ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
ΤΜΗΜΑ ΟΡΓΑΝΩΣΗΣ ΚΑΙ ΔΙΟΙΚΗΣΗΣ ΕΠΙΧΕΙΡΗΣΕΩΝ

ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ MBA - TQM

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΣΤΑΤΙΣΤΙΚΗ ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ ΠΩΛΗΣΕΩΝ ΦΥΣΙΚΩΝ ΧΥΜΩΝ ΣΕ ΠΕΡΙΟΧΗ ΤΩΝ Η.Π.Α.

ΚΩΣΤΟΠΟΥΛΟΥ ΔΙΟΝΥΣΙΑ

ΕΠΙΒΛΕΠΩΝ ΚΑΘΗΓΗΤΗΣ:
ΜΑΡΑΒΕΛΑΚΗΣ ΠΕΤΡΟΣ

ΠΑΡΑΡΤΗΜΑ Δ : **ΥΠΕΥΘΥΝΗ ΔΗΛΩΣΗ ΕΚΠΟΝΗΣΗΣ ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ**

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

ΤΜΗΜΑ ΟΡΓΑΝΩΣΗΣ ΚΑΙ ΔΙΟΙΚΗΣΗΣ ΕΠΙΧΕΙΡΗΣΕΩΝ

Μεταπτυχιακό Πρόγραμμα Σπουδών

στη «Διοίκηση Επιχειρήσεων – Ολική Ποιότητα» με διεθνή προσανατολισμό

ΒΕΒΑΙΩΣΗ ΕΚΠΟΝΗΣΗΣ ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ

(περιλαμβάνεται ως ξεχωριστή [δεύτερη] σελίδα στο σώμα της διπλωματικής εργασίας)

Δηλώνω υπεύθυνα ότι η διπλωματική εργασία για τη λήψη του μεταπτυχιακού τίτλου σπουδών, του Πανεπιστημίου Πειραιώς, στη Διοίκηση Επιχειρήσεων - Ολική Ποιότητα με διεθνή προσανατολισμό με τίτλο:

"Στατιστική Ανάλυση Δεδομένων Πωλήσεων Φυσικών Χυμών σε περιοχή των Η.Π.Α."

έχει συγγραφεί από εμένα αποκλειστικά και στο σύνολό της. Δεν έχει υποβληθεί ούτε έχει εγκριθεί στο πλαίσιο κάποιου άλλου μεταπτυχιακού προγράμματος ή προπτυχιακού τίτλου σπουδών, στην Ελλάδα ή στο εξωτερικό, ούτε είναι εργασία ή τμήμα εργασίας ακαδημαϊκού ή επαγγελματικού χαρακτήρα.

Δηλώνω επίσης υπεύθυνα ότι οι πηγές στις οποίες ανέτρεξα για την εκπόνηση της συγκεκριμένης εργασίας, αναφέρονται στο σύνολό τους, κάνοντας πλήρη αναφορά στους συγγραφείς, τον εκδοτικό οίκο ή το περιοδικό, συμπεριλαμβανομένων και των πηγών που ενδεχομένως χρησιμοποιήθηκαν από το διαδίκτυο.

Παράβαση της ανωτέρω ακαδημαϊκής μου ευθύνης αποτελεί ουσιώδη λόγο για την ανάκληση του πτυχίου μου.

Υπογραφή Μεταπτυχιακού Φοιτητή/ τριας .....

Ονοματεπώνυμο .....Διονυσία Κωστοπούλου.....

Ημερομηνία .....16/7/2018.....

ΠΕΡΙΛΗΨΗ

Στην παρούσα εργασία μελετώνται οι πωλήσεις φυσικών χυμών πορτοκαλιού στην ευρύτερη περιοχή του Σικάγο.

Στο εισαγωγικό κεφάλαιο αναφέρεται ο σκοπός της εργασίας και γίνεται μια σύντομη αναφορά στην παγκόσμια αγορά χυμών.

