

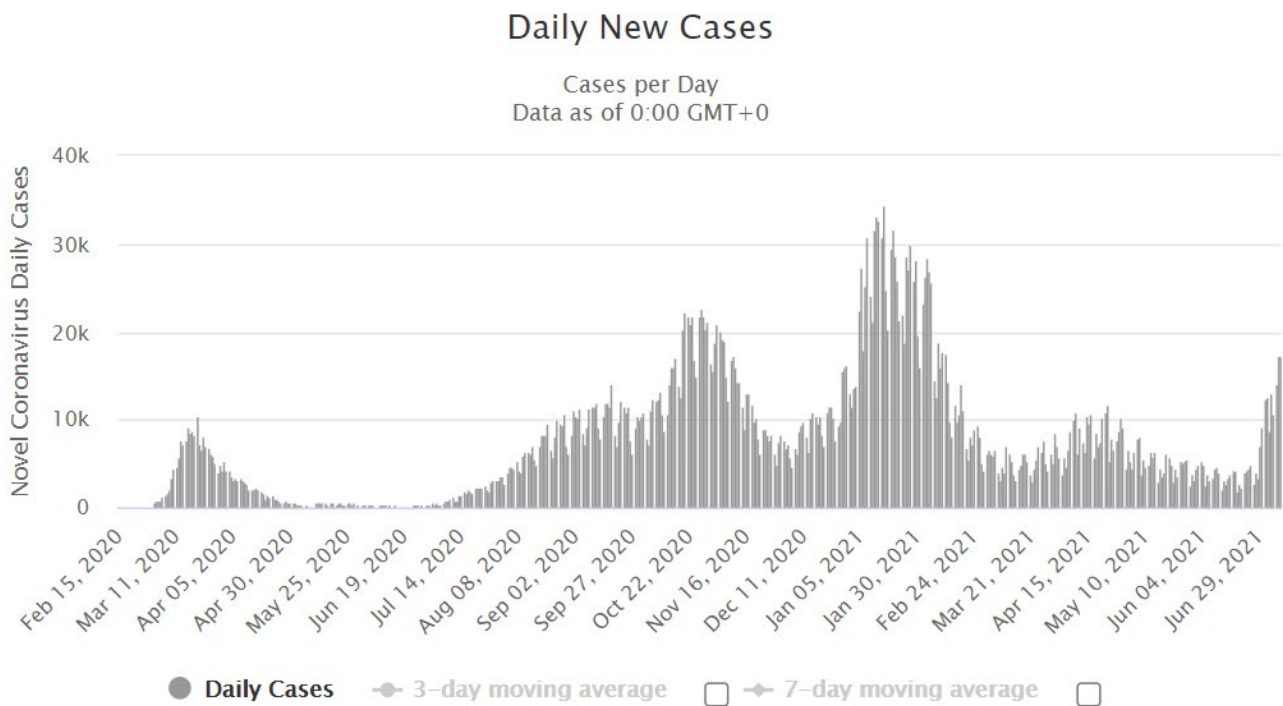
# ΕΡΓΑΣΙΑ ΣΤΗ ΣΤΑΤΙΣΤΙΚΗ 2021

Ονοματεπώνυμο: Σιταρίδης Παναγιώτης

ΑΕΜ: 10249

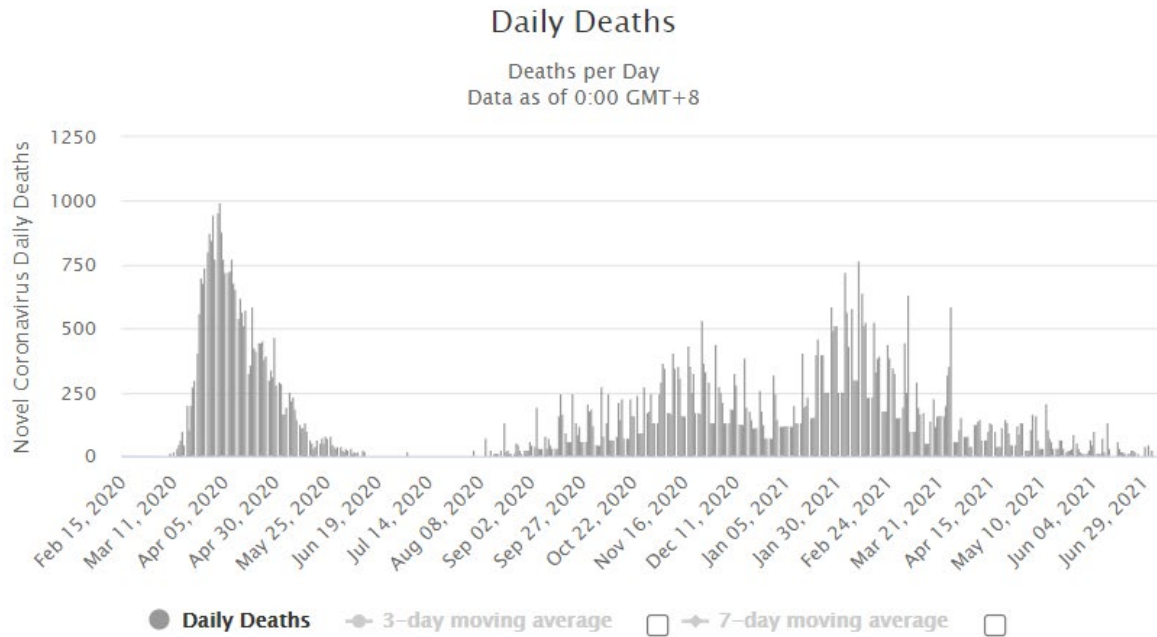
Χώρα: Ισπανία

## Daily New Cases in Spain



Από το διάγραμμα για τα ημερήσια νέα κρούσματα, όσον αφορά το 3<sup>ο</sup> κύμα, δηλαδή από τις 07/12/2020 έως τις 14/03/2021 βλέπουμε πως στις 15/01/2021 έχουμε τον μεγαλύτερο αριθμό κρουσμάτων (34.378), επομένως θα ορίσουμε την κορύφωση της καμπύλης με αριθμό κρουσμάτων 34.378.

