Κατερίνα Μοναστηριώτη

ΑΜ: 0291

Φεβρουάριος 2025

Contents

[Δεδομένα Εισόδου 3](#_Toc189068799)

[Επιθυμητή Επεξεργασία 4](#_Toc189068800)

[1. 4](#_Toc189068801)

[2 5](#_Toc189068802)

[Α 5](#_Toc189068803)

[Β 5](#_Toc189068804)

[Γ 5](#_Toc189068805)

[3 5](#_Toc189068806)

[4 6](#_Toc189068807)

[5 8](#_Toc189068808)

[Μέρος Β 9](#_Toc189068809)

[Θέμα 1 9](#_Toc189068810)

[Θέμα 2 11](#_Toc189068811)

# Δεδομένα Εισόδου

Τα δεδομένα εισόδου δημιουργήθηκαν σύμφωνα με τον αριθμό μητρώου μου ο οποίος είναι 0291. Η μεθοδολογία που ακολούθησα είναι στο Excel επέλεξα την επιλογή Δεδομένα, μετέπειτα Data Analysis. Στη συνέχεια, στο αναδυόμενο κατάλογο το στατιστικό εργαλείο επέλεξα την επιλογή Random Number Generation. Αφού το έκανα αυτό, έβαλα τις εξής τιμές στις μεταβλητές number of variables ίσον με δύο, Number of Random numbers ίσο με 291. Όσον αφορά την κατανομή επέλεξα normal. Στη μέση τιμή, έθεσα την τιμή ίση με 1. Στο Standard deviation έβαλα την τιμή ίση με 9 και στο random seed έβαλα την τιμή ίση με 91. Μετέπειτα, δημιουργήθηκαν δύο μεταβλητές με 291 τιμές στο excel και καταχώρησα την πρώτη στήλη ως Χ1 και την δεύτερη στήλη ως Χ2. Αυτή είναι η διαδικασία που ακολούθησα για να δημιουργήσω τα δεδομένα σύμφωνα με τον αριθμό μητρώου μου.

Ο τύπος της μέσης τιμής είναι ο εξής:

Ο τύπος της τυπικής απόκλισης είναι ο εξής:

# Επιθυμητή Επεξεργασία

## 1.

Για να εκτιμήσω τα παρακάτω περιγραφικά στατιστικά μέτρα χρησιμοποίησα το Excel. Για τον συντελεστή μεταβλητότητας όπου ο τύπος του είναι τυπική απόκλιση διά τη μέση τιμή έκανα τον υπολογισμό της μέσης τιμής και της τυπικής απόκλισης για τις δύο στήλες. Οι τιμές που βρήκα για τις μέσες τιμές της πρώτης στήλης είναι -0.12009163 και για την δεύτερη στήλη είναι 1.627195. Η τυπική απόκλιση για την πρώτη στήλη είναι 9.064017026 και για την δεύτερη στήλη είναι 9.178449.

Έτσι, υπολογίζοντας τον συντελεστή μεταβλητότητας για τις δύο στήλες είναι -75.4758406 για την πρώτη στήλη και 5.640658 για την δεύτερη στήλη.

Ο τύπος του συντελεστή μεταβλητότητας δίνεται από την διαίρεση της τυπικής απόκλισης:

Ο υπολογισμός της μέσης τιμές στο Excel έγινε με τη συνάρτηση AVERAGE και ο υπολογισμός της τυπικής απόκλισης έγινε με τη συνάρτηση STDEV.S

Όσον αφορά την διάμεσο υπολογίστηκε στο Excel με τη συνάρτηση MEDIAN. Για τη πρώτη μεταβλητή Χ1 η διάμεσος ισούται με -0.02079753 και για την δεύτερη μεταβλητή Χ2 η διάμεσος ισούται με 1.569411. Οι υπολογισμοί έγιναν με την συνάρτηση MEDIAN.