Στο 2ο κεφάλαιο αναφέρονται και εξηγούνται συνοπτικά οι βασικές έννοιες, εργαλεία και μεθοδολογίες της Στατιστικής που χρησιμοποιούνται στη συνέχεια για την ανάλυση των δεδομένων: οι κύριες γραφικές και αριθμητικές μέθοδοι της περιγραφικής στατιστικής, η γραμμική και πολλαπλή παλινδρόμηση και η ανάλυση διακύμανσης (ANOVA).

Στο 3ο κεφάλαιο παρουσιάζονται τα δεδομένα που διαθέτουμε καθώς και η δομή που ακολουθεί η ανάλυσή τους κατά την εξέλιξη της εργασίας.

Στο 4ο κεφάλαιο ακολουθεί η κύρια ανάλυση μέσω της χρήσης των εργαλείων που προαναφέρθηκαν. Συγκεκριμένα, παρουσιάζονται οι κατανομές των πωλήσεων και στη συνέχεια μελετάται η επίδραση τιμής, διαφήμισης, εμπορικού σήματος και δημογραφικών παραγόντων προκειμένου να διερευνηθεί κατά πόσο οι παράγοντες αυτοί επηρεάζουν τις πωλήσεις φυσικών χυμών πορτοκαλιού διαχρονικά. Τέλος, επειχειρείται η εξαγωγή ενιαίου μοντέλου που να ερμηνεύει τις διαφοροποιήσεις των πωλήσεων.

Στο τελευταίο κεφάλαιο, λαμβάνοντας υπόψη τα αποτελέσματα που προέκυψαν από την προηγειθείσα ανάλυση, παρατίθενται τα συμπεράσματα καθώς και προτάσεις για μελλοντική μελέτη.

ABSTRACT

Scope of this master thesis is the statistical data analysis of orange juice sales in Chicago area.

In the introductory chapter, the world juice market is briefly presented and linked to the objective of the thesis regarding orange juices sales.

In chapter 2, all Statistic terms, tools and methodologies that are subsequently used for data analysis are introduced and explained: main graphical and numerical methods of descriptive statistics, linear and multiple regression and analysis of variance (ANOVA).

In chapter 3, the available data as well as the whole structure of the following analysis are presented.

In chapter 4, follows the main data analysis through the aforementioned tools. First, the sales distribution is depicted and then the impact of price, advertising, brand and demographic features is calculated in order to investigate how these factors affect orange juice sales over time. Furthermore, in order to interpret the sales differentiations, the creation of a single model is attempted.

In the final chapter, after taking into account the results of the previous analysis, all the conclusions are presented followed by suggestions for future research.

ΠΕΡΙΕΧΟΜΕΝΑ

ΛΙΣΤΑ ΠΙΝΑΚΩΝ

ΛΙΣΤΑ ΔΙΑΓΡΑΜΜΑΤΩΝ ΚΑΙ ΣΧΗΜΑΤΩΝ

ΚΕΦΑΛΑΙΟ 1: ΕΙΣΑΓΩΓΗ

Η διεθνής αγορά φρούτων παρουσίασε ραγδαία ανάπτυξη τα τελευταία χρόνια, από το 1980 και μετά, λόγω της αύξησης των εισοδημάτων, τη μείωση του μεταφορικού κόστους, της ανάπτυξης της τεχνολογίας και της σύναψης διεθνών συμφωνιών. Πρώτα στη διεθνή αγορά φρούτων ως προς την αξία των πωλήσεων έρχονται τα εσπεριδοειδή.

