

Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης
Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών

ΘΕΩΡΙΑ ΠΙΘΑΝΟΤΗΤΩΝ ΚΑΙ ΣΤΑΤΙΣΤΙΚΗ

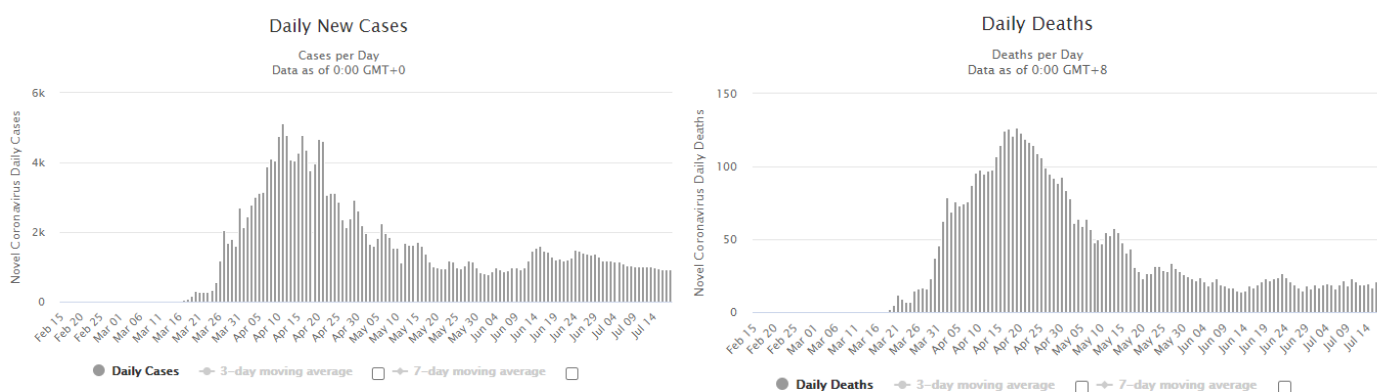
Εργασία SPSS

Φωτεινή Σαββίδου (9657)

ΣΥΛΛΟΓΗ ΔΕΔΟΜΕΝΩΝ

Χώρα Α: Τουρκία

Στην *Εικόνα 1* δίνονται τα στοιχεία για τα ημερήσια νέα κρούσματα και τους ημερήσιους νέους θανάτους στην Τουρκία, όπως έχουν καταγραφεί στον ιστοχώρο <https://www.worldometers.info/coronavirus/>.



Εικόνα 1: Γραφήματα για τα ημερήσια νέα κρούσματα (αριστερά) και τους ημερήσιους νέους θανάτους (δεξιά) στην Τουρκία.

Η κορύφωση των ημερήσιων νέων κρουσμάτων συμβαίνει στις 11 Απριλίου με 5138 νέα κρούσματα και στη συνέχεια ο αριθμός των ημερήσιων κρουσμάτων μειώνεται. Στο διάγραμμα δίνονται ακόμα 3 ημέρες με υψηλό αριθμό κρουσμάτων (16, 20 και 21 Απριλίου), αλλά δεν υπάρχουν επαρκείς πληροφορίες¹ για την προέλευση των νέων περιπτώσεων που ενδεχομένως να οδηγούσαν σε επιλογή διαφορετικής ημέρας κορύφωσης της «καμπύλης». Αντίστοιχα, η κορύφωση των ημερήσιων νέων θανάτων συμβαίνει στις 19 Απριλίου με 127 νέους θανάτους. Για τη Μελέτη Α χρησιμοποιούμε τα στοιχεία για τα ημερήσια νέα κρούσματα από τις 12 Απριλίου έως τη 1 Μαΐου και για τους ημερήσιους νέους θανάτους από τις 20 Απριλίου έως τις 9 Μαΐου. Στη συνέχεια διαιρούμε κάθε τιμή των ημερήσιων νέων θανάτων με την αντίστοιχη τιμή των ημερήσιων νέων κρουσμάτων και υπολογίζουμε το αντίστοιχο ποσοστό. Τα αποτελέσματα παρουσιάζονται στους *Πίνακες 1* και *2*.

¹ Πηγή: <https://covid19.saglik.gov.tr/?lang=en-US>

| i | Ημερήσια νέα κρούσματα | Ημερήσιοι νέοι θάνατοι |
|----|------------------------|------------------------|
| 1 | 4789 | 123 |
| 2 | 4093 | 119 |
| 3 | 4062 | 117 |
| 4 | 4281 | 115 |
| 5 | 4801 | 109 |
| 6 | 4353 | 106 |
| 7 | 3783 | 99 |
| 8 | 3977 | 95 |
| 9 | 4674 | 92 |
| 10 | 4611 | 89 |
| 11 | 3083 | 93 |
| 12 | 3116 | 84 |
| 13 | 3122 | 78 |
| 14 | 2861 | 61 |
| 15 | 2357 | 64 |
| 16 | 2131 | 59 |
| 17 | 2392 | 64 |
| 18 | 2936 | 57 |
| 19 | 2615 | 48 |
| 20 | 2188 | 50 |

Πίνακας 1: Τα δεδομένα για τα ημερήσια νέα κρούσματα και τους ημερήσιους νέους θανάτους τα διαστήματα 12/4 με 1/5 και 20/4 με 9/5 αντίστοιχα.

| i | Ποσοστό θνητότητας (%) | i | Ποσοστό θνητότητας (%) |
|----|------------------------|----|------------------------|
| 1 | 2,57 | 11 | 3,02 |
| 2 | 2,91 | 12 | 2,7 |
| 3 | 2,88 | 13 | 2,5 |
| 4 | 2,69 | 14 | 2,13 |
| 5 | 2,27 | 15 | 2,72 |
| 6 | 2,44 | 16 | 2,77 |
| 7 | 2,62 | 17 | 2,68 |
| 8 | 2,39 | 18 | 1,94 |
| 9 | 1,97 | 19 | 1,84 |
| 10 | 1,93 | 20 | 2,29 |

Πίνακας 2: Ημερήσιο ποσοστό θνητότητας στην Τουρκία την περίοδο μετά την κορύφωση του φαινομένου ως προς τα ημερήσια νέα κρούσματα (Δείγμα Α, Μελέτη Α).

