Πανεπιστήμιο Δυτικής Μακεδονίας

Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης

Κυριακή Πασχαλίδου

Αρ. Μητρώου:1547-291

Ιούνιος 2023

Περιεχόμενα

[Πίνακες 3](#_Toc138612162)

[Εικόνες 4](#_Toc138612163)

[Ερώτημα 1ο 5](#_Toc138612164)

[Ερώτημα 2ο 11](#_Toc138612165)

[Ερώτημα 3ο 13](#_Toc138612166)

[Ερώτημα 4ο 15](#_Toc138612167)

[Ερώτημα 5ο 16](#_Toc138612168)

[Ερώτημα 6ο 22](#_Toc138612169)

[Ερώτημα 7ο 23](#_Toc138612170)

# Πίνακες

[Table 1Πίνακας συσχετίσεων των μεταβλητών, Pearson correlation 5](#_Toc138612179)

[Table 2Bayesian Pearson Correlation 8](#_Toc138612180)

[Table 3Πίνακας διακυμάνσεων και συνδιακυμάνσεων μεταξύ των μεταβλητών 10](#_Toc138612181)

[Table 4Περίληψη του μοντέλου- sales 11](#_Toc138612182)

[Table 5Linear regression one variable - ANOVA 11](#_Toc138612183)

[Table 6Coefficients linear regression ove independent variable 12](#_Toc138612184)

[Table 7Model Summary - Sales 13](#_Toc138612185)

[Table 8ANOVA model H1 13](#_Toc138612186)

[Table 9Coefficients multivariable linear regression model 14](#_Toc138612187)

[Table 10Cronbach's reliability test q1-q16 15](#_Toc138612188)

[Table 11Chi-squared test 16](#_Toc138612189)

[Table 12Factor Loadings 16](#_Toc138612190)

[Table 13Factor Loadings (Structure Matrix) 17](#_Toc138612191)

[Table 14Factor Characteristics 18](#_Toc138612192)

[Table 15Factor correlations between the three factors 19](#_Toc138612193)

[Table 16Additional fit indices 19](#_Toc138612194)

[Table 17Residual matrix 19](#_Toc138612195)

[Table 18ANOVA age grouping 22](#_Toc138612196)

[Table 19Chi-squared test PCA Analysis 24](#_Toc138612197)

[Table 20Component Loadings PCA 24](#_Toc138612198)

[Table 21Component Characteristics - PCA Analysis 25](#_Toc138612199)

# Εικόνες

[Figure 1Pearson correlation οπτικοποιημένα 6](#_Toc138612200)

[Figure 2Pearson'correlation heat map 7](#_Toc138612201)

[Figure 3Correlation matrix 9](#_Toc138612202)

# Ερώτημα 1ο

Στόχος του πρώτου ερωτήματος είναι ο υπολογισμός των συσχετίσεων μεταξύ του μισθού διευθύνοντος συμβούλου (salary), την ηλικία του, τα έτη που εργάζεται στην εταιρεία, τα έτη που κατέχει τη θέση του διευθύνοντος συμβούλου στην εταιρεία, τις πωλήσεις και τα κέρδη. Μετέπειτα είναι η απεικόνιση των αντίστοιχων διαγραμμάτων για να οπτικοποιήσουμε τις συσχετίσεις αυτές και να κατασκευάσουμε τον πίνακα διακυμάνσεων-συνδιακυμάνσεων για τις παραπάνω μεταβλητές. Έτσι, λοιπόν, αρχίζοντας με τις συσχετίσεις όπως απεικονίζονται στον Table 1. Συγκρίνουμε τις συσχετίσεις μεταξύ των διαφορετικών μεταβλητών και όχι με τις ίδιες μεταβλητές με τον Pearson correlation. Η συσχέτιση μεταξύ της ηλικίας και του μισθού είναι θετική και μάλιστα ισούται με 0.115. Η συσχέτιση μεταξύ της μεταβλητής comten και του μισθού είναι θετική και μάλιστα 0.038. Αυτή τη συσχέτιση όπως λέμε στα μαθηματικά επειδή τείνει να ισούται στο μηδέν θα την λέγαμε ουδέτερη, ούτε αρνητική ή θετική. Η συσχέτιση μεταξύ της μεταβλητής ceoten και μισθού ισούτι με 0.143 όπου είναι θετική η συσχέτιση. Η συσχέτιση μεταξύ των πωλήσεων και του μισθού είναι θετική και ίση με 0.380. Η συσχέτιση των κερδών και του μισθού ισούται με 0.394 και είναι θετική. Η συσχέτιση μεταξύ της ηλικίας και της μεταβλητής comten ισούται με 0.479 και είναι θετική. Η συσχέτιση μεταξύ της ηλικίας και της μεταβλητής ceoten είναι θετική και ίση με 0.339. Η συσχέτιση μεταξύ ηλικίας και πωλήσεων είναι θετική και ίση με 0.127. Η συσχέτιση μεταξύ της ηλικίας και των κερδών είναι θετική και ίση με 0.115. Η συσχέτιση της μεταβλητής comten και ceoten είναι θετική και ίση με 0.315. Η συσχέτιση μεταξύ των μεταβλητών comten και πωλήσεων είναι θετική και ίση 0.104. Η συσχέτιση μεταξύ των μεταβλητών comten και κερδών είναι θετική και ίση με 0.144. Η συσχέτιση μεταξύ των μεταβλητών ceoten και πωλήσεων είναι αρνητική και ίση με -0.068. Η συσχέτιση μεταξύ των μεταβλητών ceoten και κερδών είναι αρνητική και ίση με -0.022. Η συσχέτιση μεταξύ των μεταβλητών sales και profits είναι θετική και μάλιστα άρρηκτα συσχετισμένες μιας και που ισούται με 0.798. Δηλαδή έχουν υψηλό δείκτη συσχέτισης.

Table 1Πίνακας συσχετίσεων των μεταβλητών, Pearson correlation

| **Pearson's Correlations** | | | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | |  | | **salary** | | **age** | | **comten** | | **ceoten** | | **sales** | | **profits** | |
| 1. salary |  | Pearson's r |  | — |  |  |  |  |  |  |  |  |  |  |  |
|  |  | p-value |  | — |  |  |  |  |  |  |  |  |  |  |  |
| 2. age |  | Pearson's r |  | 0.115 |  | — |  |  |  |  |  |  |  |  |  |
|  |  | p-value |  | 0.126 |  | — |  |  |  |  |  |  |  |  |  |
| 3. comten |  | Pearson's r |  | 0.038 |  | 0.479 |  | — |  |  |  |  |  |  |  |
|  |  | p-value |  | 0.618 |  | < .001 |  | — |  |  |  |  |  |  |  |
| 4. ceoten |  | Pearson's r |  | 0.143 |  | 0.339 |  | 0.315 |  | — |  |  |  |  |  |
|  |  | p-value |  | 0.058 |  | < .001 |  | < .001 |  | — |  |  |  |  |  |
| 5. sales |  | Pearson's r |  | 0.380 |  | 0.127 |  | 0.104 |  | -0.068 |  | — |  |  |  |
|  |  | p-value |  | < .001 |  | 0.092 |  | 0.167 |  | 0.371 |  | — |  |  |  |
| 6. profits |  | Pearson's r |  | 0.394 |  | 0.115 |  | 0.144 |  | -0.022 |  | 0.798 |  | — |  |
|  |  | p-value |  | < .001 |  | 0.128 |  | 0.056 |  | 0.775 |  | < .001 |  | — |  |
|  | | | | | | | | | | | | | | | |

Για να οπτικοποιήσουμε τα παραπάνω αποτελέσματα κάναμε διαγράμαμτα όπως φαίνονται στην Figure 1. Όπως πολύ σωστά παρατηρούμε αξίζει να σημειωθεί ότι η μεταβλητή coeten και οι μεταβλητές sales και profits είναι αρνητικά συσχετισμένες και αυτό το καταλαβαίνουμε από την ευθεία που απεικονίζεται να είναι γραμμική με αρνητική κλίση. Όλες οι άλλες γραμμές είναι ευθείες με θετική κλίση.