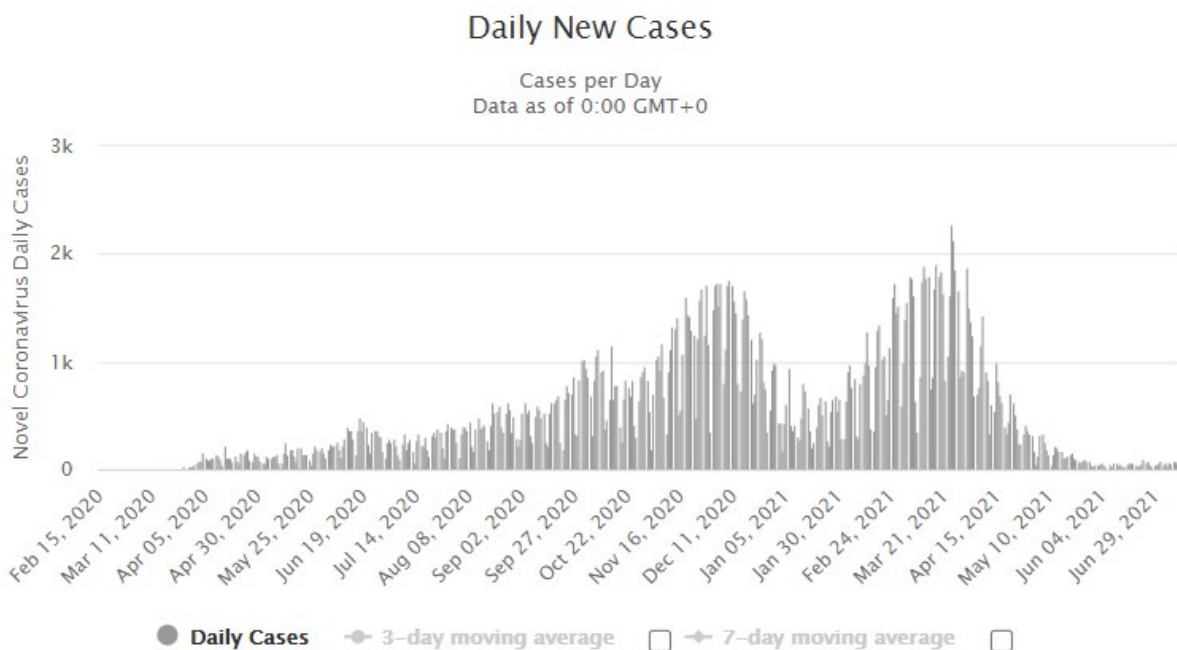
## Daily New Deaths in Spain



Από το διάγραμμα για τους ημερήσιους νέους θανάτους για το 3<sup>ο</sup> κύμα, δηλαδή από τις 07/12/2020 έως τις 14/03/2021 βλέπουμε πως στις 09/02/2021 έχουμε την μέγιστη καταγραφή ημερήσιων θανάτων(766) οπότε θα ορίσουμε την κορύφωση της καμπύλης αυτήν την μέρα με 766 θανάτους.

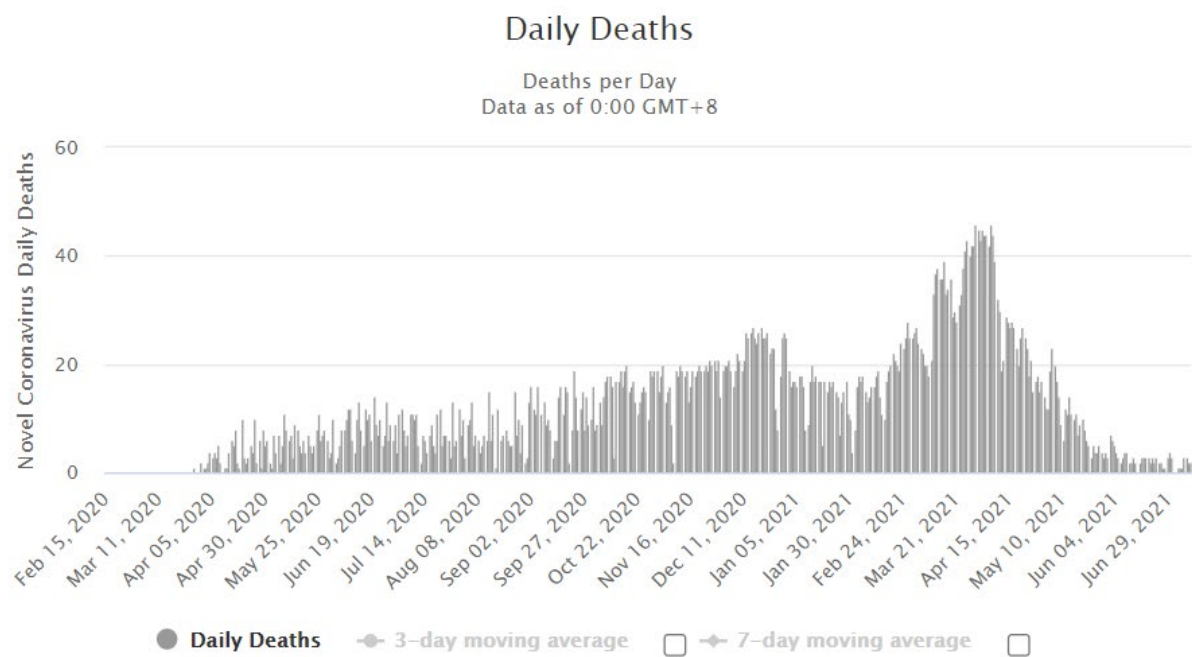
Χώρα: Μολδαβία

## Daily New Cases in Moldova



Από το διάγραμμα για τα ημερήσια νέα κρούσματα, όσον αφορά το κύμα μεταξύ 18/10/2020 και 18/01/2021, έχουμε τον μεγαλύτερο αριθμό κρουσμάτων (1.766) στις 09/12/2020, επομένως θα ορίσουμε την κορύφωση της καμπύλης με αριθμό κρουσμάτων 1.766.

## Daily New Deaths in Moldova



Από το διάγραμμα για τους ημερήσιους νέους θανάτους για το κύμα μεταξύ 18/10/2020 και 18/01/2021, βλέπουμε πως στις 09/02/2021 έχουμε την μέγιστη καταγραφή ημερήσιων θανάτων(27) σε δυο ημερομηνίες, στις 15/12/2020 και στις 19/12/2020 οπότε θα ορίσουμε την κορύφωση της καμπύλης σε αυτές τις μέρες με 27 θανάτους.