Η διάμεσος βρίσκεται πολύ εύκολα αν υπολογίσουμε και βάλουμε κατά αύξουσα σειρά τις τιμές την μεταβλητής και επιλέξουμε την κεντρική σε περίπτωση που το πλήθος είναι άρτιος αριθμός, αν είναι περιττός αριθμός επιλέγουμε την μέση τιμή των δύο κεντρικών τιμών.

Για τον υπολογισμό της ασυμμετρίας στο σύνολο δεδομένων μας για τις μεταβλητές Χ1 και Χ2 υπολογίσθηκαν skewness και kurtosis για κάθε μια από τις μεταβλητές Χ1 και Χ2. Το skewness για την μεταβλητή Χ1 ισούται με 0.079192296 και για το Χ2 ισούται με -0.27577. Το kurtosis ισούται με 0.04865061 για την πρώτη μεταβλητή και ίσο με -0.1093 για την δεύτερη μεταβλητή.

Για το ενδοτεταρτημοριακό εύρος IQR χρησιμοποίησα την συνάρτηση QUARTILE.INC η οποία παίρνει δύο ορίσματα στο Excel. Το πρώτο όρισμα είναι οι τιμές που δίνονται για την μεταβλητή Χ1 και Χ2 και το δεύτερο όρισμα παίρνει τιμές από 0 έως 3 σύμφωνα με το τεταρτημόριο του εύρος. Για την μεταβλητή Χ1 και ως όρισμα 0 παίρνει τιμή -25.5005656 ενώ για την μεταβλητή Χ2 παίρνει τιμή -25.0605. Για την μεταβλητή Χ1 και ως όρισμα 1, παίρνει τιμή -6.38437791 και για την μεταβλητή Χ2 παίρνει τιμή -4.5915. Για την μεταβλητή Χ1 και όρισμα 2 παίρνει τιμή 0.019382813 και για την μεταβλητή Χ2 και με όρισμα 2 παίρνει τιμή 1.569411. Για την μεταβλητή Χ1 και όρισμα 3 παίρνει τιμή 5.182766133 και για την μεταβλητή Χ2 παίρνει τιμη΄8.247379.

Το ενδοτεταρτημοριακό εύρος IQR βρίσκεται ως η μικρότερη τιμή γι’αυτό παίρνει το όρισμα της τιμή μηδέν. Μετέπειτα, υπολογίζεται ως το 25% των τιμών γι’αυτό παίρνει η συνάρτηση την τιμή 1, το 75% γι’αυτό το όρισμα παίρνει την τιμή 2 και τέλος ως τιμή 3 παίρνει τη μέγιστη τιμή του εύρος των τεταρτημορίων.

## 2

### Α

Δοθέντος ότι το δεύτερο δείγμα είναι πληθυσμός και όχι δείγμα τότε υπολογίστηκαν οι εξής πιθανότητες χρησιμοποιώντας τις αντίστοιχες συναρτήσεις στο Excel.

Για το α ερώτημα, υλοποιήθηκε η συνάρτηση NORM.S.DIST με δύο ορίσματα όπου το πρώτο όρισμα παίρνει τιμή την πρώτη τιμή από την μεταβλητή Χ2 και στο όρισμα cumulative το ορίσαμε να είναι ως TRUE. Η τιμή που πήραμε είναι ίση με 3.62635Ε-30.

### Β

Για το δεύτερο ερώτημα υλοποιήθηκε η συνάρτηση NORM.DIST η οποία συνάρτηση έχει 4 ορίσματα όπου το πρώτο όρισμα είναι η δεύτερη τιμή του δεύτερου δείγματος. Το δεύτερο όρισμα είναι η μέση τιμή του δείγματος θεωρούμενο ότι είναι πληθυσμός. Το τρίτο όρισμα είναι η τυπική απόκλιση και το τέταρτο όρισμα είναι το Cumulative όπου ορίστηκε ως TRUE.