**Correlation plot**

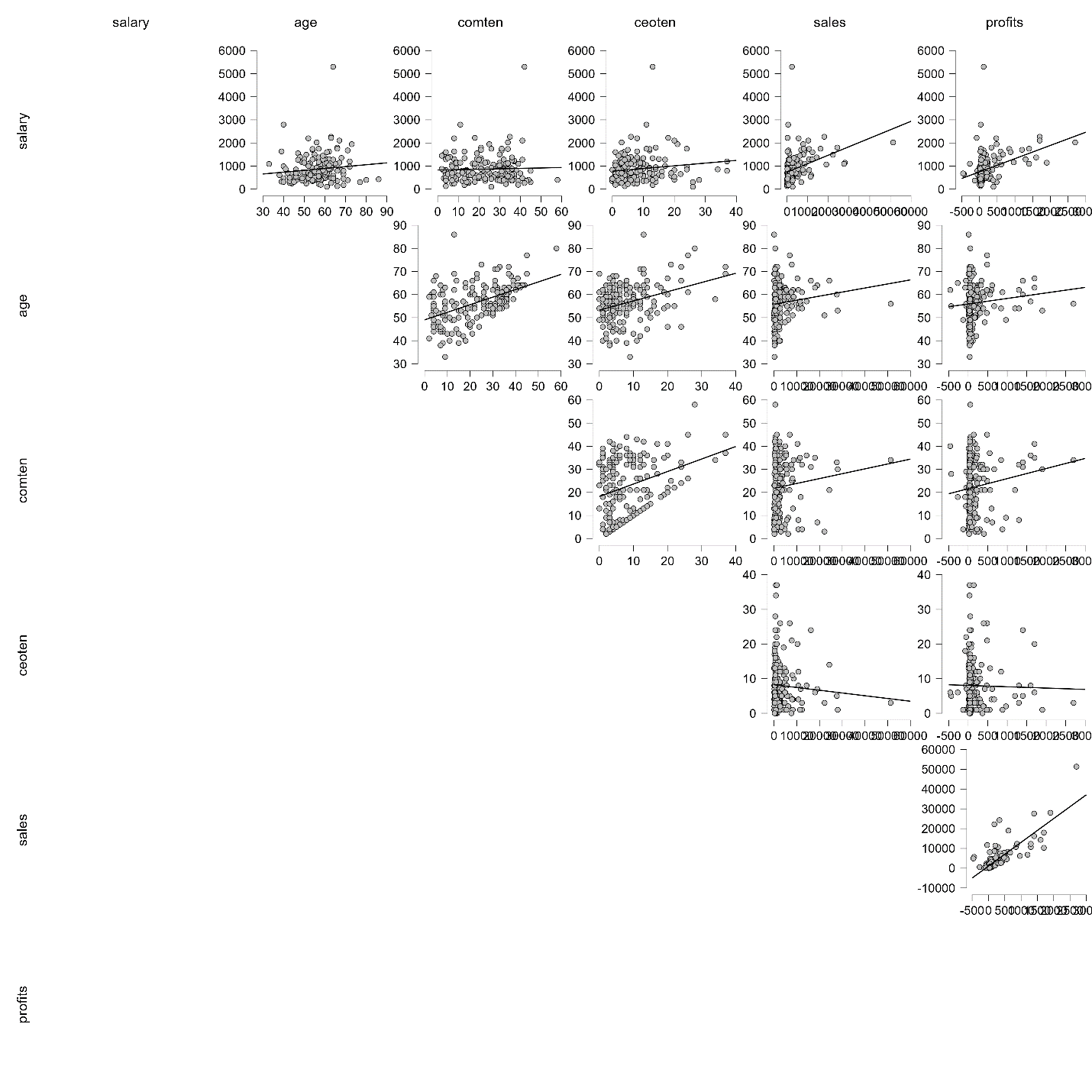


Figure Pearson correlation οπτικοποιημένα

Στην Figure 2 απεικονίζεται heatmap σύμφωνα με τις τιμές που λάβαμε από τον pearson correlation. Όσο πιο σκούρο μπλε απεικονίζεται στον χάρτη τόσο θετικότερη είναι συσχέτιση μεταξύ των μεταβλητών, ενώ όσο πιο μπεζ τόσο πιο αρνητική εμφανίζεται η συσχέτιση μεταξύ των μεταβλητών.

**Pearson's r heatmap**

A picture containing text, screenshot, font, number

Description automatically generated

Figure Pearson'correlation heat map

Στον

Table 2 απεικονίζονται τα αποτελέσματα από το Bayesian Pearson Correlation ντετερμινιστικά όπως λέμε στην στατιστική και όχι πιθανοθεωρητικά. Δηλαδή, τα δεδομένα μας ακολουθούν κάποια κατανομή. Μεταξύ των μεταβλητών salary και age Ο δείκτης συσχέτισης Pearson ισούται με 0.115 είναι θετικός. Ο παράγοντας Bayes ισούται με 0.300. Μεταξύ των μεταβλητών salary και comten η συσχέτιση pearson ισούται με 0.038 θετική και πολύ κοντά στο μηδέν. Ο παράγοντας bayes ισούται με 0.106. Η συσχέτιση μεταξύ των μεταβλητών salary και coeten ισούται με 0.143 ενώ ο παράγοντας Bayes ισούται με 0.561. Η συσχέτιση της μεταβλητής salary και sales ισούται με 0.380 ενώ ο παράγοντας bayes ισούται με 69357.696. Η συσχέτιση μεταξύ των μεταβλητών salary και profits ισούται με 0.394 είναι μια θετική συσχέτιση, ενώ ο παράγοντας bayes ισούται με 204178.682. Όσον αφορά τη μεταβλητή age και comten ο συντελεστής συσχέτισης ισούται με 0.479 ενώ ο παράγοντας bayes ισούται με 6.118 επί 10 εις την ογδόη. Η συσχέτιση μεταξύ των μεταβλητών age και coeten ισούται με 0.339 ενώ ο παράγοντας bayes ισούται με 3575.529. Για τις μεταβλητές age και sales ο συντελεστής συσχέτισης είναι 0.127 ενώ ο παράγοντας bayes ισούται με 0.385. Για τις μεταβλητές age και profits ο συντελεστής συσχέτισης ισούται με 0.115ενώ ο παράγοντας bayes ισούται με 0.296. Για τις μεταβλητές comten και ceoten ο συντελεστής συσχέτισης Pearson ισούται με 0.315, ενώ ο παράγοντας bayes ισούται με 800.609. Για τις μεταβλητές comten και sales ο συντελεστής συσχέτισης Pearson ισούται με 0.104 ενώ ο παράγοντας bayes ισούται με 0.243. Για τις μεταβλητές comten και profit ο συντελεστής συσχέτισης ισούται με 0.144 ενώ ο παράγοντας bayes ισούται με 0.573. Για τις μεταβλητές coeten και sales ο συντελεστής συσχέτισης Pearson Ισούται με -0.068,ενώ ο παράγοντας bayes ισούται με 0.140. Για τις μεταβλητές coeten και profits ο συντελεστής συσχέτισης ισούται με -0.022 ενώ ο παράγοντας bayes ισούται με 0.098. Για τις μεταβλητές sales και profits ο συντελεστής συσχέτισης Pearson είναι θετικός άρα θετική συσχέτιση και ίσος με 0.798 ενώ ο παράγοντας bayes ισούται με 1.191x10εις την 37.