## Μελέτη Α

Πίνακας για την Ισπανία:

Ημερήσια Κρούσματα	Ημερήσιοι Θάνατοι	Δείγμα Α
24.894	643	2,6
20.348	513	2,5
29.612	530	1,8
31.688	234	0,7
28.602	234	0,8
25.840	234	0,9
21.442	530	2,5

22.027	337	1,5
18.841	388	2
28.764	397	1,4
27.069	178	0,7
29.906	178	0,6
25.920	179	0,7
28.292	443	1,6
19.665	389	2
16.072	345	2,1
23.381	329	1,4
26.235	156	0,6
28.498	156	0,5
27.021	155	0,6

Πίνακας για την Μολδαβία:

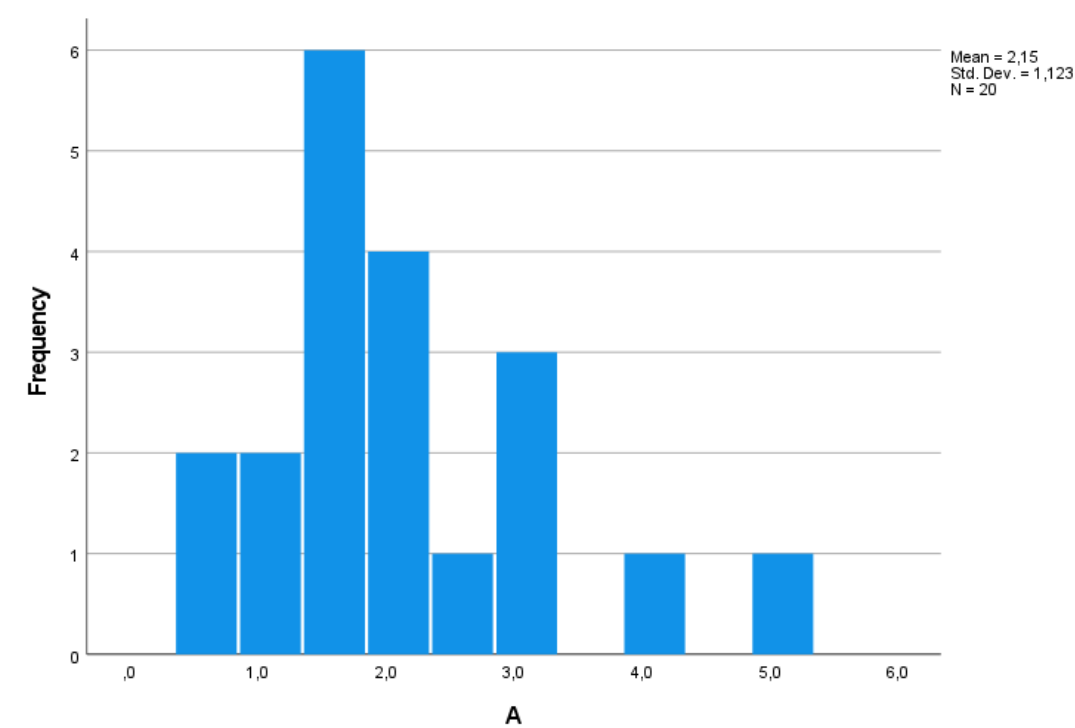
Ημερήσια Κρούσματα	Ημερήσιοι Θάνατοι	Δείγμα Β
1.715	25	1,5
1.579	25	1,6
1.459	26	1,8
795	22	2,8
735	23	3,1
1.403	23	1,6
1.673	12	0,7
1.390	8	0,6
1.441	18	1,2
1.213	25	2,1
629	26	4,1
703	23	3,3
1.021	19	1,9
1.282	16	1,2
1.222	17	1,4
815	17	2,1
746	16	2,1
359	18	5
569	18	3,2
927	16	1,7

Για να δημιουργηθούν τα δύο δείγματα κοιτάξαμε τα διαγράμματα για τα ημερήσια νέα κρούσματα και τους ημερήσιους θανάτους για την κάθε χώρα. Επιλέξαμε ως ημέρα κορύφωσης την ημέρα με τα περισσότερα καταγεγραμμένα κρούσματα και τους περισσότερους θανάτους. Πήραμε τις τιμές από τις 20 συνεχόμενες μέρες που ακολουθούσαν και για τα κρούσματα και για τους θανάτους και διαιρώντας αυτά τα νούμερα (ημερήσιοι θάνατοι προς ημερήσια κρούσματα) και πολλαπλασιάζοντας με το 100 προέκυψαν τα δείγματα Α και Β για την κάθε χώρα. Αυτά φαίνονται και στα διαγράμματα πιο πάνω και το πώς προέκυψαν τα δύο δείγματα φαίνεται στους δύο προηγούμενους πίνακες . Για την χώρα Α (Ισπανία), τόσο για τα ημερήσια κρούσματα όσο και για τους ημερήσιους θανάτους, δεν υπήρχε κάποιο θέμα. Στην χώρα Β(Μολδαβία) για τα ημερήσια κρούσματα δεν υπήρξε κάποιο θέμα, ωστόσο υπήρχαν δύο ίσες μέγιστες τιμές για τους ημερήσιους θανάτους στο κύμα που επιλέξαμε, έτσι η λήψη του δείγματος ελήφθη για 20 συνεχόμενες μέρες μετά την δεύτερη φορά που εμφανίζεται η μέγιστη τιμή.