### Γ

Για το τρίτο ερώτημα αυτής της εργασίας όπου έχει να κάνει με την πιθανότητα ανάμεσα σε δύο τιμές, υπολογίστηκε πρώτα η πιθανότητα με τη συνάρτηση NORM.DI με ορίσματα την τιμή όπως αναφέρει η άσκηση την Τρίτη τιμή του δεύτερου δείγματος και την τέταρτη τιμή του δεύτερου δείγματος.

Υπολογίστηκαν δύο πιθανότητες για το συγκεκριμένο ερώτημα. Πρώτα, υπολογίστηκε η πιθανότητα η μεταβλητή Χ να είναι μεγαλύτερη από την Τρίτη τιμή του δεύτερου δείγματος η οποία επέστρεψε τιμή ίση με 0.246881883. Δεύτερον, υπολογίστηκε η πιθανότητα η μεταβλητή Χ να είναι μικρότερη από την τέταρτη τιμή του δεύτερου δείγματος η οποία έδωσε τιμή ίση με 0.834804. Εν τέλη, η αφαίρεση αυτών των δύο τιμών μας δίνει την απάντηση για την εξής πιθανότητα και η τιμή της ισούται με 0.587921776.

## 3

Για το ν υπολογισμό του διαστήματος εμπιστοσύνης του πρώτου δείγματος με διάστημα εμπιστοσύνης 92% έγινε χρήση με το Excel. Τα βήματα που ακολουθήθηκαν είναι Data Analysis και από την λίστα έγινε επιλογή Descriptive statistics. Μετέπειτα, στο Input Range έβαλα όλες τιμές του πρώτου δείγματος και πάτησα τικ στο Summary Statistics και Confidence level for mean ισούται με 92% γιατί αυτό μου ζητάει η άσκηση. Ο εξής πίνακας δημιουργήθηκε

|  |  |
| --- | --- |
| *Column1* | |
|  |  |
| Mean | -0.12009 |
| Standard Error | 0.531342 |
| Median | -0.0208 |
| Mode | 2.90013 |
| Standard Deviation | 9.064017 |
| Sample Variance | 82.1564 |
| Kurtosis | 0.048651 |
| Skewness | 0.079603 |
| Range | 50.97932 |
| Minimum | -25.5006 |
| Maximum | 25.47876 |
| Sum | -34.9467 |
| Count | 291 |
| Confidence Level(92.0%) | 0.933484 |

Όπου δείχνει την μέση τιμή ίση με -0.12009, το τυπικό σφάλμα ίσο με 0.531342, τη διάμεσος ίση με -0.0208 και το mode ίσο με 2.90013. Η τυπική απόκλιση ίση με 9.06401. Όσον αφορά την ασυμμετρία του δείγματος η kurtosis ισούται με 0.048651 και το skewness ισούται με 0.079603. Το εύρος των τιμών ισούται με 50.97932 με ελάχιστη τιμή ίση με -25.5006 και μεγαλύτερη τιμή 25.47876. Το άθροισμα ισούται με -34.9467 και το πλήθος των τιμών ισούται με 291. Το διάστημα εμπιστοσύνης όπως ορίστηκε στο Excel είναι ίσο με 92% διαφορετικά ίσο με 0.933484.

Η εξίσωση της συνάρτησης skewness είναι η εξής:

Η εξίσωση της συνάρτησης kurtosis δίνεται από τον εξής τύπο:

## 4

Για τον υπολογισμό του διαστήματος εμπιστοσύνης της διαφοράς μ1-μ2 σε επίπεδο εμπιστοσύνης 96% υπολογίστηκαν τρία διαφορετικά στατιστικά μοντέλα. Το πρώτο στατιστικό μοντέλο που υλοποιήθηκε κάνοντας τα εξής βήματα στο Excel είναι Data Analysis και μετά από τη λίστα επιλέχθηκε η επιλογή t-test two sample assuming equal variances. Οι τιμές που δόθηκαν αναπαρίστανται στον εξής πίνακα

|  |  |  |
| --- | --- | --- |
|  | *Variable 1* | *Variable 2* |
| Mean | -0.12009 | 1.627195 |
| Variance | 82.1564 | 84.24393 |
| Observations | 291 | 291 |
| Pooled Variance | 83.20017 |  |
| Hypothesized Mean Difference | 0 |  |
| df | 580 |  |
| t Stat | -2.31065 |  |
| P(T<=t) one-tail | 0.010601 |  |
| t Critical one-tail | 1.753759 |  |
| P(T<=t) two-tail | 0.021202 |  |
| t Critical two-tail | 2.058378 |  |

Μετέπειτα, για την πλήρη εικόνα των στατιστικών μοντέλων έγινε υλοποίηση του στατιστικού μοντέλου t-test two sample assuming unequal variances και οι τιμές που δίνονται φαίνονται στον εξής πίνακα. Η επιλογή αυτή έγινε στο Excel με το Data Analysis μετέπειτα t-test sample assuming unequal variances. Στο variable 1 επιλέχθηκαν οι τιμές της μεταβλητής Χ1, στο variable 2 επιλέχθηκαν οι τιμές Χ2. Στην τιμή alpha καταχωρήθηκε η τιμή 0.04.

|  |  |  |
| --- | --- | --- |
|  | *Variable 1* | *Variable 2* |
| Mean | -0.12009 | 1.627195 |
| Variance | 82.1564 | 84.24393 |
| Observations | 291 | 291 |
| Hypothesized Mean Difference | 0 |  |
| df | 580 |  |
| t Stat | -2.31065 |  |
| P(T<=t) one-tail | 0.010601 |  |
| t Critical one-tail | 1.753759 |  |
| P(T<=t) two-tail | 0.021202 |  |
| t Critical two-tail | 2.058378 |  |

Στη συνέχεια το τρίτο και τελευταίο στατιστικό μοντέλο που υλοποιήθηκε είναι το z-test Two sample for means όπου οι τιμές αναπαρίστανται στον εξής πίνακα. Η υλοποίηση του z-test έγινε με Data Analysis, μετέπειτα επιλέχθηκε η επιλογή z-test two sample for means. Σε αυτό το στατιστικό μοντέλο επιπρόσθετα επιλέχθηκαν το πλήθος των δειγμάτων Χ1 και Χ2 στο Variable 1 Variance, Variable 2 Variance και στο alpha επιλέχθηκε το 0.04.

|  |  |  |
| --- | --- | --- |
|  | *Variable 1* | *Variable 2* |
| Mean | -0.12009 | 1.627195 |
| Known Variance | 291 | 291 |
| Observations | 291 | 291 |
| Hypothesized Mean Difference | 0 |  |
| z | -1.23552 |  |
| P(Z<=z) one-tail | 0.108319 |  |
| z Critical one-tail | 1.750686 |  |
| P(Z<=z) two-tail | 0.216638 |  |
| z Critical two-tail | 2.053749 |  |

## 5

Για τον έλεγχο της υπόθεσης της Ηο έναντι της εναλλακτικής με μέση διαφορά το -1 έγινε μια αλλαγή από μ1-μ2 σε μ2-μ1 για να πάρει την τιμή 1 καθώς στα στατιστικά μοντέλα ήταν αδύνατον να τοποθετηθεί αρνητική τιμή.