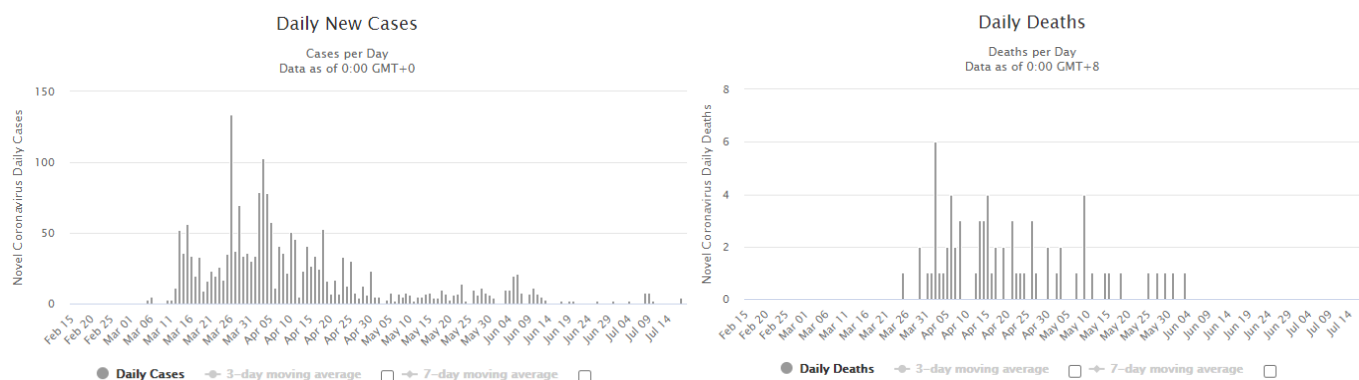
Για τη Μελέτη Β καταγράφουμε τα ημερήσια νέα κρούσματα για 10 ημέρες πριν την κορύφωση του φαινομένου, δηλαδή στο διάστημα από 1 Απριλίου (1^η ημέρα) έως 10 Απριλίου (10^η ημέρα). Τα αποτελέσματα παρουσιάζονται στον Πίνακα 3.

| Ημέρα | Ημερήσια νέα κρούσματα |
|-------|------------------------|
| 1 | 2148 |
| 2 | 2456 |
| 3 | 2786 |
| 4 | 3013 |
| 5 | 3135 |
| 6 | 3148 |
| 7 | 3892 |
| 8 | 4117 |
| 9 | 4056 |
| 10 | 4747 |

Πίνακας 3: Ημερήσια νέα κρούσματα στην Τουρκία την περίοδο πριν την κορύφωση του φαινομένου (Δείγμα Α, Μελέτη Β).

Χώρα Β: Εσθονία

Στην *Εικόνα 2* δίνονται τα στοιχεία για τα ημερήσια νέα κρούσματα και τους ημερήσιους νέους θανάτους στην Εσθονία, όπως έχουν καταγραφεί στον ιστοχώρο <https://www.worldometers.info/coronavirus/>.



Εικόνα 2: Γραφήματα για τα ημερήσια νέα κρούσματα (αριστερά) και τους ημερήσιους νέους θανάτους (δεξιά) στην Εσθονία.

Στις 26 Μαρτίου παρατηρείται η μέγιστη τιμή των ημερήσιων νέων κρουσμάτων (134 κρούσματα). Η τιμή αυτή όμως θεωρείται ως ειδική καταγραφή λόγω αλλαγής στον τρόπο συλλογής και καταγραφής των νέων κρουσμάτων². Έτσι, θεωρούμε πως η κορύφωση του φαινομένου συμβαίνει στις 3 Απριλίου με 103 νέα κρούσματα. Επίσης, στις 12 Απριλίου καταγράφηκαν 5 νέα κρούσματα. Η τιμή αυτή είναι αρκετά χαμηλή σε σχέση με τις «γειτονικές» τιμές και ενδέχεται να οδηγήσει σε λανθασμένα ποσοστά θνητότητας ως προς τα ημερήσια κρούσματα για τη χώρα. Επειδή υπάρχει κάποια επιφύλαξη σχετικά με το κατά πόσο αυτή η τιμή είναι ενδεικτική της εξέλιξης της επιδημίας στην χώρα, την απορρίπτω από το δείγμα (9^η παρατήρηση).

Για τους ημερήσιους νέους θανάτους εξετάζουμε 3 πιθανές ημέρες (2 Απριλίου με 6 νέους θανάτους, 6 και 15 Απριλίου με 4 νέους θανάτους) για την κορύφωση της «καμπύλης». Στις 2 Απριλίου παρότι παρατηρείται η μέγιστη τιμή των ημερήσιων νέων θανάτων, δεν θεωρούμε πως η ημέρα αυτή αποτελεί την κορύφωση του φαινομένου, καθώς το επόμενο διάστημα παρατηρείται αύξηση των ημερήσιων θανάτων (η σύγκριση γίνεται λαμβάνοντας υπόψη το μικρό εύρος τιμών των νέων θανάτων). Επίσης, στις 15 Απριλίου καταγράφηκαν 3 νέοι θάνατοι και ένας θάνατος που συνέβη στις 10 Απριλίου και αποδείχθηκε πως οφειλόταν στον ιό COVID-19³. Άρα, θεωρούμε ότι η κορύφωση του φαινομένου συμβαίνει στις 6 Απριλίου με 4 νέους θανάτους.