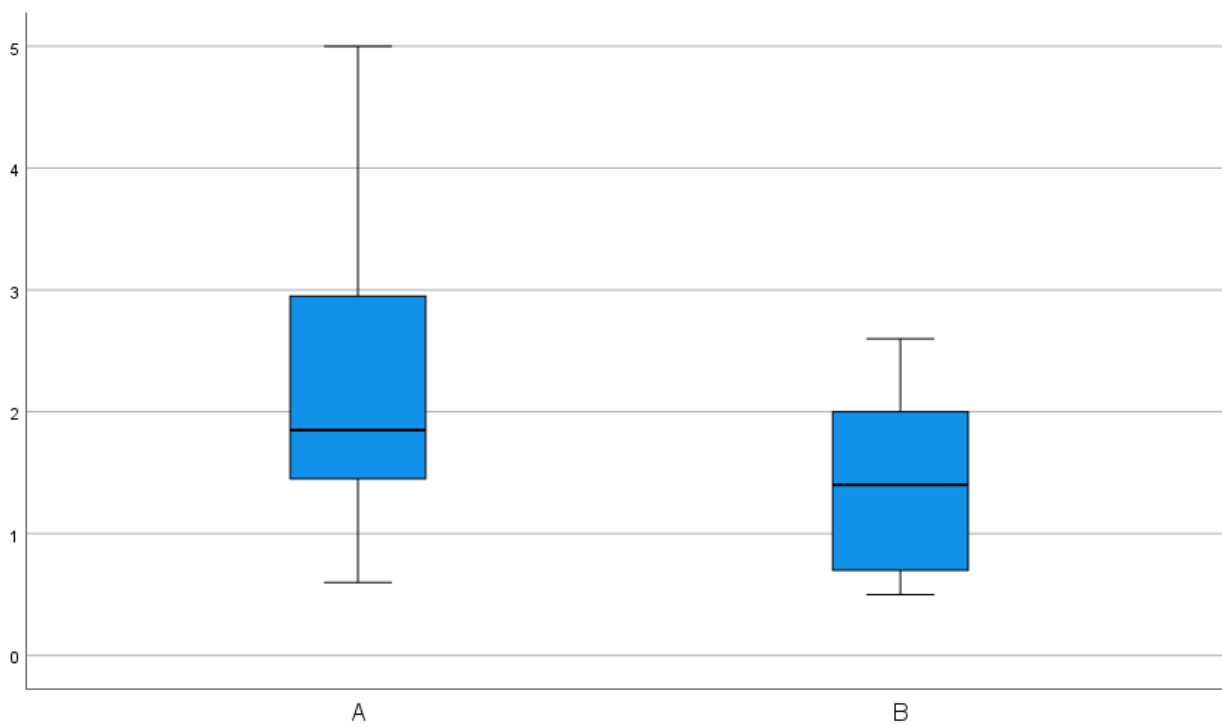
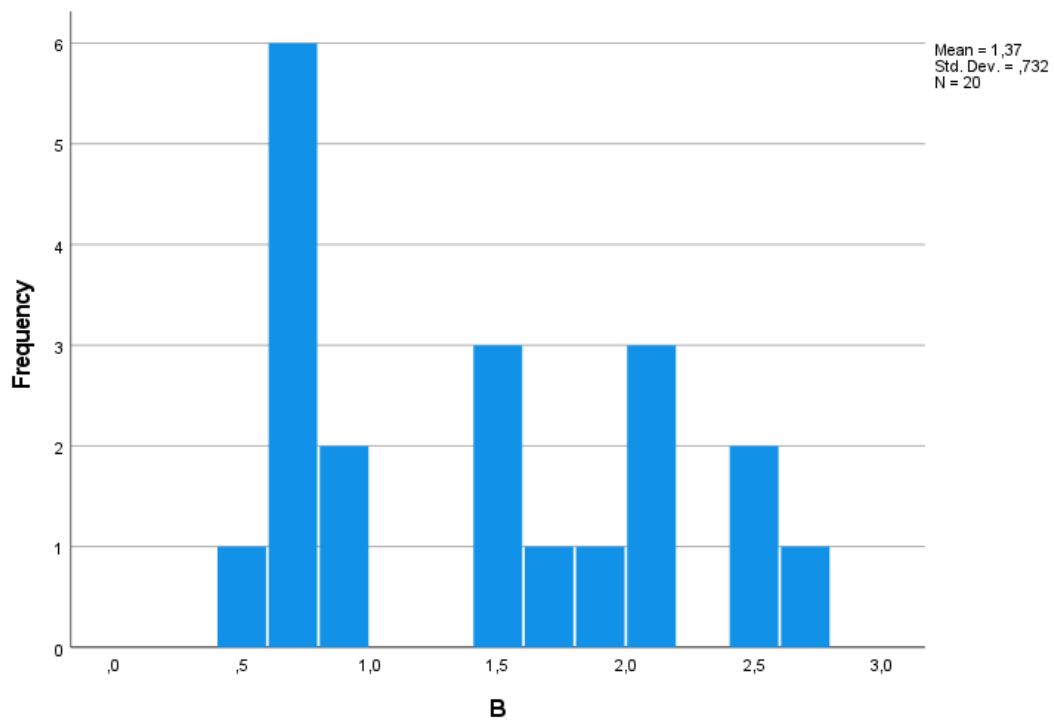
ΕΡΩΤΗΜΑ 1.

Statistics			
		A	B
N	Valid	20	20
	Missing	40	40
Mean		2,150	1,370
Median		1,850	1,400
Std. Deviation		1,1232	,7320
Variance		1,262	,536
Range		4,4	2,1
Percentiles	25	1,425	,700
	50	1,850	1,400
	75	3,025	2,000

Ιστόγραμμα Δείγματος Α.



ΙΣΤΟΓΡΑΜΜΑ ΔΕΙΓΜΑΤΟΣ Β.



#####

## EPQTHMA 2

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
A	20	2,150	1,1232	,2512
B	20	1,370	,7320	,1637

### One-Sample Test

Test Value = 0

	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
A	8,560	19	,000	2,1500	1,624	2,676
B	8,369	19	,000	1,3700	1,027	1,713

Το διάστημα εμπιστοσύνης 95% για το Α παρατηρούμε ότι είναι [1,624, 2,676] και για το Β [1,027,1,713] .Και από την μέση τιμή του Α(2,15) και του Β(1,37) βλέπουμε πως η μέση τιμή του ημερήσιου ποσοστού θνητότητας δεν μπορεί να είναι 5% διότι απέχει αρκετά από το διάστημα εμπιστοσύνης. Η επιφύλαξη που έχουμε για την εγκυρότητα των παραπάνω αποτελεσμάτων είναι σχετικά μικρή, καθώς επιλέξαμε ένα αρκετά μεγάλο διάστημα εμπιστοσύνης. Ωστόσο δεν μπορούμε να πούμε πως είμαστε και απολύτως σίγουροι για το αποτελέσματά μας, μιας και το δείγμα των 20 παρατηρήσεων είναι μικρό.