Να σημειωθεί ότι το 1/3 των εσπεριδοειδών πωλείται σε επεξεργασμένη μορφή, με το 80% της παραγωγής πορτοκαλιών να μετατρέπεται σε χυμούς. Το κυριότερο χαρακτηριστικό της παγκόσμιας αγοράς χυμών πορτοκαλιού είναι η γεωγραφική συγκέντρωση της παραγωγής, όπου κυριαρχούν η πολιτεία της Φλόριντα των Η.Π.Α. και η πολιτεία του Σάο Πάολο στη Βραζιλία. Η παραγωγή χυμών πορτοκαλιού από τις δύο αυτές πολιτείες ξεπερνά το 80% της παγκόσμιας παραγωγής. Η βασική τους διαφορά είναι ότι η Βραζιλία εξάγει το 99% της παραγωγής της, σε αντίθεση με τη Φλόριντα της οποίας η παραγωγή καταναλώνεται κατά 90% εγχώρια και μόνο το 10% εξάγεται.

Η Ευρωπαϊκή Ένωση αντιθέτως είναι ο μεγαλύτερος εισαγωγέας χυμών πορτοκαλιού (80%) και ακολουθούν ο Καναδάς και η Ιαπωνία. Η Ε.Ε. εισάγει κυρίως από τη Βραζιλία ενώ οι Η.Π.Α. και ο Καναδάς καταναλώνουν κυρίως από την παραγωγή της Φλόριντα.

Οι χυμοί φρούτων καταναλώνονται κυρίως στις βιομηχανοποιημένες χώρες, όχι μόνο γιατί οι καταναλωτές σε αυτές τις χώρες έχουν υψηλό εισόδημα αλλά και επειδή έχουν αυξημένες ανησυχίες για την υγιεινή διατροφή.

Σκοπός λοιπόν της παρούσας εργασίας είναι η στατιστική ανάλυση ορισμένων μεταβλητών που ενδεχομένως επηρεάζουν τις πωλήσεις φυσικών χυμών πορτοκαλιού και η εξαγωγή συμπερασμάτων για τον τρόπο που αυτές διαμορφώνονται.

ΚΕΦΑΛΑΙΟ 2: ΜΕΘΟΔΟΛΟΓΙΕΣ

Σε αυτό το κεφάλαιο θα αναφερθούν οι βασικές έννοιες και μεθοδολογίες της Στατιστικής, οι οποίες θα χρησιμοποιηθούν στο επόμενο κεφάλαιο για την ανάλυση και την εξαγωγή συμπερασμάτων. Αυτές είναι οι μεθοδολογίες της περιγραφικής στατιστικής, η γραμμική παλινδρόμηση και η ανάλυση διακύμανσης.

2.1 ΠΕΡΙΓΡΑΦΙΚΗ ΣΤΑΤΙΣΤΙΚΗ – ΠΕΡΙΓΡΑΦΙΚΑ ΜΕΤΡΑ ΔΕΔΟΜΕΝΩΝ

Οι μεταβλητές μιας στατιστικής έρευνας αποτελούνται συνήθως από ένα μεγάλο πλήθος δεδομένων. Δεδομένα εννοούμε τις αριθμητικές πληροφορίες που συλλέγουμε και στη συνέχεια επεξεργαζόμαστε για να πάρουμε μια απόφαση ή να εξάγουμε κάποιο συμπέρασμα.
Για να παρουσιάσουμε το δείγμα μας, να εντοπίσουμε χαρακτηριστικά των δεδομένων και να εξάγουμε κάποια αρχικά συμπεράσματα, οργανώνουμε τα δεδομένα μας αρχικά σε πίνακες και στη συνέχεια χρησιμοποιούμε γραφικές και αριθμητικές μεθόδους.

Οι γραφικές μέθοδοι παρουσίασης περιλαμβάνουν το ιστόγραμμα, το διάγραμμα διασποράς, το θηκόγραμμα, το πολύγωνο συχνοτήτων κ.ά., ενώ στις αριθμητικές μεθόδους έχουμε τους πίνακες συχνοτήτων, τα μέτρα θέσης, τα μέτρα διασποράς, τα μέτρα σχήματος κ.ά. Από αυτά θα αναλύσουμε και θα χρησιμοποιήσουμε τα παρακάτω:

ΜΕΤΡΑ ΘΕΣΗΣ (ή ΚΕΝΤΡΙΚΗΣ ΤΑΣΗΣ)