² Πηγή: <https://www.terviseamet.ee/et/uudised/eile-tuvastasid-laborid-2-paeva-testide-pohjal-134-nakatunut>

³ Πηγή: <https://www.terviseamet.ee/et/uudised/viimase-oopaevaga-lisandus-29-positiivset-covid-19-analuusi>

Για τη Μελέτη Α χρησιμοποιούμε τα στοιχεία για τα ημερήσια νέα κρούσματα από τις 4 έως 24 Απριλίου (χωρίς την τιμή στις 12 Απριλίου) και για τους ημερήσιους νέους θανάτους από τις 7 έως 27 Απριλίου (αγνοούμε την 9^η παρατήρηση, δηλαδή την τιμή στις 15 Απριλίου). Στη συνέχεια διαιρούμε κάθε τιμή των ημερήσιων νέων θανάτων με την αντίστοιχη τιμή των ημερήσιων νέων κρουσμάτων και υπολογίζουμε το αντίστοιχο ποσοστό. Τα αποτελέσματα παρουσιάζονται στους Πίνακες 4 και 5.

| i | Ημερήσια νέα κρούσματα | Ημερήσιοι νέοι θάνατοι |
|----|------------------------|------------------------|
| 1 | 78 | 2 |
| 2 | 58 | 3 |
| 3 | 11 | 0 |
| 4 | 41 | 0 |
| 5 | 36 | 0 |
| 6 | 22 | 1 |
| 7 | 51 | 3 |
| 8 | 46 | 3 |
| 9 | 23 | 1 |
| 10 | 41 | 2 |
| 11 | 27 | 0 |
| 12 | 34 | 2 |
| 13 | 25 | 0 |
| 14 | 53 | 3 |
| 15 | 16 | 1 |
| 16 | 7 | 1 |
| 17 | 17 | 1 |
| 18 | 7 | 0 |
| 19 | 33 | 3 |
| 20 | 13 | 1 |

Πίνακας 4: Τα δεδομένα για τα ημερήσια νέα κρούσματα και τους ημερήσιους νέους θανάτους τα διαστήματα 4/4 με 24/4 και 7/4 με 27/4 αντίστοιχα.

| i | Ποσοστό θνητότητας (%) | i | Ποσοστό θνητότητας (%) |
|----|------------------------|----|------------------------|
| 1 | 2,56 | 11 | 0 |
| 2 | 5,17 | 12 | 5,88 |
| 3 | 0 | 13 | 0 |
| 4 | 0 | 14 | 5,66 |
| 5 | 0 | 15 | 6,25 |
| 6 | 4,55 | 16 | 14,29 |
| 7 | 5,88 | 17 | 5,88 |
| 8 | 6,52 | 18 | 0 |
| 9 | 4,35 | 19 | 9,09 |
| 10 | 4,88 | 20 | 7,69 |

Πίνακας 5: Ημερήσιο ποσοστό θνητότητας στην Εσθονία την περίοδο μετά την κορύφωση του φαινομένου ως προς τα ημερήσια νέα κρούσματα (Δείγμα Β, Μελέτη Α).

Για τη Μελέτη Β καταγράφουμε τα ημερήσια νέα κρούσματα για 10 ημέρες πριν την κορύφωση του φαινομένου. Απορρίπτουμε την τιμή στις 26 Μαρτίου, καθώς όπως αναφέρεται παραπάνω είναι ειδική καταγραφή. Τα ημερήσια νέα κρούσματα το διάστημα από 23 Μαρτίου (1^η ημέρα) έως 2 Απριλίου (10^η ημέρα) παρουσιάζονται στον Πίνακα 6.

| Ημέρα | Ημερήσια νέα κρούσματα |
|-------|------------------------|
| 1 | 26 |
| 2 | 17 |
| 3 | 35 |
| 4 | 37 |
| 5 | 70 |
| 6 | 34 |
| 7 | 36 |
| 8 | 30 |
| 9 | 34 |
| 10 | 79 |

Πίνακας 6: Ημερήσια νέα κρούσματα στην Εσθονία την περίοδο πριν την κορύφωση του φαινομένου (Δείγμα Β, Μελέτη Β).

ΜΕΛΕΤΗ Α

Ζήτημα 1:

Στον Πίνακα Α.1 παρουσιάζονται τα μέτρα κεντρικής τάσης (μέση τιμή, διάμεσος) και μεταβλητότητας (διασπορά, τυπική απόκλιση, εύρος δεδομένων, πρώτο και τρίτο τεταρτομόριο) για τα δύο δείγματα και στα Σχήματα Α.1.1, Α.1.2 και Α.1.3 τα ιστογράμματα και το θηκόγραμμα των δύο μεταβλητών που μελετάμε.