### ΕΡΩΤΗΜΑ 3.

#### Group Statistics

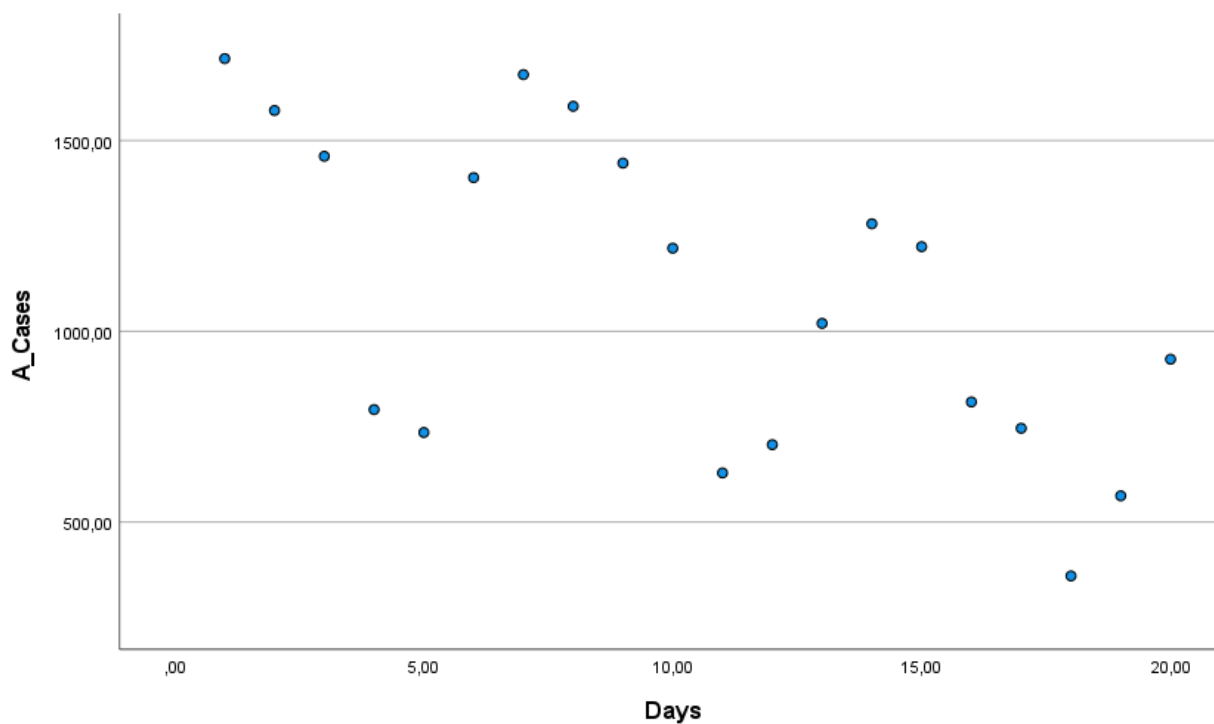
	Country	N	Mean	Std. Deviation	Std. Error Mean
A_B	A	20	2,1500	1,12320	,25116
	B	20	1,3700	,73205	,16369

#### Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means					
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference Lower Upper
A_B	Equal variances assumed	1,726	,197	2,602	38	,013	,78000	,29979	,17311 1,38689
	Equal variances not assumed			2,602	32,674	,014	,78000	,29979	,16984 1,39016

Από τον πρώτο πίνακα βλέπουμε πως οι τυπικές αποκλίσεις έχουν αρκετή διαφορά, οπότε κοιτάμε τη 2η γραμμή του δεύτερου πίνακα(και τη πρώτη γραμμή να παίρναμε θα φτάναμε στα ίδια συμπεράσματα). Βλέπουμε πως το εύρος εμπιστοσύνης που βρίσκουμε είναι μόνο θετικές τιμές, δηλαδή δεν περιέχουν το 0 (η διαφορά τους είναι στατιστικά σημαντική). Άρα για το δείγμα μας η μέση τιμή του ημερήσιου ποσοστού θνητότητας δεν μπορεί να είναι ίδια στις δύο περιοχές, στο διάστημα εμπιστοσύνης μας.

### ΕΡΩΤΗΜΑ 4.



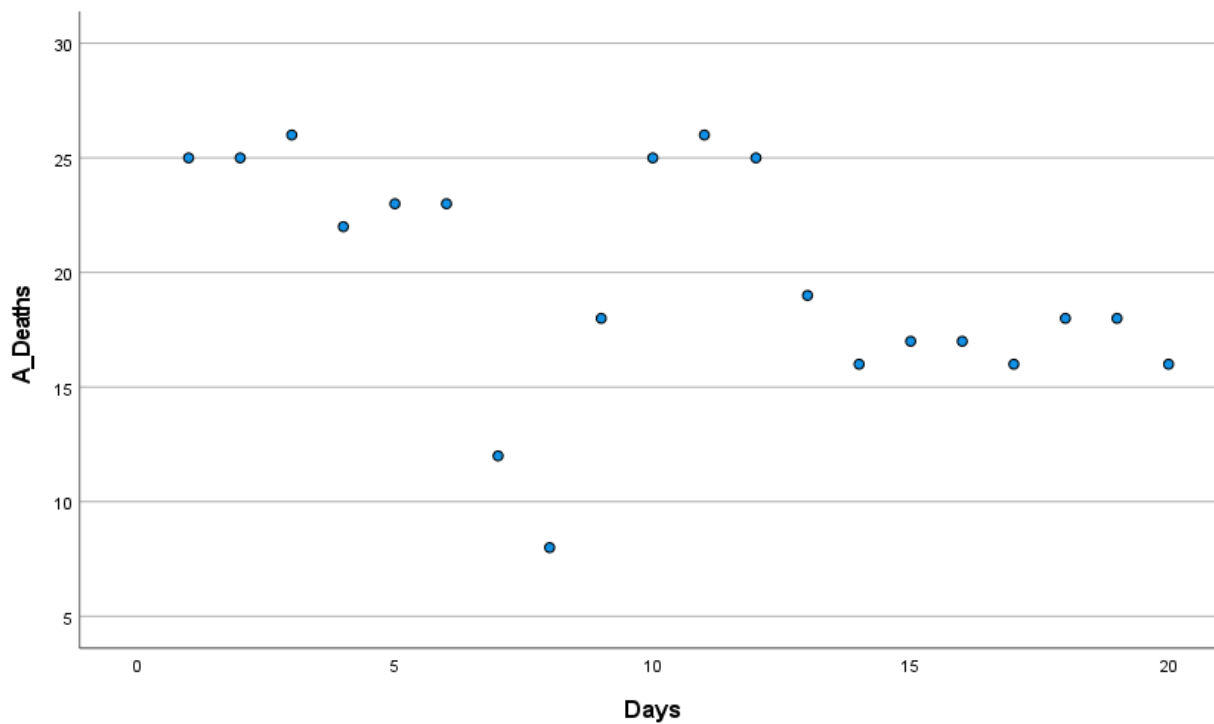
### Correlations

		A_Cases	Days
A_Cases	Pearson Correlation	1	-,628**
	Sig. (2-tailed)		,003
	N	20	20
Days	Pearson Correlation	-,628**	1
	Sig. (2-tailed)	,003	
	N	20	20