Έτσι, συνεχίζοντας τον έλεγχο των υποθέσεων έγινε χρήση στο Excel, Data Analysis έγινε υλοποίηση το στατιστικό μοντέλο t-test: two sample assuming equal variances με τη μόνη διαφορά ότι συμπληρώθηκε η τιμή 1 στο hypothesized mean και έγινε μετάταξη των variable 1 και variable 2 δηλαδή έγινε αντιστροφή των δύο μεταβλητών για να υπολογισθεί η μηδενική υπόθεση Η0 έναντι της εναλλακτικής Η1 με μέση διαφορά το 1 και διάστημα εμπιστοσύνης alpha ίσο με 0.05

Για το στατιστικό μοντέλο t test Two sample assuming Equal Variances μου δόθηκε ο εξής πίνακας.

|  |  |  |
| --- | --- | --- |
|  | *Variable 1* | *Variable 2* |
| Mean | 1.627195 | -0.12009 |
| Variance | 84.24393 | 82.1564 |
| Observations | 291 | 291 |
| Pooled Variance | 83.20017 |  |
| Hypothesized Mean Difference | 1 |  |
| df | 580 |  |
| t Stat | 0.988227 |  |
| P(T<=t) one-tail | 0.161727 |  |
| t Critical one-tail | 1.647485 |  |
| P(T<=t) two-tail | 0.323454 |  |
| t Critical two-tail | 1.964063 |  |

Συνεχίζοντας, για το στατιστικό μοντέλο t-test Two sample assuming Unequal Variances μου δόθηκε ο εξής πίνακας με hypothesized mean ίση με 1 και διάστημα εμπιστοσύνης ίσο με 0.05.

|  |  |  |
| --- | --- | --- |
|  | *Variable 1* | *Variable 2* |
| Mean | 1.627195 | -0.12009 |
| Variance | 84.24393 | 82.1564 |
| Observations | 291 | 291 |
| Hypothesized Mean Difference | 1 |  |
| df | 580 |  |
| t Stat | 0.988227 |  |
| P(T<=t) one-tail | 0.161727 |  |
| t Critical one-tail | 1.647485 |  |
| P(T<=t) two-tail | 0.323454 |  |
| t Critical two-tail | 1.964063 |  |

Εν τέλη, για το στατιστικό μοντέλο z-test two sample for means με hypothesized mean ίση με 1 και διάστημα εμπιστοσύνης ίσο με 0.05, μου δόθηκε ο εξής πίνακας αποτελεσμάτων

|  |  |  |
| --- | --- | --- |
|  | *Variable 1* | *Variable 2* |
| Mean | 1.627195 | -0.12009 |
| Known Variance | 291 | 291 |
| Observations | 291 | 291 |
| Hypothesized Mean Difference | 1 |  |
| z | 0.528411 |  |
| P(Z<=z) one-tail | 0.298607 |  |
| z Critical one-tail | 1.644854 |  |
| P(Z<=z) two-tail | 0.597214 |  |
| z Critical two-tail | 1.959964 |  |

# Μέρος Β

## Θέμα 1

Οι παρατηρήσεις οι οποίες είναι υπό μελέτη είναι στο σύνολο 10. Το α ερώτημα της άσκησης μας ρωτάει να βρούμε το μοντέλο γραμμικής παλινδρόμησης που συνδέει το κόστος με το κέρδος της εκμετάλλευσης και να ελέγξουμε τη σημαντικότητα του. Έτσι, λοιπόν κάνοντας Data Analysis, regression για να βρούμε το γραμμικό μοντέλο προκύπτει ο εξής πίνακας.

|  |  |
| --- | --- |
| SUMMARY OUTPUT | |
|  |  |
| *Regression Statistics* | |
| Multiple R | 0.092512 |
| R Square | 0.008559 |
| Adjusted R Square | -0.11537 |
| Standard Error | 1.597838 |
| Observations | 10 |