Είναι τα μέτρα εκείνα που προσδιορίζουν ένα κεντρικό σημείο γύρω από το οποίο τείνουν να συγκεντρώνονται τα δεδομένα. Έτσι, για κάθε συγκεκριμένο σύνολο δεδομένων είναι δυνατόν να επιλεγεί κάποια τυπική τιμή ή μέσο που θα «αντιπροσωπεύει» τη συμπεριφορά των τιμών.
Έστω η ποσοτική μεταβλητή Χ η οποία σε τυχαίο δείγμα μεγέθους n έδωσε τις παρατηρήσεις x_1, …, x_n.
Οι τρεις συνηθέστεροι τρόποι μέτρησης της κεντρικής τάσης ενός δείγματος είναι:

- η δειγματική Μέση Τιμή (sample mean value), που ορίζεται ως το άθροισμα όλων των παρατηρήσεων διαιρεμένο με το πλήθος των παρατηρήσεων

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

- η δειγματική Διάμεσος (sample median), που ορίζεται ως η κεντρική τιμή στις κατά αύξουσα σειρά διατεταγμένες παρατηρήσεις,
- η Κορυφή που είναι η παρατήρηση με τη μεγαλύτερη συχνότητα εμφάνισης.

Η δειγματική μέση τιμή είναι το σημαντικότερο εκ των τριών και θα τη χρησιμοποιήσουμε στη στατιστική συμπερασματολογία (σε επόμενο κεφάλαιο) για να βγάλουμε συμπεράσματα για τη μέση τιμή του πληθυσμού. Για τον υπολογισμό της μέσης τιμής χρησιμοποιούνται όλες οι τιμές του δείγματος, ενώ για τη διάμεσο μόνο η τάξη τους. Γι αυτό και η μέση τιμή επηρεάζεται από μακρινές τιμές αλλά η διάμεσος όχι.

ΜΕΤΡΑ ΔΙΑΣΠΟΡΑΣ

Ως μέτρα διασποράς (ή μεταβλητότητας) εννοούμε τα μέτρα εκείνα που μας δείχνουν πόσο συγκεντρωμένες είναι οι παρατηρήσεις γύρω από την κεντρική τιμή, δηλαδή προσδιορίζουν τη μεταβλητότητα που παρουσιάζει ένα σύνολο δεδομένων. Τα κυριότερα μέτρα διασποράς ενός δείγματος είναι:

- το δειγματικό Εύρος, που ορίζεται ως η διαφορά της ελάχιστης από τη μέγιστη τιμή του δείγματος,
- η δειγματική Διακύμανση, που μετράει τη μεταβλητότητα των παρατηρήσεων γύρω από τη Μέση Τιμή και ορίζεται από τον τύπο

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

- η δειγματική Τυπική Απόκλιση, που ορίζεται ως $s = \sqrt{s^2}$,

- το Ενδοτεταρτημοριακό Εύρος, που ορίζεται ως η διαφορά IQR = $Q_3 - Q_1$, όπου Q1: το σημείο κάτω από το οποίο βρίσκεται το 25% των δεδομένων και Q3: το σημείο κάτω από το οποίο βρίσκεται το 75% των δεδομένων.