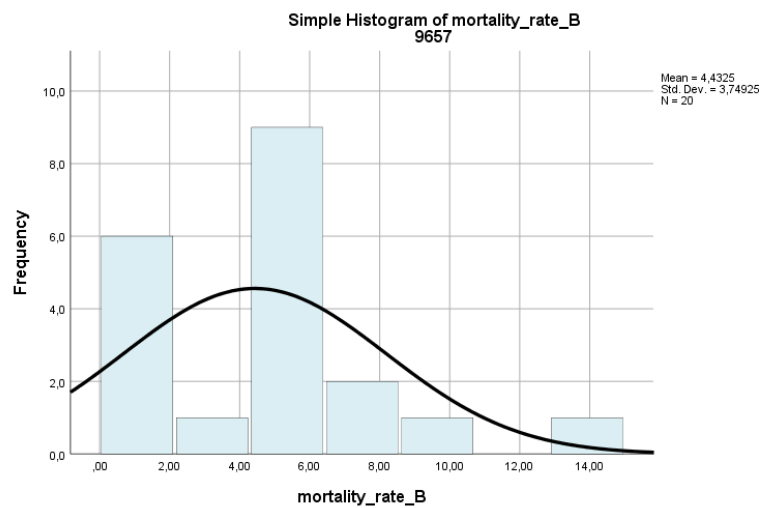
| Statistics | | AEM: 9657 | |
|----------------|---------|------------------|------------------|
| | | mortality_rate_A | mortality_rate_B |
| N | Valid | 20 | 20 |
| | Missing | 0 | 0 |
| Mean | | 2,4630 | 4,4325 |
| Median | | 2,5350 | 5,0250 |
| Std. Deviation | | ,35609 | 3,74925 |
| Variance | | ,127 | 14,057 |
| Range | | 1,18 | 14,29 |
| Percentiles | 25 | 2,1650 | ,0000 |
| | 50 | 2,5350 | 5,0250 |
| | 75 | 2,7150 | 6,1575 |

Πίνακας Α.1: Συνοπτικά μέτρα κεντρικής τάσης και μεταβλητότητας για τα δύο δείγματα.

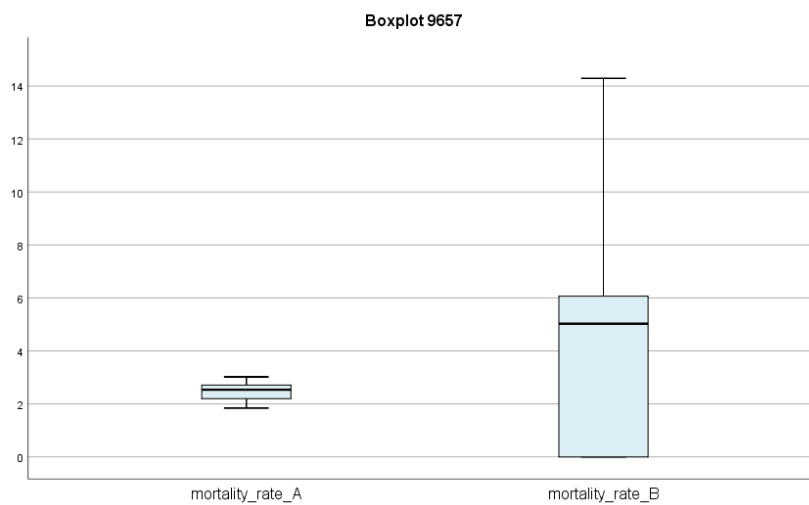
Παρατηρούμε πως η μέση τιμή του ημερήσιου ποσοστού θνητότητας στη χώρα Β είναι περίπου διπλάσια της μέσης τιμής στη χώρα Α. Επίσης, το εύρος των τιμών του δείγματος της χώρας Β είναι πολύ μεγαλύτερο από αυτό της χώρας Α και παρατηρείται μεγάλη διαφορά στην τυπική απόκλιση των δύο δειγμάτων. Οι τιμές του δείγματος Β παρουσιάζουν μεγαλύτερη διασπορά από τις τιμές του δείγματος Α.



Σχήμα A.1.1: Ιστόγραμμα του ημερήσιου ποσοστού θνητότητας της χώρας A και η καμπύλης της κανονικής κατανομής.



Σχήμα A.1.2: Ιστόγραμμα του ημερήσιου ποσοστού θνητότητας της χώρας B και η καμπύλης της κανονικής κατανομής.



Σχήμα A.1.3: Θηκόγραμμα του ημερήσιου ποσοστού θνητότητας στις χώρες A και B.

Για τη χώρα Α μπορούμε να θεωρήσουμε, με κάποια επιφύλαξη λόγω του μικρού μεγέθους του δείγματος, πως το ημερήσιο ποσοστό θνητότητας ακολουθεί κανονική κατανομή. Στο ιστόγραμμα φαίνεται πως οι παρατηρήσεις συγκεντρώνονται στις κεντρικές κλάσεις και το θηκόγραμμα παρουσιάζει σημαντική συμμετρία. Οι δύο μύστακες έχουν παρόμοιο μήκος και η διάμεσος είναι περίπου στο κέντρο της θήκης.

Με βάση το ιστόγραμμα και το θηκόγραμμα του ημερήσιου ποσοστού θνητότητας της χώρας Β δεν μπορούμε να θεωρήσουμε πως η μεταβλητή ακολουθεί κανονική κατανομή, καθώς δεν παρουσιάζει χαρακτηριστική συμμετρία. Οι μύστακες δεν έχουν το ίδιο μήκος και η διάμεσος τείνει προς το τρίτο τεταρτομόριο.

Επίσης, από το θηκόγραμμα φαίνεται πως η κεντρική τάση και η διασπορά των τιμών της δεύτερης μεταβλητής είναι μεγαλύτερη σε σχέση με της πρώτης μεταβλητής.

Ζήτημα 2:

Το 95% διάστημα εμπιστοσύνης της μέσης τιμής του ημερήσιου ποσοστού θνητότητας στη χώρα Α είναι: [2,2963, 2,6297] και στη χώρα Β: [2,6778, 6,1872]. Η τιμή 10 δεν περιλαμβάνεται σε κανένα από τα δύο διαστήματα, οπότε η μέση τιμή και των δύο μεταβλητών δεν μπορεί να πάρει την τιμή 10% σε επίπεδο εμπιστοσύνης 95%.

Επίσης, το εύρος του διαστήματος εμπιστοσύνης της χώρας Α είναι 0,33 και της χώρας Β είναι 3,51. Παρατηρούμε ότι το εύρος του διαστήματος εμπιστοσύνης στην πρώτη περίπτωση είναι πολύ μικρότερο από το εύρος του δεύτερου διαστήματος και άρα η εκτίμηση της μέσης τιμής του ημερήσιου ποσοστού θνητότητας της χώρας Α είναι πιο ακριβής.