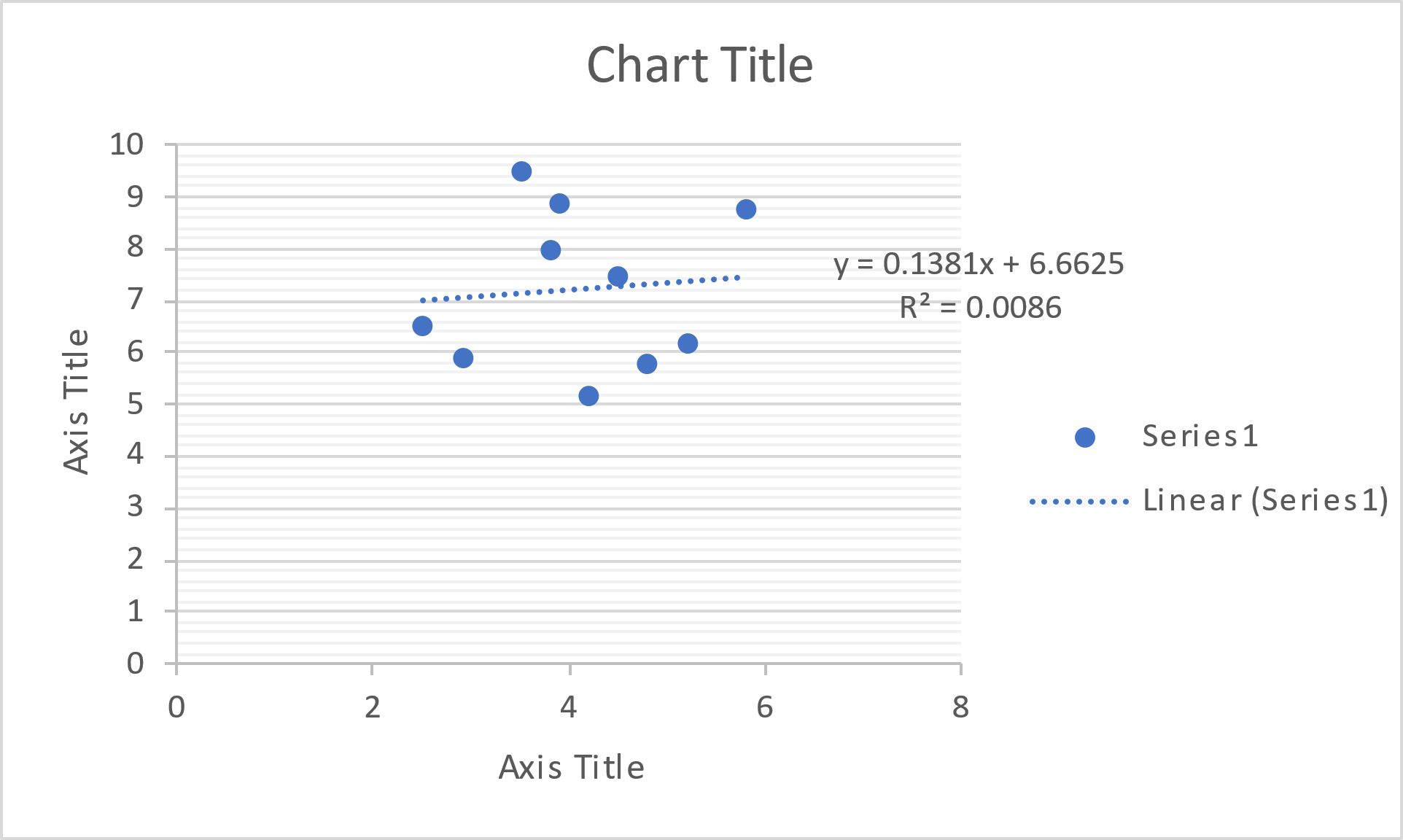
Σε αυτόν τον πίνακα ο πολλαπλασιαστής R ισούται ίσιος με 0.092513. Το τετράγωνο R ισούται με 0.008559 ενώ το τυπικό σφάλμα ισούται με 1.597838. Οι παρατηρήσεις είναι στο σύνολο 10.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ANOVA |  |  |  |  |  |
|  | *df* | *SS* | *MS* | *F* | *Significance F* |
| Regression | 1 | 0.176314 | 0.176314 | 0.069059 | 0.799353 |
| Residual | 8 | 20.42469 | 2.553086 |  |  |
| Total | 9 | 20.601 |  |  |  |