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

Από το διάγραμμα της διασποράς για τη χώρα Α φαίνεται πως υπάρχει μια εξάρτηση ανάμεσα στις μέρες και τα κρούσματα. Αν και υπάρχουν διακυμάνσεις σε μερικές μέρες όσο αυξάνονται οι μέρες φαίνεται να μειώνονται τα κρούσματα. Η εξάρτηση είναι γραμμική και αρνητική όμως δεν είναι ισχυρή γεγονός που το αποδεικνύει και το πινακάκι της συσχέτισης όπου βλέπουμε πως ο συντελεστής συσχέτισης είναι στο -0,628





### Correlations

		A_Deaths	Days
A_Deaths	Pearson Correlation	1	-,445*
	Sig. (2-tailed)		,049
	N	20	20
Days	Pearson Correlation	-,445*	1
	Sig. (2-tailed)	,049	
	N	20	20

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

Από το 2<sup>ο</sup> διάγραμμα για την χώρα Α φαίνεται πως υπάρχει μια εξάρτηση ανάμεσα στις μέρες και τους θανάτους. Υπάρχουν και πάλι διακυμάνσεις σε μερικές μέρες ωστόσο, όσο αυξάνονται οι μέρες φαίνεται να μειώνονται οι θάνατοι. Η εξάρτηση είναι γραμμική και αρνητική όμως δεν είναι ισχυρή όπως φαίνεται και στον πίνακα της συσχέτισης όπου βλέπουμε πως ο συντελεστής συσχέτισης είναι στο -0,445.

## ΕΡΩΤΗΣΗ 5.

**Model Summary**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,628 <sup>a</sup>	,394	,361	330,908

a. Predictors: (Constant), Days

**Coefficients<sup>a</sup>**

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1555,205	153,717		10,117	,000
	Days	-43,920	12,832	-,628	-3,423	,003

a. Dependent Variable: A\_Cases

Παρατηρούμε ότι το R Square είναι 0,394 που σημαίνει ότι το μοντέλο είναι 39% κατάλληλο για προβλέψεις το οποίο είναι αρκετά χαμηλό

Από την μέθοδο ελαχίστων τετραγώνων και από τα πινακάκια το γραμμικό μοντέλο είναι :

**Κρούσματα = 1555,205 - 43,920\* ημέρες**

**Model Summary**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,445 <sup>a</sup>	,198	,154	4,623

a. Predictors: (Constant), Days

**Coefficients<sup>a</sup>**

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	23,721	2,147		11,047	,000
	Days	-,378	,179	-,445	-2,110	,049

a. Dependent Variable: A\_Deaths

Βλέπουμε πως το R Square είναι 0,198 που σημαίνει ότι το μοντέλο είναι περίπου 20% κατάλληλο , ποσοστό επίσης αρκετά χαμηλό

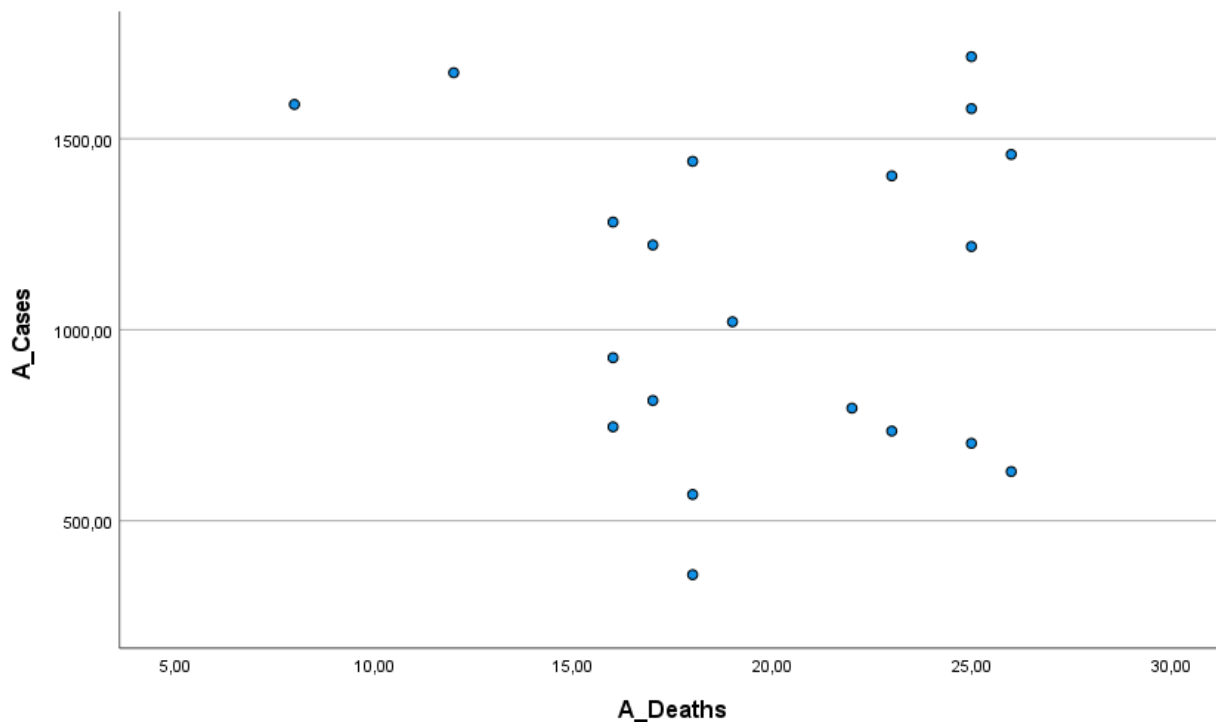
Από την μέθοδο ελαχίστων τετραγώνων και από τα πινακάκια το γραμμικό μοντέλο είναι :

**Θάνατοι = 23,721 -0,378 \* ημέρες**

Από τους δύο πίνακες παρατηρούμε ότι οι δύο συντελεστές διαφέρουν μεταξύ τους (ο ένας είναι **1555,205** και ο άλλος **23,721**). Αυτή η διαφορά τους σημαίνει ότι ο ρυθμός με τον οποίο αυξάνονται τα κρούσματα είναι μεγαλύτερος από τον ρυθμό που μειώνονται οι θάνατοι.

ΕΡΩΤΗΣΗ 6.