One-Sample Test

AEM: 9657

Test Value = 0

| | t | df | Sig. (2-tailed) | Mean Difference | 95% Confidence Interval of the Difference | |
|------------------|--------|----|-----------------|-----------------|---|--------|
| | | | | | Lower | Upper |
| mortality_rate_A | 30,933 | 19 | ,000 | 2,46300 | 2,2963 | 2,6297 |
| mortality_rate_B | 5,287 | 19 | ,000 | 4,43250 | 2,6778 | 6,1872 |

Πίνακας Α.2: Το 95% διάστημα εμπιστοσύνης της μέσης τιμής του ημερήσιου ποσοστού θνητότητας στις χώρες Α και Β.

Ζήτημα 3:

Οι δειγματικές διασπορές που υπολογίστηκαν με βάση τα δύο δείγματα ($s_1^2=0,127$, $s_2^2=14,057$) διαφέρουν σημαντικά, οπότε υποθέτουμε πως το ημερήσιο ποσοστό θνητότητας στις δύο χώρες δεν έχει την ίδια διασπορά. Σε αυτήν την περίπτωση το 95% διάστημα εμπιστοσύνης της διαφοράς των μέσων τιμών ($\mu_1-\mu_2$) είναι: $[-3,73, -0,209]$. Η τιμή 0 δεν περιέχεται στο διάστημα εμπιστοσύνης, άρα η μέση τιμή του ημερήσιου ποσοστού θνητότητας δεν μπορεί να είναι ίδια στις δύο χώρες σε επίπεδο εμπιστοσύνης 95%. Συγκεκριμένα, η μέση τιμή στη χώρα Β είναι μεγαλύτερη της μέσης τιμής στη χώρα Α κατά μία τιμή στο διάστημα $[0,209, 3,73]$.

Αν θεωρήσουμε ότι οι μεταβλητές έχουν ίδια διασπορά, τότε τα συμπεράσματά μας δεν διαφοροποιούνται.

| Independent Samples Test | | | | | | | | | |
|--------------------------|-----------------------------|---|------|------------------------------|--------|-----------------|-----------------|-----------------------|---|
| 9657 | | | | | | | | | |
| | | 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 |
| mortality_rate | Equal variances assumed | 23,752 | ,000 | -2,339 | 38 | ,025 | -1,96950 | ,84213 | -3,67430 - ,26470 |
| | Equal variances not assumed | | | -2,339 | 19,343 | ,030 | -1,96950 | ,84213 | -3,72999 - ,20901 |

Πίνακας Α.3: Το 95% διάστημα εμπιστοσύνης της διαφοράς των μέσων τιμών ($\mu_1-\mu_2$) του ημερήσιου ποσοστού θνητότητας στις χώρες Α και Β.

Ζήτημα 4:

Σύμφωνα με τα παραπάνω αποτελέσματα, φαίνεται πως η χώρα Α χαρακτηρίζεται από ημερήσιο ποσοστό θνητότητας στο διάστημα $[2, 3]$. Από το ιστόγραμμα βρέθηκε πως οι περισσότερες παρατηρήσεις συγκεντρώνονται στο κεντρικό τμήμα. Επίσης, αν θεωρήσουμε πως η μεταβλητή ακολουθεί κανονική κατανομή, τότε η πιθανότητα το ημερήσιο ποσοστό θνητότητας να λαμβάνει τιμή στο διάστημα $[2, 3]$ είναι:

$$P(2 < X < 3) = \Phi\left(\frac{3 - 2,463}{0,35609}\right) - \Phi\left(\frac{2 - 2,463}{0,35609}\right) = \Phi(1,51) - \Phi(-1,3) = 0,9345 - (1 - 0,9032) = 0,838 \text{ ή } 83,8\%$$

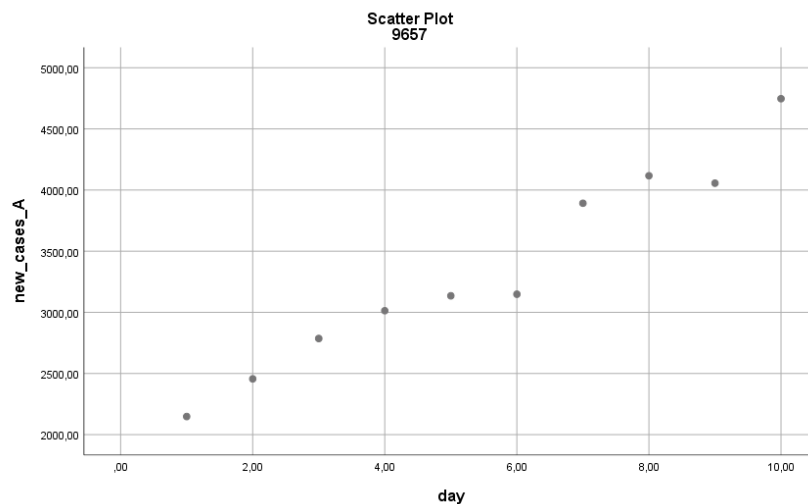
οπότε μπορούμε να εκτιμήσουμε πως το ημερήσιο ποσοστό θνητότητας στη χώρα Α θα κυμαίνεται από 2% έως 3% την περίοδο μετά την κορύφωση.

Για τη χώρα Β οι περισσότερες παρατηρήσεις του δείγματος συγκεντρώνονται στο διάστημα $[0, 6]$. Ωστόσο, φαίνεται πως η κατανομή της μεταβλητής δεν είναι συμμετρική και οι τιμές του δείγματος παρουσιάζουν μεγάλη διασπορά. Άρα, δεν μπορούμε να εκτιμήσουμε με μεγάλη ακρίβεια τα ποσοστά θνητότητας που θα χαρακτηρίζουν τη χώρα την περίοδο μετά την κορύφωση.