Από τον πίνακα ΑΝΟΒΑ παρατηρούμε ότι οι βαθμοί ελευθερίας για το μοντέλο παλινδρόμησης μας είναι 1. Το SS ισούται με 0.176314, το στατιστικό μοντέλο F ισούται με 0.069059. Το σφάλμα έχει 8 βαθμούς ελευθερίας, ενώ το SS ισούται με 20.42469 και το MS ισούται με 2.553086.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* | *Lower 95.0%* | *Upper 95.0%* |
| Intercept | 6.662536 | 2.217699 | 3.004256 | 0.016961 | 1.548513 | 11.77656 | 1.548513 | 11.77656 |
| X Variable 1 | 0.138069 | 0.525394 | 0.262791 | 0.799353 | -1.07349 | 1.34963 | -1.07349 | 1.34963 |

Στον παραπάνω πίνακα είναι σημαντικό να λάβουμε υπόψιν τις σταθερές του μοντέλου παλινδρόμησης όπου ο συντελεστής του x ισούται να είναι 6.662536 και ο σταθερός όρος ισούται με 0.138069. Το μοντέλο εξελίσσεται όταν παριστάνεται στο εξής διάγραμμα.



Το β ερώτημα της άσκησης μας ζητάει να βρούμε τον συντελεστή συσχέτισης του Pearson. Ο συντελεστής συσχέτισης του Pearson βρίσκεται πολύ εύκολα από την λίστα του Data Analysis στο Excel.

Τα αποτελέσματα του συντελεστή συσχέτισης εμφανίζονται στον παρακάτω πίνακα.

|  |  |
| --- | --- |
| Pearson |  |
| 0.09 | 0.09 |
| no linear correclation | |
| A value of 0 denotes no linear correlation | |

Όπου παρατηρούμε ότι είναι πολύ κοντά στο μηδέν όπου αυτό σημαίνει ότι δεν υπάρχει γραμμική συσχέτιση.

## Θέμα 2

Για το θέμα 2 παρατηρούμε τον εξής πίνακα στο Excel.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Μέγεθος επιχείρησης |  |  |
| Είδος περιοχής | Μικρού μεγέθους | Μεγάλου μεγέθους |  |
|  | (< 10 κλίνες) | (> 10 κλίνες) |  |
| Πεδινή | 15 | 52 | 67 |
| Ορεινή | 18 | 8 | 26 |
| Ημιορεινή | 13 | 14 | 27 |
|  | 46 | 74 | 120 |
|  |  |  |  |
|  | Μέγεθος επιχείρησης |  |  |
| Είδος περιοχής | Μικρού μεγέθους | Μεγάλου μεγέθους |  |
|  | (< 10 κλίνες) | (> 10 κλίνες) |  |
| Πεδινή | 25.7 | 41.3 |  |
| Ορεινή | 10.0 | 16.0 |  |
| Ημιορεινή | 10.4 | 16.7 |  |
|  |  |  |  |
|  |  |  |  |
| Εύρεση της σημαντικότητας p για την ανεξαρτησία | | |  |
| ΣΥΝΑΡΤΗΣΗ : | =CHITEST(ΠΑΡΑΤΗΡΗΣΕΙΣ;ΑΝΑΜΕΝΟΜΕΝΕΣ) | | |
|  | = CHITEST(B4:C6;B13:C15) | |  |
| P-ΤΙΜΗ= | 0,000082 |  |  |
|  |  |  |  |
|  | 0.0000825 |  |  |

Με επίπεδο σημαντικότητας 5% παρατηρούμε ότι η p-value ισούται με 0.000082 όπου είναι μικρότερη τιμή από το 0.05 συνεπώς υπάρχει στατιστικώς σημαντική διαφρορά μεταξύ το μέγεθος των αγροτουριστικών μονάδων από το είδος περιοχής άρα αυτό σημαίνει ότι επηρεάζεται από την ίδια την περιοχή.