ΙΣΤΟΓΡΑΜΜΑ

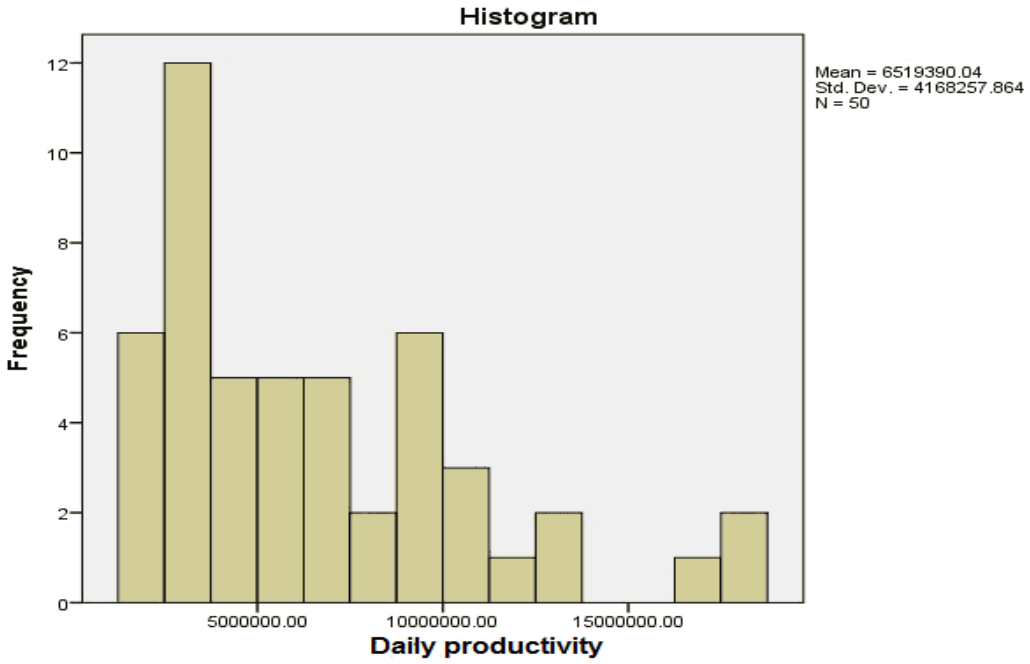
Τις περισσότερες φορές τα δεδομένα που διαθέτουμε χρειάζεται να ομαδοποιηθούν, δηλαδή να οριστεί μία ομάδα δεδομένων (κλάση) τα οποία βρίσκονται μέσα σε συγκεκριμένα όρια. Έχοντας ομαδοποιήσει τα αριθμητικά δεδομένα μπορούμε να τα παραστήσουμε γραφικά σε ένα ιστόγραμμα το οποίο χρησιμοποιείται για την απεικόνιση των κατανομών συχνοτήτων των δεδομένων.

Το ιστόγραμμα είναι ένα διάγραμμα με κάθετες στήλες που έχουν ως βάση τα διαστήματα που έχουν οριστεί ως κλάσεις και ύψος ανάλογο με τον αριθμό (συχνότητα) των παρατηρήσεων που ανήκουν στα διαστήματα αυτά. Δηλαδή, στον κάθετο άξονα του ιστογράμματος μπορούμε να τοποθετήσουμε τη συχνότητα f_i ή το ποσοστό p_i ή τα αθροιστικά τους αντίστοιχα εάν μας ενδιαφέρει το πλήθος ή το ποσοστό των παρατηρήσεων, που οι τιμές τους είναι μικρότερες ή ίσες ορισμένης τιμής x_i.

Ένα μέτρο που χαρακτηρίζει τις κατανομές συχνοτήτων, άρα και την παρουσίαση αυτών στο ιστόγραμμα, είναι η λοξότητα ή ασυμμετρία. Εάν τα δεδομένα κατανέμονται συμμετρικά, τότε όσο απομακρυνόμαστε από τη μέση τιμή είτε προς τα πάνω είναι προς τα κάτω συναντάμε περίπου τον ίδιο αριθμό παρατηρήσεων

Για ένα πλήθος δεδομένων έχουμε:

- Δεξιά λοξότητα ή θετική ασυμμετρία όταν: μέση τιμή > διάμεσος > κορυφή
- Αριστερή λοξότητα ή αρνητική ασυμμετρία όταν: μέση τιμή < διάμεσος < κορυφή
- Μηδενική ασυμμετρία (συμμετρία) όταν: μέση τιμή ≈ διάμεσος ≈ κορυφή