Για υστέρηση T = 10:



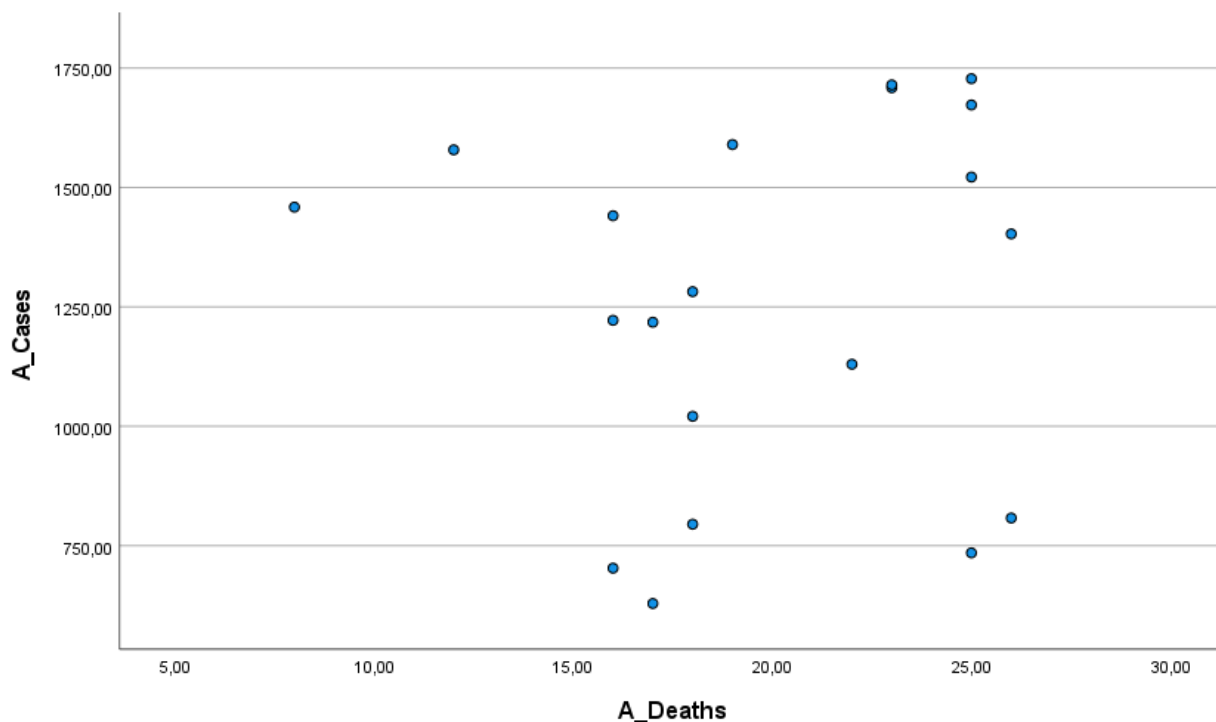
**Correlations**

		A_Deaths	A_Cases
A_Deaths	Pearson Correlation	1	-,103
	Sig. (2-tailed)		,665
	N	20	20
A_Cases	Pearson Correlation	-,103	1
	Sig. (2-tailed)	,665	
	N	20	20

Ημερήσια Κρούσματα	Ημερήσιοι Θάνατοι
1.715	25
1.579	25
1.459	26
795	22
735	23
1.403	23
1.673	12
1.390	8
1.441	18
1.213	25
629	26
703	23
1.021	19
1.282	16
1.222	17
815	17
746	16
359	18
569	18
927	16

Παρατηρούμε από τον πίνακα ότι ο συντελεστής συσχέτισης είναι αρνητικός, επομένως η γραμμική συσχέτιση είναι αρνητική, δηλαδή οι θάνατοι μειώνονται καθώς τα κρούσματα αυξάνονται. Ο συντελεστής αποκλίνει αρκετά από την αρνητική μονάδα, επομένως η συσχέτιση κρίνεται ασθενής και δεν μπορεί να γίνει αξιόπιστη πρόβλεψη για τους θανάτους με βάση τα ημερήσια κρούσματα 10 μέρες πριν.

Για υστέρηση  $T = 15$ :



### Correlations

		A_Deaths	A_Cases
A_Deaths	Pearson Correlation	1	,098
	Sig. (2-tailed)		,681
	N	20	20
A_Cases	Pearson Correlation	,098	1
	Sig. (2-tailed)	,681	
	N	20	20

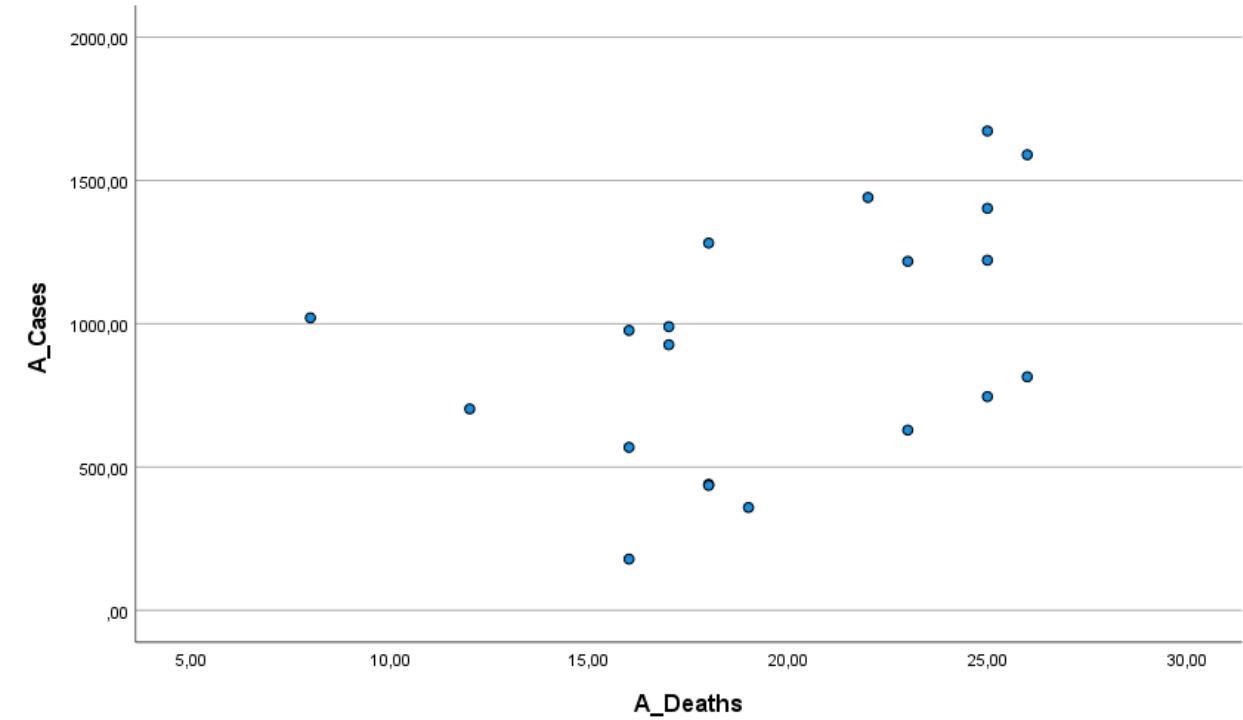
Ημερήσια Κρούσματα	Ημερήσιοι Θάνατοι
1.728	643
808	513
1.130	530
1.709	234
1.766	234
1.715	234
1.579	530
1.459	337
795	388
735	397
1.403	178
1.673	178
1.390	179
1.441	443
1.213	389
629	345
703	329
1.021	156
1.282	156
1.222	155