Table 2Bayesian Pearson Correlation

| **Bayesian Pearson Correlations** | | | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | |  | | **salary** | | **age** | | **comten** | | **ceoten** | | **sales** | | **profits** | |
| 1. salary |  | Pearson's r |  | — |  |  |  |  |  |  |  |  |  |  |  |
|  |  | BF₁₀ |  | — |  |  |  |  |  |  |  |  |  |  |  |
| 2. age |  | Pearson's r |  | 0.115 |  | — |  |  |  |  |  |  |  |  |  |
|  |  | BF₁₀ |  | 0.300 |  | — |  |  |  |  |  |  |  |  |  |
| 3. comten |  | Pearson's r |  | 0.038 |  | 0.479 |  | — |  |  |  |  |  |  |  |
|  |  | BF₁₀ |  | 0.106 |  | 6.118×10+8 |  | — |  |  |  |  |  |  |  |
| 4. ceoten |  | Pearson's r |  | 0.143 |  | 0.339 |  | 0.315 |  | — |  |  |  |  |  |
|  |  | BF₁₀ |  | 0.561 |  | 3575.529 |  | 800.609 |  | — |  |  |  |  |  |
| 5. sales |  | Pearson's r |  | 0.380 |  | 0.127 |  | 0.104 |  | -0.068 |  | — |  |  |  |
|  |  | BF₁₀ |  | 69357.696 |  | 0.385 |  | 0.243 |  | 0.140 |  | — |  |  |  |
| 6. profits |  | Pearson's r |  | 0.394 |  | 0.115 |  | 0.144 |  | -0.022 |  | 0.798 |  | — |  |
|  |  | BF₁₀ |  | 204178.682 |  | 0.296 |  | 0.573 |  | 0.098 |  | 1.191×10+37 |  | — |  |
|  | | | | | | | | | | | | | | | |

Στην Figure 3 απεικονίζεται ο πίνακας συσχέτισης. Παρατηρούμε πολύ ορθά ότι όλες οι συσχετίσεις είναι θετικές εκτός από αυτές των μεταβλητών coeten και sales, coeten και profits, όπου είναι αρνητικές σύμφωνα με την γραμμική καμπύλη, ευθεία που απεικονίζεται.

**Bayesian Correlation Matrix Plot**

A picture containing text, diagram, map

Description automatically generated

Figure 3Correlation matrix

Στον Table 3παρουσιάζονται τα αποτελέσματα του πίνακα διακυμάνσεων και συνδιακυμάνσεων με την ανάλυση ANOVA που πραγματοποιήθηκε. Τα τετραγωνικά αθροίσματα είναι 1.288x106 με 1 βαθμό ελευθερίας και μέσο τετραγωνικό σφάλμα να ισούται με 1.288x106. Το στατικό F-test ισούται με 4.533 και η p-value ισούται με 0.035. Την τιμή p-value την συγκρίνουμε με το 0.05 για να ορίσουμε αν υπάρχει στατιστικά σημαντική διαφορά στα δεδομένα μας. Δηλαδή, το διάστημα εμπιστοσύνης που ορίσαμε είναι 95%. Αν η τιμή p-value είναι μικρότερη του 0.05 τότε σημαίνει ότι για τα δεδομένα μας υπάρχει στατιστικώς σημαντική διαφορά και απορρίπτεται η μηδενική υπόθεση. Αν η p-value είναι μεγαλύτερη από το 0.05 τότε σημαίνει ότι δεν υπάρχει στατιστικώς σημαντική διαφορά στα δεδομένα μας και το στατιστικό τεστ και η μηδενική υπόθεση είναι αποδεκτά. Συνεχίζοντας, για την μεταβλητή comten το τετραγωνικό άθροισμα ισούται με 119644.864, οι βαθμοί ελευθερίας είναι ένας, το μέσο τετράγωνο ισούται με 119644.864, το στατιστικό F ισούται με 0.421 και η p-value ισούται με 0.517. Για την μεταβλητή ceoten το τετραγωνικό άθροισμα ισούται με 1.820x106, οι βαθμοί ελευθερίας ισούται με 1 , το μέσο τετράγωνο ισούται με 1.820x106, το στατιστικό F ισούται με 6.405 και η τιμή p-value ισούται με 0.012. Για την μεταβλητή sales το τετραγωνικό άθροισμα ισούται με 931558.385, οι βαθμοί ελευθερίας ισούται με 1, το μέσο τετράγωνο ισούται με 931558.385, το στατιστικό F ισούται με 3.278 και η p-value ισουται με 0.072. Τα residuals δηλαδή τα υπόλοιπα όπως λέμε στην στατιστική, τα σφάλματα δδηλαδή, έχουν άθροισμα τετραγώνων 4.859χ107 με 171 βαθμούς ελευθερίας, και μέσο τετράγωνο να ισούται με 284152.747.

Table 3Πίνακας διακυμάνσεων και συνδιακυμάνσεων μεταξύ των μεταβλητών

| **ANCOVA - salary** | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Cases** | | **Sum of Squares** | | **df** | | **Mean Square** | | **F** | | **p** | |
| profits |  | 1.288×10+6 |  | 1 |  | 1.288×10+6 |  | 4.533 |  | 0.035 |  |
| college |  | 119644.864 |  | 1 |  | 119644.864 |  | 0.421 |  | 0.517 |  |
| comten |  | 368636.635 |  | 1 |  | 368636.635 |  | 1.297 |  | 0.256 |  |
| ceoten |  | 1.820×10+6 |  | 1 |  | 1.820×10+6 |  | 6.405 |  | 0.012 |  |
| sales |  | 931558.385 |  | 1 |  | 931558.385 |  | 3.278 |  | 0.072 |  |
| Residuals |  | 4.859×10+7 |  | 171 |  | 284152.747 |  |  |  |  |  |
|  | | | | | | | | | | | |
| *Note.*  Type III Sum of Squares | | | | | | | | | | | |

# Ερώτημα 2ο

Στο δεύτερο ερώτημα στόχος είναι η υλοποίηση ενός μοντέλου γραμμικής παλινδρόμησης. Σε τέτοια μοντέλα είναι σημαντικός ο εντοπισμός των συντελεστών για να υλοποιηθεί και να στηθεί το μοντέλο. Το μοντέλο μας έχει την εξής μορφή.

*(1)*

Στην εξίσωση 1με y απεικονίζουμε την προβλεπόμενη μεταβλητή της εξαρτημένης μεταβλητής όπου σε αυτή την περίπτωση είναι η μεταβλητή οι πωλήσεις με box απεικονίζουμε το συντελεστής παλινδρόμησης bo της ανεξάρτητης μεταβλητής Χ όπου σε αυτή την περίπτωση είναι ο μισθός του διευθύνοντα συμβούλου και με ε απεικονίζουμε το σφάλμα μεταξύ της πραγματικής τιμής και της προβλεπόμενης.

Σύμφωνα με τον Table 4 έχουμε τα εξής στατιστική μέτρα. Για το στατιστικό μοντέλο R έχουμε ότι το μοντέλο Η0 η τιμή του ισούται με 0, για το R2 ισούται με 0,για το adjusted- R2 ισούται με 0 ενώ το μέσο τετραγωνικό σφάλμα RMSE ισούται με 6088.654. Για το μοντέλο Η1 το στατιστικό μέτρο R ισούται με 0.380, το R2 ισούται με 0.145, το adjusted R2 ισούται με 0.140 ενώ το μέσο τετραγωνικό σφάλμα RMSE ισούται με 5647.429.

Table 4Περίληψη του μοντέλου- sales

| ***Model Summary - sales*** | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***Model*** | | ***R*** | | ***R²*** | | ***Adjusted R²*** | | ***RMSE*** | |
| *H₀* |  | *0.000* |  | *0.000* |  | *0.000* |  | *6088.654* |  |
| *H₁* |  | *0.380* |  | *0.145* |  | *0.140* |  | *5647.429* |  |
|  | | | | | | | | | |

Σύμφωνα με τον Table 5 για το μοντέλο Η1, το άθροισμα τετραγώνων για το συντελεστή παλινδρόμησης ισούται με 9.433×10+8, με έναν βαθμό ελευθερίας, μέσο τετράγωνο 9.433×10+8, το στατιστικό F ισούται με 39.576 και p-value ισούται να είναι μικρότερο του 0.001 όπου το συγκρίνουμε με το 0.05 για να εξετάσουμε αν υπάρχει στατιστικώς σημαντική διαφορά, σε αυτή την περίπτωση αν το μοντέλο μας γίνεται αποδεκτό για προβλέψεις ή όχι. Σε αυτή την περίπτωση είναι αποδεκτό. Το σφάλμα, residual δηλαδή, έχει άθροισμα τετραγώνων ίσο με 5.581×10+9, με 175 βαθμούς ελευθερίας, και μέσο τετραγώνων 3.189×10+7. Συνολικά, το άθροισμα έχει άθροισμα τετραγώνων ίσο με 6.525×10+9 με 176 βαθμούς ελευθερίας.