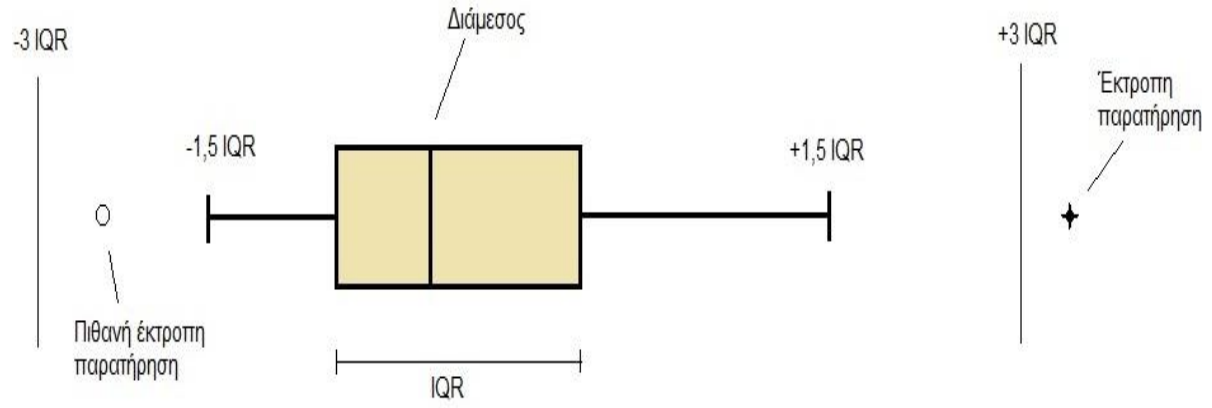
Σχήμα 1: Παράδειγμα ιστογράμματος με κατανομή δεξιάς λοξότητας / θετική ασυμμετρία – συχνότητα ημερήσιας παραγωγής

ΘΗΚΟΓΡΑΜΜΑ

Το θηκόγραμμα είναι ένας απλός τρόπος οπτικής των δεδομένων που συνοψίζει την κεντρική τάση (διάμεσος), τη μεταβλητότητα (ενδοτεταρτημοριακό εύρος) τις ακραίες τιμές καθώς και τη σχηματική μορφή (λοξότητα) της κατανομής.

Για την κατασκευή του βρίσκουμε αρχικά τη διάμεσο και τα σημεία Q_1 και Q_3. Στη συνέχεια κατασκευάζουμε ένα ορθογώνιο μήκους IQR, που μας δίνει το κεντρικό διάστημα με το 50% των παρατηρήσεων, και πλάτους αυθαίρετης επιλογής και τοποθετούμε μέσα σε αυτό τη διάμεσο. Οι απολήξεις του θηκογράμματος κατασκευάζονται με την έκταση της γραμμής από το σημείο Q_3 έως τη μεγαλύτερη παρατήρηση και από το Q_1 έως τη μικρότερη παρατήρηση, μόνο εάν οι παρατηρήσεις αυτές βρίσκονται εντός μιας απόστασης που ισούται με 1,5 IQR (εσωτερικά όρια). Εάν μία ή περισσότερες παρατηρήσεις είναι μακρύτερα από αυτή την απόσταση σημειώνονται ως πιθανές έκτροπες παρατηρήσεις. Εάν οι παρατηρήσεις αυτές

βρίσκονται σε απόσταση μεγαλύτερη από 3 IQR (εξωτερικά όρια) από την αντίστοιχη ακμή, τότε σημειώνονται ως έκτροπες παρατηρήσεις.