ΜΕΛΕΤΗ Β

Ζήτημα 5:

Από το διάγραμμα διασποράς για τη χώρα Α (Σχήμα Β.1.1) παρατηρούμε ότι υπάρχει ισχυρή γραμμική εξάρτηση του αριθμού των ημερήσιων νέων κρουσμάτων από τον αύξοντα αριθμό της ημέρας, καθώς τα σημεία στο διάγραμμα διασποράς βρίσκονται «κοντά» σε μία νοητή γραμμή. Το συμπέρασμα αυτό επιβεβαιώνεται και από τον συντελεστή συσχέτισης, οποίος είναι ίσος με $r=0.979$. Άρα, αν γνωρίζουμε τον αριθμό της ημέρας μπορούμε να υπολογίσουμε με μεγάλη ακρίβεια τον αριθμό των νέων κρουσμάτων.



Σχήμα Β.1.1: Διάγραμμα διασποράς του αριθμού των ημερήσιων νέων κρουσμάτων σε σχέση με τον αύξοντα αριθμό της ημέρας τη χρονική περίοδο πριν την κορύφωση για τη χώρα Α.

Correlations

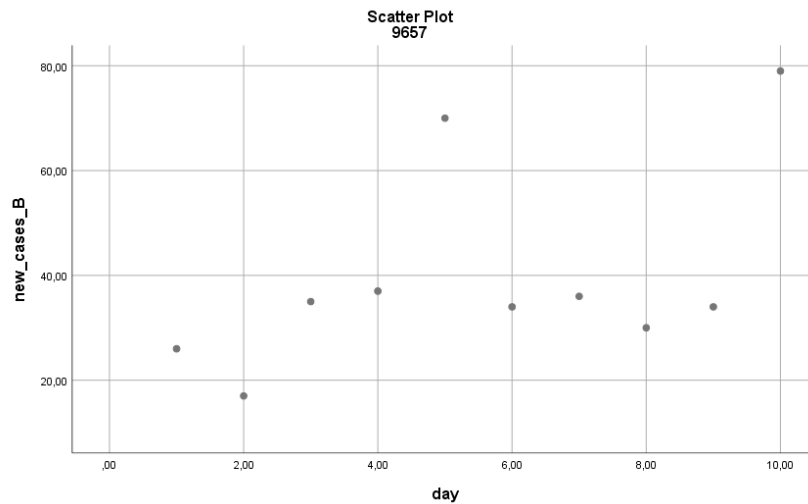
AEM: 9657

| | | new_cases_A | day |
|-------------|---------------------|-------------|--------|
| new_cases_A | Pearson Correlation | 1 | ,979** |
| | Sig. (2-tailed) | | ,000 |
| | N | 10 | 10 |
| day | Pearson Correlation | ,979** | 1 |
| | Sig. (2-tailed) | ,000 | |
| | N | 10 | 10 |

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

Πίνακας Β.1.1: Ο συντελεστής συσχέτισης για τον αύξοντα αριθμό της ημέρας και τον αριθμό των ημερήσιων νέων κρουσμάτων στη χώρα Α.

Από το διάγραμμα διασποράς (Σχήμα Β.1.2) της χώρας Β παρατηρούμε ότι υπάρχει ασθενής γραμμική εξάρτηση του αριθμού των ημερήσιων νέων κρουσμάτων από τον αύξοντα αριθμό της ημέρας, καθώς τα σημεία στο διάγραμμα διασποράς είναι αρκετά «σκορπισμένα» γύρω από μία νοητή γραμμή. Ο συντελεστής συσχέτισης είναι $r=0,505$, οπότε συμπεραίνουμε ότι αν γνωρίζουμε τον αριθμό της ημέρας δεν μπορούμε να υπολογίσουμε με μεγάλη ακρίβεια τον αριθμό των ημερήσιων νέων κρουσμάτων στη χώρα.



Σχήμα Β.1.2: Διάγραμμα διασποράς του αριθμού των ημερήσιων νέων κρουσμάτων σε σχέση με τον αύξοντα αριθμό της ημέρας τη χρονική περίοδο πριν την κορύφωση για τη χώρα Β.

Correlations

AEM: 9657

| | | new_cases_B | day |
|-------------|---------------------|-------------|------|
| new_cases_B | Pearson Correlation | 1 | ,505 |
| | Sig. (2-tailed) | | ,137 |
| | N | 10 | 10 |
| day | Pearson Correlation | ,505 | 1 |
| | Sig. (2-tailed) | ,137 | |
| | N | 10 | 10 |

Πίνακας Β.1.2: Ο συντελεστής συσχέτισης για τον αύξοντα αριθμό της ημέρας και τον αριθμό των ημερήσιων νέων κρουσμάτων στη χώρα Β.

Ζήτημα 6:

Το μοντέλο γραμμικής παλινδρόμησης για τη χώρα Α είναι:

$$new_cases = 1886,6 + 266,036t$$

και η τυπική απόκλιση των σφαλμάτων παλινδρόμησης είναι $\sigma=178,83844$, τιμή χαμηλή σε σχέση με το επίπεδο τιμών των ημερήσιων νέων κρουσμάτων. Επίσης, από τον συντελεστή παλινδρόμησης προκύπτει ότι κάθε μέρα ο αριθμός των νέων κρουσμάτων αυξάνεται περίπου κατά την τιμή 266, τιμή που συμφωνεί με τα στοιχεία του δείγματος. Το γραμμικό αυτό μοντέλο παλινδρόμησης κρίνεται αρκετά ακριβές.