Table 5Linear regression one variable - ANOVA

| **ANOVA** | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Model** | |  | | **Sum of Squares** | | **df** | | **Mean Square** | | **F** | | **p** | |
| H₁ |  | Regression |  | 9.433×10+8 |  | 1 |  | 9.433×10+8 |  | 29.576 |  | < .001 |  |
|  |  | Residual |  | 5.581×10+9 |  | 175 |  | 3.189×10+7 |  |  |  |  |  |
|  |  | Total |  | 6.525×10+9 |  | 176 |  |  |  |  |  |  |  |
|  | | | | | | | | | | | | | |
| *Note.*  The intercept model is omitted, as no meaningful information can be shown. | | | | | | | | | | | | | |

Σύμφωνα με Table 6 το Η0 έχει ως σταθερά να ισούται με 3529.463 με κανονικό σφάλμα 457.651, το στατιστικό t test ισούται με 7.712 και η p-value ισούται να είναι μικρότερη του 0.05 σεένα διάστημα εμπιστοσύνης 95%. Το μοντέλο Η1 έχει ως σταθερά 118.031 και για συντελεστή έχει το 118.031 με σφάλμα 757.420 και το t -test ισούται με τη σταθερά ίσο με 0.156 και η p-value ισούται με 0.876. Η μεταβλητή salary έχει συντελεστή 3.940 με σφάλμα 0.724, t-test να ισούται με 5.438 και η p-value να ισούται να είναι μικρότερη από το 0.001. Οπότε το μοντέλο μας είναι της μορφής όπως απεικονίζεται στην εξίσωση 2.

*(2)*

Table Coefficients linear regression ove independent variable

| **Coefficients** | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Model** | |  | | **Unstandardized** | | **Standard Error** | | **Standardized** | | **t** | | **p** | |
| H₀ |  | (Intercept) |  | 3529.463 |  | 457.651 |  |  |  | 7.712 |  | < .001 |  |
| H₁ |  | (Intercept) |  | 118.031 |  | 757.420 |  |  |  | 0.156 |  | 0.876 |  |
|  |  | salary |  | 3.940 |  | 0.724 |  | 0.380 |  | 5.438 |  | < .001 |  |
|  | | | | | | | | | | | | | |

# Ερώτημα 3ο

Στο ερώτημα αυτό στόχος είναι η υλοποίηση ενός γραμμικού μοντέλο παλινδρόμησης το οποίο είναι όπως λέμε στην στατιστική πολυμεταβλητό, δηλαδή περιέχει περισσότερες από μία μεταβλητές. Αν η p-τιμή είναι μικρότερη από το επίπεδο σημαντικότητας (π.χ., p <0,05), τότε ο ερευνητής μπορεί να συμπεράνει ότι η παρατηρούμενη επίδραση αντανακλά στην πραγματικότητα τα χαρακτηριστικά του πληθυσμού και όχι μόνο δειγματοληπτικό σφάλμα. Ένας ερευνητής μπορεί στη συνέχεια να αναφέρει ότι το αποτέλεσμα επιτυγχάνει στατιστική σημαντικότητα, απορρίπτοντας έτσι την μηδενική υπόθεση.

Σύμφωνα με τον Table 7 για το μοντέλο Η0 το στατιστικό R ισούται με 0, το R2 ισούται με 0, το adjusted R2 ισούται με 0 και το RMSE ισούται με 6088.654. Όσον αφορά το μοντέλο Η1, το R ισούται με 0.804, το R2 ισούται με 0.647, το adjusted R2 ισούται με 0.638 και το RMSE ισούται με 3661.790.

Table Model Summary - Sales

| **Model Summary - sales** | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Model** | | **R** | | **R²** | | **Adjusted R²** | | **RMSE** | |
| H₀ |  | 0.000 |  | 0.000 |  | 0.000 |  | 6088.654 |  |
| H₁ |  | 0.804 |  | 0.647 |  | 0.638 |  | 3661.790 |  |
|  | | | | | | | | | |

Σύμφωνα με τον Table 8 το μοντέλο Η1 έχει ως παλινδρόμηση το τετραγωνικά αθροίσματα ίσα με 4.218×10+9, βαθμούς ελευθερίας 4, το μέσο τετράγωνο ίσο με 1.055×10+9, το στατιστικό F ίσο με 78.649 και p-value να είναι μικρότερη του 0.001. Το σφάλμα έχει άθροισμα τετραγώνων ίσο με 2.306×10+9, 172 βαθμούς ελευθερίας, μέσο τετράγωνο 1.341×10+7, ενώ συνολικά το άθροισμα τετραγώνων ισούται με 6.525×10+9 με 176 βαθμούς ελευθερίας.

Table ANOVA model H1

| **ANOVA** | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Model** | |  | | **Sum of Squares** | | **df** | | **Mean Square** | | **F** | | **p** | |
| H₁ |  | Regression |  | 4.218×10+9 |  | 4 |  | 1.055×10+9 |  | 78.649 |  | < .001 |  |
|  |  | Residual |  | 2.306×10+9 |  | 172 |  | 1.341×10+7 |  |  |  |  |  |
|  |  | Total |  | 6.525×10+9 |  | 176 |  |  |  |  |  |  |  |
|  | | | | | | | | | | | | | |
| *Note.*  The intercept model is omitted, as no meaningful information can be shown. | | | | | | | | | | | | | |

Σύμφωνα με το Table 9 το Η0 μοντέλο έχει ως unstandardized ίσο με 3529.463, standard error ίσο με 457.651. το στατιστικό t-test ισούται με 7.712 και η τιμή p-value ισούται να είναι μικρότερη του 0.001.

Όσον αφορά το μοντέλο Η1 η σταθερά ισούται με 656.844, το standard error ισούται με 705.648, το στατιστικό t-test ισούται με 0.931 και η τιμή p-value ισούται με 0.353. Για τη μεταβλητή salary , η τιμή του συντελεστή ισούται με 0.937, το σφάλμα ισούται με 0.520 , t-test ισούται με 1.802 και η p-value ισούται με 0.073. Για τη μεταβλητή comten η τιμή του συντελεστή ισούται με 6.652, το σφάλμα ισούται με 24.034, το t-test ισούται με 0.277 ενώ η p-value τιμή ισούται με 0.782. Για την μεταβλητή ceoten η τιμή του συντελεστή ισούται με -58.296, το σφάλμα ισούται με 41.461, το t-test ισούται με -1.406 και η p-value τιμή ισούται με 0.162. Για την μεταβλητή profits η τιμή του συντελεστή ισούται με 11.430, το σφάλμα ισούται με 0.757, το t-test ισούται με 15.096 και η p-value ισούται να είναι μικρότερο του 0.001. Γενικά μιλώντας, το πολύ-μεταβλητό μοντέλο γραμμικής παλινδρόμησης, ισούται να είναι όπως απεικονίζεται στην εξίσωση 3.