Σχήμα 2: Θηκόγραμμα

Η χρησιμότητα του θηκογράμματος συνοψίζεται στα παρακάτω:

- για τον εντοπισμό της μεταβλητότητας των δεδομένων με βάση το μήκος του πλαισίου (IQR) και το μήκος των απολήξεων
- για τον εντοπισμό πιθανής λοξότητας της κατανομής του συνόλου των δεδομένων.
- για τον εντοπισμό πιθανών έκτροπων παρατηρήσεων και έκτροπων παρατηρήσεων
- για τη σύγκριση δύο ή περισσότερων συνόλων δεδομένων

2.2 ΕΛΕΓΧΟΙ ΥΠΟΘΕΣΕΩΝ

Ως μηδενική υπόθεση καλούμε έναν ισχυρισμό για την τιμή μιας πληθυσμιακής παραμέτρου. Ο ισχυρισμός αυτός θεωρείται αληθής εκτός και αν αποδειχθεί στατιστικά ότι δεν ισχύει. Η απόρριψη της μηδενικής υπόθεσης λέγεται εναλλακτική υπόθεση. Για παράδειγμα, εάν η πληθυσμιακή παράμετρος που μας ενδιαφέρει είναι η μέση τιμή και θεωρούμε ότι ισούται με 5, έχουμε συμβολικά:

$\mathbf{H_0 : \mu = 5}$ (μηδενική υπόθεση)

$\mathbf{H_1 : \mu \neq 5}$ (εναλλακτική υπόθεση)

Άλλοι ισχυρισμοί μπορεί να αφορούν τη μεγαλύτερη ή μικρότερη τιμή μιας παραμέτρου:

$\mathbf{H_0 : \mu \geq 5}$ (μηδενική υπόθεση)

$\mathbf{H_1 : \mu < 5}$ (εναλλακτική υπόθεση)

Εάν εξαχθεί συμπέρασμα επιβεβαίωσης της μηδενικής υπόθεσης, για να αξιολογηθεί η αξιοπιστία του συμπεράσματος αυτού χρησιμοποιείται η τιμή p που δηλώνει την αξιοπιστία της μηδενικής υπόθεσης. Όταν για καθορισμένο επίπεδο σημαντικότητας α η τιμή $p<\alpha$, τότε η H_0 απορρίπτεται, ενώ στην αντίθετη περίπτωση δεν μπορεί να απορριφθεί.

Η τιμή p προσδιορίζεται χρησιμοποιώντας συναρτήσεις ελέγχου και τυποποιημένους πίνακες. Στην περίπτωση που η μηδενική υπόθεση έχει τη μορφή:

- $\mu \geq \chi$, ο έλεγχος είναι αριστερόπλευρος γιατί η περιοχή απόρριψης προκύπτει στην αριστερή πλευρά του διαγράμματος κατανομής, δηλαδή το $\bar{X}$ είναι σημαντικά μικρότερο της τιμής χ
- $\mu \leq \chi$, ο έλεγχος είναι δεξιόπλευρος γιατί η περιοχή απόρριψης προκύπτει στη δεξιά πλευρά του διαγράμματος κατανομής, δηλαδή το $\bar{X}$ είναι σημαντικά μεγαλύτερο της τιμής χ
- $\mu = \chi$, ο έλεγχος είναι δίπλευρος γιατί η περιοχή απόρριψης προκύπτει και στις δύο πλευρές του διαγράμματος κατανομής.

