑ 7

Για την αρχική υστέρηση $\tau=12$:

Model Summary AEM:10038

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,525 ^a	,276	,236	35,861

a. Predictors: (Constant), Daily_Cases

AEM:10038 Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	121,097	14,966		8,092	<,001
	Daily_Cases	,004	,002	,525	2,620	,017

a. Dependent Variable: Deaths

Έστω ευθεία $y = b + ax$ όπου y ημερήσιοι νέοι θάνατοι, x ημερήσια νέα κρούσματα, b κλίση της ευθείας και a σταθερός όρος της ευθείας.

Από τον δεύτερο πίνακα προκύπτει ότι $b=121,097$ και $a=0,004$.

Η τυπική απόκλιση σφαλμάτων παλινδρόμησης είναι 35,861 και το R-square ισούται με 27,6%, συνεπώς το μοντέλο δεν είναι αρκετά κατάλληλο για προβλέψεις.

Για υστέρηση 8 ημερών(τ-4):

Model Summary AEM:10038

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,162 ^a	,026	-,028	41,588

a. Predictors: (Constant), DailyCases

AEM:10038 Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	143,060	18,479		7,742	<,001
	DailyCases	,002	,003	,162	,698	,494

a. Dependent Variable: Deaths

Έστω ευθεία $y = b + ax$ όπου y ημερήσιοι νέοι θάνατοι, x ημερήσια νέα κρούσματα, b κλίση της ευθείας και a σταθερός όρος της ευθείας.

Από τον δεύτερο πίνακα προκύπτει ότι $b=143.06$ και $a=0.002$.

Η τυπική απόκλιση σφαλμάτων παλινδρόμησης είναι 41.588 και το R-square ισούται με 2.6%, συνεπώς το μοντέλο δεν είναι αρκετά κατάλληλο για προβλέψεις.

Για υστέρηση 3 ημερών(τ-9):

Model Summary AEM:10038

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,535 ^a	,287	,247	35,596

a. Predictors: (Constant), NewCases

AEM:10038 Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	106,894	19,305		5,537	<,001
	NewCases	,012	,005	,535	2,690	,015

a. Dependent Variable: Deaths

Έστω ευθεία $y = b + ax$ όπου y ημερήσιοι νέοι θάνατοι, x ημερήσια νέα κρούσματα, b κλίση της ευθείας και a σταθερός όρος της ευθείας.

Από τον δεύτερο πίνακα προκύπτει ότι $b=106.894$ και $a=0.012$.

Η τυπική απόκλιση σφαλμάτων παλινδρόμησης είναι 35.596 και το R-square ισούται με 28.7%, συνεπώς το μοντέλο δεν είναι αρκετά κατάλληλο για προβλέψεις.

Γενικά παρατηρείται ότι οι εξισώσεις ευθειών των μοντέλων γραμμικής παλινδρόμησης δε διαφέρουν σημαντικά μεταξύ των διαφορετικών τιμών υστέρησης (οι κλίσεις και οι σταθεροί όροι των ευθειών είναι παρόμοιοι). Επίσης και η τυπική απόκλιση σφαλμάτων παλινδρόμησης είναι παρόμοια και για τις τρεις δοκιμές. Η μεγαλύτερη διαφορά φαίνεται στο R-square. Όλα τα μοντέλα κρίνονται ακατάλληλα για προβλέψεις, αν έπρεπε να επιλέξουμε όμως κάποιο από τα τρία, τότε θα επιλέγαμε το τρίτο καθώς έχει τη μεγαλύτερη τιμή R-square (28,7%).