*(3)*

Table Coefficients multivariable linear regression model

| **Coefficients** | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Model** | |  | | **Unstandardized** | | **Standard Error** | | **Standardized** | | **t** | | **p** | |
| H₀ |  | (Intercept) |  | 3529.463 |  | 457.651 |  |  |  | 7.712 |  | < .001 |  |
| H₁ |  | (Intercept) |  | 656.844 |  | 705.648 |  |  |  | 0.931 |  | 0.353 |  |
|  |  | salary |  | 0.937 |  | 0.520 |  | 0.090 |  | 1.802 |  | 0.073 |  |
|  |  | comten |  | 6.652 |  | 24.034 |  | 0.013 |  | 0.277 |  | 0.782 |  |
|  |  | ceoten |  | -58.296 |  | 41.461 |  | -0.068 |  | -1.406 |  | 0.162 |  |
|  |  | profits |  | 11.430 |  | 0.757 |  | 0.759 |  | 15.096 |  | < .001 |  |
|  | | | | | | | | | | | | | |

# Ερώτημα 4ο

Στο τέταρτο ερώτημα στόχος είναι να υπολογιστεί ο δείκτης Cronbach’s για τις μεταβλητές q1-q16 οι οποίες είναι μεταβλητές όπου οι τιμές τους είναι από το 0 έως το 5. Τιμές κοντά στο 0 δείχνουν ότι δεν είναι καθόλου ικανοποιημένος ενώ τιμές κοντά στο 5 δείχνουν ότι είναι πολύ ικανοποιημένος.

Σύμφωνα με Table 10 ο συντελεστής Cronbach’s a μάς δίνει τιμή 0.544, όπου σύμφωνα με τη στατιστική δεν είναι αποδεκτός όρος οπότε τα ερωτηματολόγια που μας δόθηκαν δεν είναι τόσο αξιόπιστα. Τιμές κοντά στο 0.7 και πάνω από το στατιστικό τεστ του Cronbach μάς λένε ότι τα ερωτηματολόγια είναι αποδεκτά. Ένα 95% κάτω όριο του δείκτη Cronbach είναι 0.438, ενώ ένα 95% άνω όριο του δείκτη Cronbach είναι 0.635.

Table 10Cronbach's reliability test q1-q16

**Unidimensional Reliability**

| **Frequentist Scale Reliability Statistics** | | | |
| --- | --- | --- | --- |
| **Estimate** | | **Cronbach's α** | |
| Point estimate |  | 0.544 |  |
| 95% CI lower bound |  | 0.438 |  |
| 95% CI upper bound |  | 0.635 |  |
|  | | | |
| *Note.*  The following items correlated negatively with the scale: q1, q4, q9, q13. | | | |

# Ερώτημα 5ο

Σύμφωνα με το Table 11 το στατιστικό chi-squared έχει τιμή ίση με 130.461, 75 βαθμούς ελευθερίας και p-value μικρότερο του 0.001.

Table Chi-squared test

| **Chi-squared Test** | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **Value** | | **df** | | **p** | |
| Model |  | 130.461 |  | 75 |  | < .001 |  |
|  | | | | | | | |

Σύμφωνα με Table 12 τα ερωτηματολόγια q6, q11, q7, q15, q8, q12, q16 αντιστοιχούν στον πρώτο παράγοντα με τιμές 0.728, 0.551, 0.512, σ0.495, 0.482, 0.447, 0.419. Τα ερωτηματολόγια q10, q14, q3 έχουν τιμές 1.058, 0.509, 0.469, ενώ τα ερωτηματολογια q9, q1, q13, q4 έχουν τιμές0.728, 0.727, 0.528, 0.431 αντίστοιχα. Τα ερωτηματολόγια q2 και q5 δεν ανήκουν σε κάποιο παράγοντα ανάλυσης.

Table Factor Loadings

| **Factor Loadings** | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **Factor 1** | | **Factor 2** | | **Factor 3** | | **Uniqueness** | |
| q6 |  | 0.728 |  |  |  |  |  | 0.567 |  |
| q11 |  | 0.551 |  |  |  |  |  | 0.715 |  |
| q7 |  | 0.512 |  |  |  |  |  | 0.666 |  |
| q15 |  | 0.495 |  |  |  |  |  | 0.651 |  |
| q8 |  | 0.482 |  |  |  |  |  | 0.780 |  |
| q12 |  | 0.447 |  |  |  |  |  | 0.735 |  |
| q16 |  | 0.419 |  |  |  |  |  | 0.790 |  |
| q10 |  |  |  | 1.058 |  |  |  | -0.002 |  |
| q14 |  |  |  | 0.509 |  |  |  | 0.716 |  |
| q3 |  |  |  | 0.469 |  |  |  | 0.664 |  |
| q9 |  |  |  |  |  | 0.738 |  | 0.433 |  |
| q1 |  |  |  |  |  | 0.727 |  | 0.492 |  |
| q13 |  |  |  |  |  | 0.528 |  | 0.662 |  |
| q4 |  |  |  |  |  | 0.431 |  | 0.702 |  |
| q2 |  |  |  |  |  |  |  | 0.873 |  |
| q5 |  |  |  |  |  |  |  | 0.866 |  |
|  | | | | | | | | | |
| *Note.*  Applied rotation method is promax. | | | | | | | | | |

Σύμφωνα με Table 13 τα factor loadings είναι τα εξής. Το ερωτηματολόγιο q1 ισούται με 0.661. Το ερωτηματολόγιο q2 δεν ανήκει κάπου. Το ερωτηματολόγιο q3 ισούται με 0.527. Το ερωτηματολόγιο q4 ισούται 0.505. Το ερωτηματολόγιο q5 δεν ανήκει σε κάποιο από τους παράγοντες γι’αυτό και δεν έχει τιμή. Το ερωτηματολόγιο q6 έχει τιμή ίση με 0.644. Το ερωτηματολόγιο q7 ισούται με 0.543. Το ερωτηματολόγιο q8 ισούται με 0.461. Το ερωτηματολόγιο q9 ισούται με 0.749. Το ερωτηματολόγιο q10 ισούται με 0.987. Το ερωτηματολόγιο q11 ισούται με 0.531. Το ερωτηματολόγιο q12 ισούται με 0.504. Το ερωτηματολόγιο q13 ισούται με 0.569. Το ερωτηματολόγιο q14 ισούται με 0.531. Το ερωτηματολόγιο q14 ισούται με 0.531. Το ερωτηματολόγιο q15 ισούται με 0.574. Το ερωτηματολόγιο q16 ισούται με 0.452.

Table 13Factor Loadings (Structure Matrix)

| **Factor Loadings (Structure Matrix)** | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **Factor 1** | | **Factor 2** | | **Factor 3** | |
| q1 |  |  |  |  |  | 0.661 |  |
| q2 |  |  |  |  |  |  |  |
| q3 |  |  |  | 0.527 |  |  |  |
| q4 |  |  |  |  |  | 0.505 |  |
| q5 |  |  |  |  |  |  |  |
| q6 |  | 0.644 |  |  |  |  |  |
| q7 |  | 0.543 |  |  |  |  |  |
| q8 |  | 0.461 |  |  |  |  |  |
| q9 |  |  |  |  |  | 0.749 |  |
| q10 |  |  |  | 0.987 |  |  |  |
| q11 |  | 0.531 |  |  |  |  |  |
| q12 |  | 0.504 |  |  |  |  |  |
| q13 |  |  |  |  |  | 0.569 |  |
| q14 |  |  |  | 0.531 |  |  |  |
| q15 |  | 0.574 |  |  |  |  |  |
| q16 |  | 0.452 |  |  |  |  |  |
|  | | | | | | | |
| *Note.*  Applied rotation method is promax. | | | | | | | |

Στον Table 14 απεικονίζονται τα χαρακτηριστικά κάθε ένα από τους τους τρεις παράγοντες που δημιουργήθηκαν. Για τον παράγοντα τον πρώτο η ιδιοτιμή ισούται με 3.800, για τον δεύτερο παράγοντα η ιδιοτιμή ισούται με 1.904, ενώ για τον τρίτο παράγοντα η ιδιοτιμή ισούται με 1.580.Μετέπειτα ο κάθε παράγοντας χωρίζεται σε δύο κατηγορίες, σε unrotated solution και σε rotated solution. Στο unrotated solution το sum sq. loadings για τον πρώτο παράγοντα ισούται με 3.189, για τον δεύτερο 1.379 και για τον τρίτο 1.128, το proportion variable ισούται με 0.199 για τον πρώτο παράγοντα, 0.086 για τον δεύτερο παράγοντα και 0.071 για τον τρίτο παράγοντα. Αθροιστικά, για τον πρώτο παράγοντα το άθροισμα ισούται με 0.199, 0.285για το δεύτερο παράγοντα και 0.356 για τον τρίτο παράγοντα. Όσον αφορά την κατηγορία rotated solution, το sum sq. loadings ισούται με 2.237 για τον πρώτο παράγοντα, 1.790 για τον δεύτερο παράγοντα και 0.104 για τον τρίτο παράγοντα. Το proportion variable ισούται με 0.140 για τον πρώτο παράγοντα, 0.112 για το δεύτερο παράγοντα και 0.104 για τον τρίτο παράγοντα. Αθροιστικά, έχουμε ότι το άθροισμα ισούται με 0.140 για τον πρώτο παράγοντα, 0.252 για το δεύτερο παράγοντα κι 0.356 για τον τρίτο παράγοντα.