Το μοντέλο γραμμικής παλινδρόμησης για τη χώρα Β είναι:

$$new_cases = 22,067 + 3,224t$$

και η τυπική απόκλιση του σφάλματος είναι $\sigma=17,71282$, τιμή αρκετά υψηλή σε σχέση με το επίπεδο τιμών των ημερήσιων νέων κρουσμάτων στη χώρα Β. Το μοντέλο παλινδρόμησης δεν είναι αρκετά ακριβές και κρίνεται ακατάλληλο για προβλέψεις, καθώς, όπως προέκυψε και από το Ζήτημα 5, η γραμμική εξάρτηση του αριθμού των νέων κρουσμάτων από την ημέρα είναι ασθενής.

Coefficients

AEM: 9657^a

| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
|-------|------------|-----------------------------|------------|---------------------------|--------|------|
| | | B | Std. Error | Beta | | |
| 1 | (Constant) | 1886,600 | 122,170 | | 15,442 | ,000 |
| | day | 266,036 | 19,689 | ,979 | 13,512 | ,000 |

a. Dependent Variable: new_cases_A

Model Summary

AEM: 9657

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
|-------|-------------------|----------|-------------------|----------------------------|
| 1 | ,979 ^a | ,958 | ,953 | 178,83844 |

a. Predictors: (Constant), day

Πίνακας Β.2.1: Οι παράμετροι του μοντέλου γραμμικής παλινδρόμησης για τη χώρα Α.

| Coefficients AEM: 9657 ^a | | | | | |
|--|------------|-----------------------------|------------|---------------------------|-------|
| | | Unstandardized Coefficients | | Standardized Coefficients | |
| Model | | B | Std. Error | Beta | t |
| 1 | (Constant) | 22,067 | 12,100 | | 1,824 |
| | day | 3,224 | 1,950 | ,505 | 1,653 |

a. Dependent Variable: new_cases_B

| Model Summary AEM: 9657 | | | | |
|----------------------------|-------------------|----------|-------------------|----------------------------|
| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
| 1 | ,505 ^a | ,255 | ,162 | 17,71282 |

a. Predictors: (Constant), day

Πίνακας Β.2.2: Οι παράμετροι του μοντέλου γραμμικής παλινδρόμησης για τη χώρα Β.

Τα δύο μοντέλα γραμμικής παλινδρόμησης διαφέρουν. Στο πρώτο μοντέλο ο σταθερός όρος είναι πολύ μεγαλύτερος σε σχέση με αυτόν του δεύτερου μοντέλου, αφού στη χώρα Α τα ημερήσια νέα κρούσματα είναι περισσότερα από τα νέα κρούσματα στη χώρα Β. Επίσης, η κλίση της ευθείας στην πρώτη περίπτωση είναι μεγαλύτερη σε σχέση με τη δεύτερη περίπτωση, δηλαδή καθημερινά παρατηρείται μεγαλύτερη αύξηση στον αριθμό των νέων κρουσμάτων στη χώρα Α σε σχέση με τη χώρα Β, γεγονός που επιβεβαιώνεται και από τις τιμές του δείγματος.

Ζήτημα 7:

Για τη χώρα Α ο αριθμός νέων κρουσμάτων την ημέρα της κορύφωσης που υπολογίζεται από το μοντέλο γραμμικής παλινδρόμησης είναι (για $t=11$):

$$new_cases = 1886,6 + 266,036 \cdot 11 = 4812,996 \cong 4813$$

και ο πραγματικός αριθμός των νέων κρουσμάτων είναι 5138. Άρα, το σχετικό σφάλμα της εκτίμησης είναι:

$$\frac{5138 - 4813}{5138} = 0,063 \text{ ή } 6,63\%$$

Για τη χώρα Β ο αριθμός νέων κρουσμάτων την ημέρα της κορύφωσης που υπολογίζεται από το μοντέλο γραμμικής παλινδρόμησης είναι (για $t=11$):

$$new_cases = 22,067 + 3,224 \cdot 11 = 57,531 \cong 58$$

και ο πραγματικός αριθμός των νέων κρουσμάτων είναι 103. Άρα, το σχετικό σφάλμα της εκτίμησης είναι:

$$\frac{103 - 57,531}{103} = 0,441 \text{ ή } 44,1\%$$

Ζήτημα 8:

Χρησιμοποιούμε πολυωνυμική παλινδρόμηση 3^{ου} βαθμού για την πρόβλεψη της αύξησης των ημερήσιων νέων κρουσμάτων στις δύο χώρες, μέχρι την ημέρα της κορύφωσης του φαινομένου. Η γενική μορφή του μοντέλου παλινδρόμησης που θα εκτιμήσουμε είναι:

$$y = b_0 + b_1x + b_2x^2 + b_3x^3$$

Παρόλο που η γραμμική παλινδρόμηση έδωσε ικανοποιητικά αποτελέσματα στην πρόβλεψη της αύξησης των ημερήσιων νέων κρουσμάτων στη χώρα Α, κατασκευάζουμε και το μοντέλο πολυωνυμικής παλινδρόμησης. Το πολυώνυμο 3^{ου} βαθμού δίνεται από τον τύπο:

$$new_cases = 1810,8 + 386,279t - 34,696t^2 + 2,48t^3$$

Ο αριθμός των νέων κρουσμάτων την ημέρα της κορύφωσης θα είναι (για t=11):

$$new_cases = 1810,8 + 386,279 \cdot 11 - 34,696 \cdot 11^2 + 2,48 \cdot 11^3 = 5162,533$$

οπότε το σχετικό σφάλμα της εκτίμησης θα είναι:

$$\frac{|5138 - 5162,533|}{5138} = 0,0048 \text{ ή } 0,48\%$$

δηλαδή η εκτίμηση είναι πιο ακριβής σε σχέση με το γραμμικό μοντέλο.