Για υστέρηση T = 5:

Ο συντελεστής συσχέτισης είναι θετικός, επομένως και η γραμμική συσχέτιση είναι θετική, δηλαδή οι θάνατοι αυξάνονται καθώς αυξάνονται τα ημερήσια κρούσματα. Ο συντελεστής αποκλίνει αρκετά από την μονάδα, επομένως η συσχέτιση καθίσταται ασθενής και δεν έχουμε αξιόπιστη πρόβλεψη για τους θανάτους 15 μέρες πριν.

Ημερήσια Κρούσματα	Ημερήσιοι Θάνατοι
1.715	25
1.579	25
1.459	26

795	22
735	23
1.403	23
1.673	12
1.390	8
1.441	18
1.213	25
629	26
703	23
1.021	19
1.282	16
1.222	17
990	17
977	16
440	18
436	18
179	16



Correlations

		A_Deaths	A_Cases
A_Deaths	Pearson Correlation	1	,429
	Sig. (2-tailed)		,059
	N	20	20
A_Cases	Pearson Correlation	,429	1
	Sig. (2-tailed)	,059	
	N	20	20

Από τον πίνακα ο συντελεστής συσχέτισης φαίνεται θετικός, επομένως και η γραμμική συσχέτιση είναι θετική, δηλαδή οι θάνατοι αυξάνονται όσο αυξάνονται τα ημερήσια κρούσματα. Ο συντελεστής αποκλίνει

αρκετά από τη μονάδα και η συσχέτιση παραμένει ασθενής. Η πρόβλεψη δεν χαρακτηρίζεται από μεγάλο βαθμό αξιοπιστίας, ωστόσο κρίνεται αρκετά πιο έμπιστη από τις προηγούμενες.

## ΕΡΩΤΗΜΑ 7.

**Model Summary**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,103 <sup>a</sup>	,011	-,044	5,13508

a. Predictors: (Constant), A\_Cases

**Coefficients<sup>a</sup>**

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	21,121	3,319		6,363	,000
	A_Cases	-,001	,003	-,103	-,440	,665

a. Dependent Variable: A\_Deaths

Από τους πίνακες οι οποίοι περιέχουν τις εκτιμήσεις των παραμέτρων γραμμικής παλινδρόμησης (σταθερός όρος, κλίση και τυπική απόκλιση σφαλμάτων παλινδρόμησης) για τους θανάτους από τα ημερήσια κρούσματα με υστέρηση 10 μέρες παρατηρούμε ότι το R Square είναι 0,011, που σημαίνει ότι το μοντέλο είναι

1,1% κατάλληλο για προβλέψεις, ποσοστό που δεν μας καλύπτει. Από την μέθοδο των ελαχίστων τετραγώνων το γραμμικό μοντέλο είναι:

**Θάνατοι = 21,121 – 0,001 \* κρούσματα.**

**Model Summary**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,098 <sup>a</sup>	,010	-,045	5,13775

a. Predictors: (Constant), A\_Cases

Coefficients <sup>a</sup>					
Model		Unstandardized Coefficients		Standardized Coefficients	Sig.
		B	Std. Error	Beta	
1	(Constant)	18,077	4,163		,000
	A_Cases	,001	,003	,098	,681

a. Dependent Variable: A\_Deaths

Από τους παραπάνω πίνακες που περιέχουν τις εκτιμήσεις των παραμέτρων της γραμμικής παλινδρόμησης (σταθερός όρος, κλίση και τυπική απόκλιση σφαλμάτων παλινδρόμησης) για τους θανάτους από τα ημερήσια κρούσματα με υστέρηση 15 μέρες παρατηρούμε ότι το R Square είναι 0,01, που σημαίνει ότι το μοντέλο είναι 1% κατάλληλο για προβλέψεις, επίσης μια τιμή που δεν μας καλύπτει. Από την μέθοδο των ελαχίστων τετραγώνων το γραμμικό μοντέλο είναι:

**Θάνατοι = 18,077 + 0,001 \* κρούσματα.**

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,429 <sup>a</sup>	,184	,139	4,66245

a. Predictors: (Constant), A\_Cases

Coefficients <sup>a</sup>					
Model		Unstandardized Coefficients		Standardized Coefficients	Sig.
		B	Std. Error	Beta	
1	(Constant)	15,051	2,552		,000
	A_Cases	,005	,003	,429	,059

a. Dependent Variable: A\_Deaths

Από τους δύο τελευταίους πίνακες με τις εκτιμήσεις των παραμέτρων της γραμμικής παλινδρόμησης (σταθερός όρος, κλίση και τυπική απόκλιση σφαλμάτων παλινδρόμησης) για τους θανάτους από τα ημερήσια κρούσματα με υστέρηση 15 μέρες παρατηρούμε ότι το R Square είναι 0,184, που σημαίνει ότι το μοντέλο είναι περίπου 18% κατάλληλο για προβλέψεις και πιθανότατα θα μας οδηγήσει σε ένα μη ικανοποιητικό αποτέλεσμα. Από την μέθοδο των ελαχίστων τετραγώνων το γραμμικό μοντέλο είναι:

**Θάνατοι = 15,051 + 0,005 \* κρούσματα.**



Τελειώνοντας, κανένα από τα τρία μοντέλα δεν πληροί τις κατάλληλες προϋποθέσεις για την εφαρμογή του, ωστόσο καλύτερο από τα τρία καθίσταται το τρίτο με αξιοπιστία 18% σε σύγκριση με τα υπόλοιπα δύο αξιοπιστίας 1,1% και 1%.