Table Factor Characteristics

| **Factor Characteristics** | | | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | | | **Unrotated solution** | | | | | | **Rotated solution** | | | | | |
|  | | **Eigenvalues** | | **SumSq. Loadings** | | **Proportion var.** | | **Cumulative** | | **SumSq. Loadings** | | **Proportion var.** | | **Cumulative** | |
| Factor 1 |  | 3.800 |  | 3.189 |  | 0.199 |  | 0.199 |  | 2.237 |  | 0.140 |  | 0.140 |  |
| Factor 2 |  | 1.904 |  | 1.379 |  | 0.086 |  | 0.285 |  | 1.790 |  | 0.112 |  | 0.252 |  |
| Factor 3 |  | 1.580 |  | 1.128 |  | 0.071 |  | 0.356 |  | 1.665 |  | 0.104 |  | 0.356 |  |
|  | | | | | | | | | | | | | | | |

Σύμφωνα με τον Table 15 για τρεις παράγοντες factor 1, factor 2, factor 3 υπολογίσαμε την συσχέτιση μεταξύ αυτών. Η συσχέτιση μεταξύ των ίδιων παραγόντων προφανώς βγαίνει 1 γιατί υπάρχει άμεση συσχέτιση. Η συσχέτιση μεταξύ των παραγόντων 1 και 2 ισούται με 0.444 θετική στο πρόσημο και αρκετά μεγάλη δηλαδή υπάρχει θετική επίδραση μεταξύ των παραγόντων 1 και 2. Αρνητική είναι η συσχέτιση μεταξύ των παραγόντων 1 και 3 μιας και που ισούται με -0.350. Αρνητική συσχέτιση μεταξύ των παραγόντων 2 και 3 μιας και που η συσχέτιση ισούται με -0.187.

Table Factor correlations between the three factors

| **Factor Correlations** | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **Factor 1** | | **Factor 2** | | **Factor 3** | |
| Factor 1 |  | 1.000 |  | 0.444 |  | -0.350 |  |
| Factor 2 |  | 0.444 |  | 1.000 |  | -0.187 |  |
| Factor 3 |  | -0.350 |  | -0.187 |  | 1.000 |  |
|  | | | | | | | |

Σύμφωνα με Table 16 το RMSEA ισούται με 0.064. Ένα 90% διάστημα εμπιστοσύνης όσον αφορά το RMSEA ισούται το διάστημα μεταξύ 0.046 και 0.083. Το SRMR ισούται με 0.050, το TLI ισούται με 0.833, το CFI ισούται με 0.89 και το BIC ισούται με -257.750.

Table Additional fit indices

| **Additional fit indices** | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **RMSEA** | | **RMSEA 90% confidence** | | **SRMR** | | **TLI** | | **CFI** | | **BIC** | |
| 0.064 |  | 0.046 - 0.083 |  | 0.050 |  | 0.833 |  | 0.897 |  | -257.750 |  |
|  | | | | | | | | | | | |

Παρουσιάζεται ο πίνακας με τα τετραγωνικά σφάλματα για κάθε ένα μεταξύ των ερωτηματολόγιων όπως απεικονίζεται στον Table 17.

Table Residual matrix

| **Residual Matrix** | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **q1** | | **q2** | | **q3** | | **q4** | | **q5** | | **q6** | | **q7** | | **q8** | | **q9** | | **q10** | | **q11** | | **q12** | | **q13** | | **q14** | | **q15** | | **q16** | |
| q1 |  | 0.492 |  | -0.031 |  | 0.027 |  | -0.001 |  | 0.052 |  | 0.013 |  | -0.035 |  | 0.015 |  | -0.007 |  | 0.001 |  | 0.006 |  | -0.002 |  | 0.019 |  | -0.039 |  | -0.018 |  | 0.040 |  |
| q2 |  | -0.031 |  | 0.873 |  | -0.012 |  | 0.012 |  | 0.018 |  | -0.049 |  | 0.016 |  | 0.059 |  | -0.017 |  | 0.001 |  | 0.165 |  | 0.007 |  | 0.020 |  | -0.010 |  | -0.111 |  | -0.054 |  |
| q3 |  | 0.027 |  | -0.012 |  | 0.664 |  | 0.026 |  | 0.000 |  | -0.019 |  | -0.029 |  | -0.084 |  | 0.003 |  | -0.017 |  | 0.049 |  | -0.004 |  | -0.056 |  | 0.031 |  | 0.062 |  | 0.006 |  |
| q4 |  | -0.001 |  | 0.012 |  | 0.026 |  | 0.702 |  | -0.016 |  | -0.077 |  | 0.062 |  | 0.012 |  | -0.004 |  | 0.001 |  | 0.020 |  | -0.008 |  | -0.001 |  | -0.059 |  | 0.060 |  | -0.058 |  |
| q5 |  | 0.052 |  | 0.018 |  | 0.000 |  | -0.016 |  | 0.866 |  | -0.227 |  | 0.074 |  | 0.069 |  | 0.007 |  | 0.044 |  | 0.013 |  | 0.029 |  | -0.001 |  | -0.154 |  | 0.048 |  | 0.102 |  |
| q6 |  | 0.013 |  | -0.049 |  | -0.019 |  | -0.077 |  | -0.227 |  | 0.567 |  | 0.020 |  | 0.036 |  | 0.020 |  | 0.047 |  | -0.042 |  | 0.000 |  | -0.042 |  | 0.077 |  | 0.030 |  | -0.066 |  |
| q7 |  | -0.035 |  | 0.016 |  | -0.029 |  | 0.062 |  | 0.074 |  | 0.020 |  | 0.666 |  | 0.057 |  | 0.022 |  | -0.033 |  | -0.014 |  | 0.049 |  | -0.015 |  | 0.032 |  | -0.024 |  | -0.061 |  |
| q8 |  | 0.015 |  | 0.059 |  | -0.084 |  | 0.012 |  | 0.069 |  | 0.036 |  | 0.057 |  | 0.780 |  | 0.019 |  | 0.020 |  | -0.036 |  | -0.094 |  | -0.004 |  | -0.000 |  | 0.030 |  | -0.008 |  |
| q9 |  | -0.007 |  | -0.017 |  | 0.003 |  | -0.004 |  | 0.007 |  | 0.020 |  | 0.022 |  | 0.019 |  | 0.433 |  | -0.013 |  | -0.047 |  | -0.046 |  | 0.009 |  | 0.029 |  | -0.013 |  | 0.048 |  |
| q10 |  | 0.001 |  | 0.001 |  | -0.017 |  | 0.001 |  | 0.044 |  | 0.047 |  | -0.033 |  | 0.020 |  | -0.013 |  | -0.002 |  | 0.012 |  | -0.000 |  | 0.019 |  | 0.034 |  | -0.039 |  | -0.027 |  |
| q11 |  | 0.006 |  | 0.165 |  | 0.049 |  | 0.020 |  | 0.013 |  | -0.042 |  | -0.014 |  | -0.036 |  | -0.047 |  | 0.012 |  | 0.715 |  | 0.125 |  | 0.044 |  | -0.106 |  | -0.086 |  | -0.014 |  |
| q12 |  | -0.002 |  | 0.007 |  | -0.004 |  | -0.008 |  | 0.029 |  | 0.000 |  | 0.049 |  | -0.094 |  | -0.046 |  | -0.000 |  | 0.125 |  | 0.735 |  | 0.026 |  | -0.035 |  | -0.066 |  | -0.038 |  |
| q13 |  | 0.019 |  | 0.020 |  | -0.056 |  | -0.001 |  | -0.001 |  | -0.042 |  | -0.015 |  | -0.004 |  | 0.009 |  | 0.019 |  | 0.044 |  | 0.026 |  | 0.662 |  | 0.004 |  | 0.002 |  | 0.017 |  |
| q14 |  | -0.039 |  | -0.010 |  | 0.031 |  | -0.059 |  | -0.154 |  | 0.077 |  | 0.032 |  | -0.000 |  | 0.029 |  | 0.034 |  | -0.106 |  | -0.035 |  | 0.004 |  | 0.716 |  | 0.016 |  | -0.014 |  |
| q15 |  | -0.018 |  | -0.111 |  | 0.062 |  | 0.060 |  | 0.048 |  | 0.030 |  | -0.024 |  | 0.030 |  | -0.013 |  | -0.039 |  | -0.086 |  | -0.066 |  | 0.002 |  | 0.016 |  | 0.651 |  | 0.204 |  |
| q16 |  | 0.040 |  | -0.054 |  | 0.006 |  | -0.058 |  | 0.102 |  | -0.066 |  | -0.061 |  | -0.008 |  | 0.048 |  | -0.027 |  | -0.014 |  | -0.038 |  | 0.017 |  | -0.014 |  | 0.204 |  | 0.790 |  |
|  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