Model Summary and Parameter Estimates

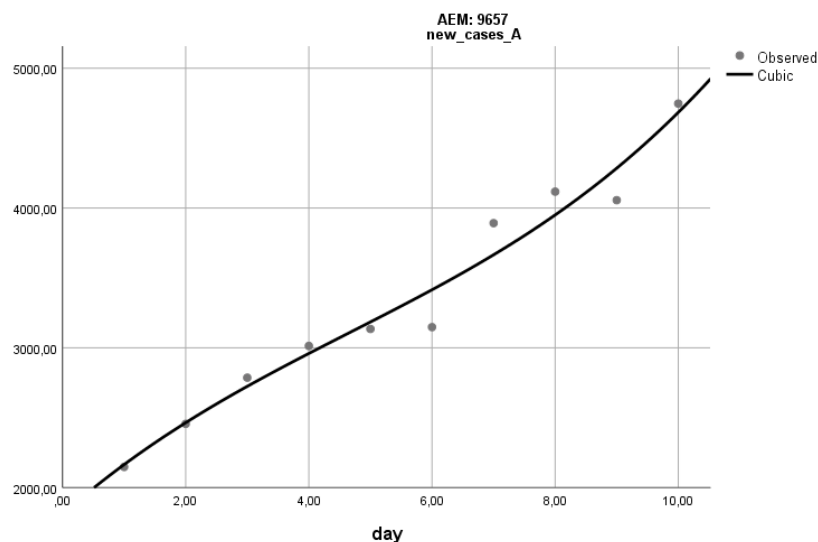
AEM: 9657

Dependent Variable: new_cases_A

| Model Summary | | | | | | Parameter Estimates | | | |
|---------------|----------|--------|-----|-----|------|---------------------|---------|---------|-------|
| Equation | R Square | F | df1 | df2 | Sig. | Constant | b1 | b2 | b3 |
| Cubic | ,964 | 54,332 | 3 | 6 | ,000 | 1810,800 | 386,279 | -34,696 | 2,480 |

The independent variable is day.

Πίνακας Β.4.1: Οι παράμετροι του μοντέλου πολυωνυμικής παλινδρόμησης 3^{ου} βαθμού για τη χώρα Α.



Σχήμα Β.4.1: Το διάγραμμα διασποράς για τη χώρα Α και το πολυώνυμο 3^{ου} βαθμού που εκτιμήθηκε.

Για τη χώρα Β εκτιμούμε την αύξηση των ημερήσιων νέων κρουσμάτων ως την ημέρα της κορύφωσης με το πολυώνυμο 3^{ου} βαθμού που δίνεται από τον τύπο:

$$new_cases = -19,033 + 40,71t - 8,295t^2 + 0,51t^3$$

Ο αριθμός των νέων κρουσμάτων την ημέρα της κορύφωσης θα είναι (για $t=11$):

$$new_cases = -19,033 + 40,71 \cdot 11 - 8,295 \cdot 11^2 + 0,51 \cdot 11^3 = 103,892$$

οπότε το σχετικό σφάλμα της εκτίμησης θα είναι:

$$\frac{|103 - 103,892|}{103} = 0,0087 \text{ ή } 0,87\%$$

δηλαδή η εκτίμηση είναι πολύ πιο ακριβής σε σχέση με το γραμμικό μοντέλο.

Model Summary and Parameter Estimates

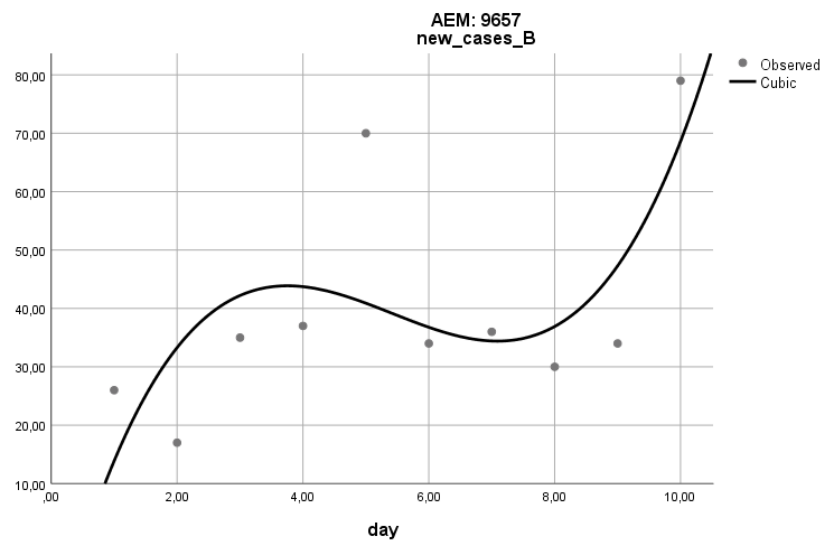
AEM: 9657

Dependent Variable: new_cases_B

| Model Summary | | | | | | Parameter Estimates | | | |
|---------------|----------|-------|-----|-----|------|---------------------|--------|--------|------|
| Equation | R Square | F | df1 | df2 | Sig. | Constant | b1 | b2 | b3 |
| Cubic | ,496 | 1,965 | 3 | 6 | ,221 | -19,033 | 40,710 | -8,295 | ,510 |

The independent variable is day.

Πίνακας Β.4.2: Οι παράμετροι του μοντέλου πολυωνυμικής παλινδρόμησης 3^{ου} βαθμού για τη χώρα Β.



Σχήμα Β.4.2: Το διάγραμμα διασποράς για τη χώρα Β και το πολυώνυμο 3^{ου} βαθμού που εκτιμήθηκε.