# Ερώτημα 6ο

Στόχος του έκτου ερωτήματος είναι ότι για τον 1ο παράγοντα που θα προκύψει από την παραγοντική ανάλυση, εφαρμόζουμε ανάλυση διακύμανσης ANOVA για να ελέγξουμε στατιστικά σημαντική σχέση με την ηλικιακή ομάδα του διευθύνοντος συμβούλου. Χωρίζουμε την ηλικία σε 4 ομάδες (<=52, 53-57, 58-62, >62). Έτσι , λοιπόν προκύπτουν τα εξής αποτελέσματα.

Σύμφωνα με Table 18 εφαρμόσαμε Ανάλυση διακύμανσης αφού κατηγοριοποιήσαμε την μεταβλητή age σε <=52, 53-57, 58-62, >62. Αφού συνέβη αυτό, παίρνουμε τα εξής αποτελέσματα. Το τετραγωνικό άθροισμα ισούται με 395.800, με έναν βαθμό ελευθερίας, και μέσα τετράγωνο 395.800, με F-test να ισούται 5.730 και p-value να ισούται 0.018. Το σφάλμα residual σε αυτή την περίπτωση ισούται με 12087.567, με 175 βαθμούς ελευθερίας, 69.072 μέσα τετραγωνικό και f-test να ισούται 69.072.

Table 18ANOVA age grouping

**ANOVA**

| **ANOVA - age** | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Cases** | | **Sum of Squares** | | **df** | | **Mean Square** | | **F** | | **p** | |
| college |  | 395.800 |  | 1 |  | 395.800 |  | 5.730 |  | 0.018 |  |
| Residuals |  | 12087.567 |  | 175 |  | 69.072 |  |  |  |  |  |
|  | | | | | | | | | | | |
| *Note.*  Type III Sum of Squares | | | | | | | | | | | |

# Ερώτημα 7ο

Στόχος του 7ου ερωτήματος είναι να εφαρμόσουμε ανάλυση κατά συστάδες για τις μεταβλητές q1-q16 και να ερμηνεύσουμε τα αποτελέσματα. Πρώτα ας μελετήσουμε τι είναι ανάλυση κατά συστάδες. Αυτή η μεθοδολογία θα πραγματοποιηθεί με τη λεγόμενη μέθοδο Principal Components Analysis. Το Principal Component Analysis (PCA) είναι μια μέθοδος μάθησης χωρίς επίβλεψη\* που χρησιμοποιεί μοτίβα που υπάρχουν σε δεδομένα υψηλών διαστάσεων (δεδομένα με πολλές ανεξάρτητες μεταβλητές) για να μειώσει την πολυπλοκότητα των δεδομένων διατηρώντας παράλληλα τις περισσότερες πληροφορίες. Επειδή ο πρωταρχικός στόχος του PCA είναι να μειώσει τον αριθμό των μεταβλητών που απαιτούνται για την περιγραφή ενός συνόλου δεδομένων, είναι πιο χρήσιμο όταν υπάρχουν πάρα πολλές μεταβλητές στα δεδομένα για εύκολη εξερεύνηση/οπτικοποίηση. Οι μεταβλητές σε ένα σύνολο δεδομένων μπορεί να παρουσιάζουν πολυσυγγραμμικότητα, που σημαίνει ότι υπάρχει σημαντικός βαθμός συσχέτισης μεταξύ δύο ή περισσότερων μεταβλητών. Αυτό σημαίνει ότι οι τιμές μιας μεταβλητής μπορούν να περιγραφούν από τις τιμές μιας άλλης. Ωστόσο, για πολλά στατιστικά μοντέλα, είναι σημαντικό οι μεταβλητές να είναι ανεξάρτητες μεταξύ τους (εξ ου και ο κοινός όρος "ανεξάρτητες μεταβλητές). Οι αναλύσεις γίνονται δύσκολες ή ακόμα και αδύνατες. Τα κύρια στοιχεία που δημιουργούνται από το PCA δεν παρουσιάζουν συγγραμμικότητα. Ένας άλλος τρόπος να το πούμε αυτό είναι ότι καθένα από τα κύρια στοιχεία είναι τελείως ορθογώνια μεταξύ τους (η συσχέτιση τους με άλλα κύρια στοιχεία είναι μηδενική). Όταν τα κύρια στοιχεία χρησιμοποιούνται ως είσοδος στην πολλαπλή παλινδρόμηση, το PCA μπορεί να βοηθήσει στην εξάλειψη προβλημάτων με την υπερπροσαρμογή (πρόβλημα που παρουσιάζεται όταν ένα μοντέλο ταιριάζει πολύ στα δεδομένα του δείγματος και θα έχει κακή απόδοση κατά την πρόβλεψη τιμών από τον μεγαλύτερο πληθυσμό από τον οποίο δειγματοληψία δεδομένων). Αυτό συμβαίνει συχνά επειδή υπάρχουν πάρα πολλές μεταβλητές στα δεδομένα σε σύγκριση με τον αριθμό των παρατηρήσεων. Σε αυτές τις περιπτώσεις, ο θόρυβος (τυχαίο σφάλμα) στα δεδομένα θα έχει πολύ μεγάλο αντίκτυπο στο μοντέλο. Επειδή το PCA μπορεί να χρησιμοποιηθεί για τη μείωση του αριθμού των μεταβλητών, μπορεί να βοηθήσει στην αντιμετώπιση προβλημάτων με την υπερπροσαρμογή. Η εκτέλεση PCR απαιτεί την επιλογή μιας εξαρτημένης μεταβλητής. Αυτή η εξαρτημένη μεταβλητή δεν πρέπει επίσης να περιλαμβάνεται στο PCA. Από προεπιλογή, το Prism επιλέγει όλες τις (συνεχείς) μεταβλητές που θα συμπεριληφθούν στο PCA, επομένως δεν υπάρχουν διαθέσιμες μεταβλητές για επιλογή ως εξαρτημένη μεταβλητή για την PCR. Ως αποτέλεσμα, η PCR είναι γκριζαρισμένη. Μόλις αποεπιλεγεί μια μεταβλητή από τη λίστα των μεταβλητών που θα είναι μέρος του PCA, η επιλογή για PCR θα είναι διαθέσιμη. Όταν έχετε αμφιβολίες, τυποποιήστε τα δεδομένα σας. Το κέντρο δεδομένων περιλαμβάνει πρώτα τον προσδιορισμό της μέσης τιμής για κάθε μεταβλητή και, στη συνέχεια, την αφαίρεση αυτής της μέσης τιμής από κάθε τιμή της μεταβλητής. Στο σύνολο δεδομένων που προκύπτει, κάθε μεταβλητή έχει μέσο όρο μηδέν. Σημειώστε ότι το κεντράρισμα από μόνο του δεν αλλάζει την τυπική απόκλιση μιας μεταβλητής. Η τυποποίηση δεδομένων περιλαμβάνει πρώτα το κεντράρισμα (βλ. παραπάνω) των μεταβλητών. Στη συνέχεια, προσδιορίζεται η τυπική απόκλιση για κάθε μεταβλητή και κάθε κεντρική τιμή διαιρείται με την τυπική απόκλιση της μεταβλητής της. Αυτό οδηγεί σε ένα σύνολο δεδομένων για το οποίο κάθε μεταβλητή έχει μέσο όρο μηδέν και τυπική απόκλιση 1 (και επομένως διακύμανση 1). Είναι σπάνιο να εκτελείται PCA σε δεδομένα που δεν είναι ούτε κεντραρισμένα ούτε τυποποιημένα (αν και γίνεται σε μικρό αριθμό κλάδων). Το Prism δεν προσφέρει αυτήν την επιλογή. Το PCA λειτουργεί αναλύοντας τη διακύμανση ενός συνόλου δεδομένων. Οι μεταβλητές με μεγαλύτερες διακυμάνσεις έχουν μεγαλύτερο αντίκτυπο στα αποτελέσματα της PCA. Ωστόσο, οι διαφορές στη διακύμανση μπορεί απλώς να οφείλονται σε διαφορές στις κλίμακες μέτρησης (για παράδειγμα, οι μετρήσεις μήκους σε χιλιοστά θα έχουν μεγαλύτερη διακύμανση από τις μετρήσεις του ίδιου μήκους σε μέτρα λόγω μόνο της κλίμακας μέτρησης). Σε ορισμένες περιπτώσεις, μπορεί να είναι σημαντικό να διατηρηθεί η σχέση διακυμάνσεων στο σύνολο δεδομένων, αλλά γενικά συνιστάται η τυποποίηση των δεδομένων (ορίζοντας τη διακύμανση κάθε μεταβλητής ίση με 1, βλέπε παραπάνω).

Σύμφωνα με Table 19 η τιμή του στατιστικού χ τετραγώνου όπως λέμε στην στατιστική ισούται με 195.326. Οι βαθμοί ελευθερίας ισούνται με 75 και η p-value τιμή για ένα διάστημα εμπιστοσύνης 95% ισούται να είναι μικρότερη από 0.001.

Table 19Chi-squared test PCA Analysis

| **Chi-squared Test** | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **Value** | | **df** | | **p** | |
| Model |  | 195.326 |  | 75 |  | < .001 |  |
|  | | | | | | | |

Όπως παρουσιάζονται στον Table 20για τρεις συστάδες που δημιουργήθηκαν από την PCA ανάλυση έχουμε ότι για την πρώτη συστάδα η τιμή για το q6 ερωτηματολόγιο κυμαίνεται γύρω στο 0.811, για το q11 κυμαίνεται γύρω στο 0.668, για το q8 ερωτηματολόγιο γύρω στο 0.599, για το q7 ερωτηματολόγιο γύρω στο 0.559, για το q12 ερωτηματολόγιο γύρω στο 0.510, για το q15 γύρω στο 0.505 για το q16 ερωτηματολόγιο γύρω στο 0.463 και για το q2 ερωτηματολόγιο γύρω στο 0.409. Για τη δεύτερη συστάδα το ερωτηματολόγιο q1 συγκεντρώνει τις τιμές του στο 0.814, το q9 στο 0.795, το q13 στο 0.665, το q4 στο 0.559. Για την Τρίτη συστάδα το ερωτηματολόγιο q10 συγκεντρώνει τις τιμές του στο 0.898, το q14 συγκεντρώνει τις τιμές του στο 0.689, το q3 συγκεντρώνει τις τιμές του στο 0.648 και το q5 συγκεντρώνει τις τιμές του στο 0.497.

Table Component Loadings PCA

| **Component Loadings** | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **RC1** | | **RC2** | | **RC3** | | **Uniqueness** | |
| q6 |  | 0.811 |  |  |  |  |  | 0.442 |  |
| q11 |  | 0.668 |  |  |  |  |  | 0.589 |  |
| q8 |  | 0.599 |  |  |  |  |  | 0.660 |  |
| q7 |  | 0.559 |  |  |  |  |  | 0.563 |  |
| q12 |  | 0.510 |  |  |  |  |  | 0.652 |  |
| q15 |  | 0.505 |  |  |  |  |  | 0.584 |  |
| q16 |  | 0.463 |  |  |  |  |  | 0.721 |  |
| q2 |  | 0.409 |  |  |  |  |  | 0.804 |  |
| q1 |  |  |  | 0.814 |  |  |  | 0.346 |  |
| q9 |  |  |  | 0.795 |  |  |  | 0.352 |  |
| q13 |  |  |  | 0.665 |  |  |  | 0.513 |  |
| q4 |  |  |  | 0.559 |  |  |  | 0.568 |  |
| q10 |  |  |  |  |  | 0.898 |  | 0.213 |  |
| q14 |  |  |  |  |  | 0.689 |  | 0.527 |  |
| q3 |  |  |  |  |  | 0.648 |  | 0.466 |  |
| q5 |  |  |  |  |  | 0.497 |  | 0.714 |  |
|  | | | | | | | | | |
| *Note.*  Applied rotation method is promax. | | | | | | | | | |

Σύμφωνα με Table 21 η συστάδα η πρώτη έχει ιδιοτιμή ίση με 3.800, η δεύτερη 1.904,ενώ η Τρίτη ίση με 1.580. Το Proportion variance ισούται με 0.237 για την πρώτη συστάδα, 0.119 για τη δεύτερη και 0.099 για την Τρίτη συστάδα. Αθροιστικά στην κατηγορία unrotated solution για την πρώτη συστάδα ισούται με 0.237, για την δεύτερη συστάδα ισούται με 0.356 και για την τρίτητ συστάδα ισούται με 0.455. Για την κατηγορία rotated solution έχουμε sum squares loadings για την πρώτη συστάδα ισούται με 2.866, 2.253 για την δεύτερη συστάδα και 2.164 για την Τρίτη συστάδα. Το proportion variance ισούται με 0.179 για την πρώτη συστάδα, 0.141 για την δεύτερη συστάδα και 0.135 για την Τρίτη συστάδα. Αθροιστικά, έχουμε ότι για την πρώτη συστάδα ισούται με 0.179, για την δεύτερη ισούται με 0.320 και για την Τρίτη συστάδα ισούται με 0.455.

Table Component Characteristics - PCA Analysis

| **Component Characteristics** | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **Unrotated solution** | | | | | | **Rotated solution** | | | | | |
|  | | **Eigenvalue** | | **Proportion var.** | | **Cumulative** | | **SumSq. Loadings** | | **Proportion var.** | | **Cumulative** | |
| Component 1 |  | 3.800 |  | 0.237 |  | 0.237 |  | 2.866 |  | 0.179 |  | 0.179 |  |
| Component 2 |  | 1.904 |  | 0.119 |  | 0.356 |  | 2.253 |  | 0.141 |  | 0.320 |  |
| Component 3 |  | 1.580 |  | 0.099 |  | 0.455 |  | 2.164 |  | 0.135 |  | 0.455 |  |
|  | | | | | | | | | | | | | |