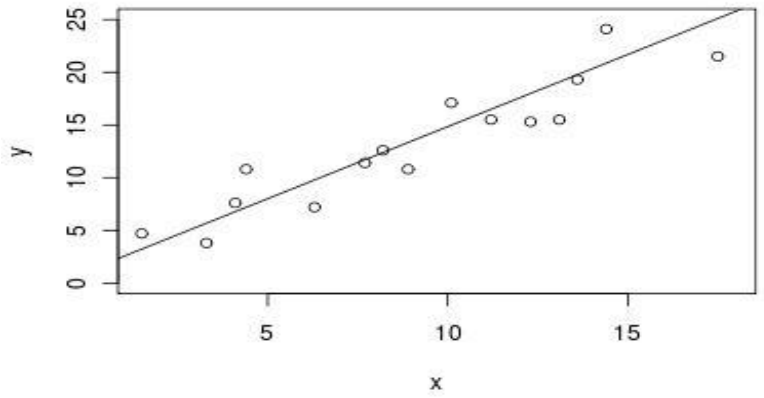
2.3 ΠΑΛΙΝΔΡΟΜΗΣΗ

Ο κλάδος της Στατιστικής που εξετάζει εάν υπάρχει στατιστική σχέση δύο ή περισσότερων μεταβλητών και στη συνέχεια περιγράφει τη σχέση αυτή λέγεται ανάλυση παλινδρόμησης (regression analysis). Η παλινδρόμηση στην οποία έχουμε μια μόνο ανεξάρτητη μεταβλητή ονομάζεται απλή παλινδρόμηση, ενώ αν έχουμε περισσότερες από μια ανεξάρτητες μεταβλητές ονομάζεται πολλαπλή παλινδρόμηση. Εξαρτημένη είναι η μεταβλητή της οποίας τις μεταβολές θέλουμε να ερμηνεύσουμε, ενώ ανεξάρτητη είναι η μεταβλητή που πιστεύουμε ότι επιδρά στην εξαρτημένη, προκαλεί τις μεταβολές της και επομένως χρησιμοποιείται για να ερμηνεύσουμε τη μεταβλητότητα της εξαρτημένης.

Γραμμική Παλινδρόμηση

Η απλούστερη περίπτωση απλής παλινδρόμησης είναι η απλή γραμμική παλινδρόμηση κατά την οποία η μεταβλητή Y προσεγγίζεται ικανοποιητικά από μια γραμμική συνάρτηση του X.

Στην απλή γραμμική παλινδρόμηση μοντελοποιείται η σχέση μεταξύ της εξαρτημένης μεταβλητής, που συμβολίζεται με Y, και της ανεξάρτητης μεταβλητής, που συμβολίζεται με X. Για να συμβεί αυτό κατασκευάζεται διάγραμμα με διάσπαρτα σημεία (διάγραμμα διασποράς), που είναι τυχαία επιλεγμένες παρατηρήσεις των X και Y, και προσαρμόζεται σε αυτό μια ευθεία γραμμή που περιγράφει τη σχέση των δύο μεταβλητών X και Y.



Σχήμα 3: Γραφική απεικόνιση γραμμικής παλινδρόμησης (διάγραμμα διασποράς)

Λόγω της αβεβαιότητας που διέπει τη ρεαλιστική πραγματικότητα και λόγω του πλήθους των παραγόντων που επηρεάζουν τη διαδικασία εξαγωγής των δεδομένων, το μοντέλο που περιγράψαμε εμπεριέχει κάποια σφάλματα. Επομένως, το μοντέλο εμπεριέχει μια συστηματική συνιστώσα και μια τυχαία συνιστώσα, αυτή των σφαλμάτων. Έτσι, το μοντέλο της απλής γραμμικής παλινδρόμησης ενός πληθυσμού δίνεται από τη σχέση:

$$Y = \beta_0 + \beta_1 X + \varepsilon,$$

όπου β_0: η σταθερά του πληθυσμού, β_1: η κλίση του πληθυσμού και ε: σφάλμα, δηλαδή η διαφορά μεταξύ της πραγματικής τιμής της Y και της τιμής της πρόβλεψης που προκύπτει από το μοντέλο.

Η παραπάνω σχέση περιγράφει τον πληθυσμό, δηλαδή την υποθετική πληθυσμιακή σχέση των μεταβλητών X και Y. Αντλώντας λοιπόν δειγματικά δεδομένα η σχέση αυτή θα προσδιοριστεί εκτιμώντας τα β_0 και β_1. Έτσι για κάθε σημείο δεδομένων ξεχωριστά προκύπτει η εκτιμώμενη εξίσωση παλινδρόμησης:

$$y_i = b_0 + b_1 x_i + e_i$$

Εδώ έχουμε το b_0 ως εκτιμητή του β_0, το b_1 ως εκτιμητή του β_1 και το e_i ως εκτιμήσεις των πραγματικών σφαλμάτων του πληθυσμού ε_i.

Τα σφάλματα e_i είναι η απόσταση του κάθε σημείου δεδομένων από την προσαρμοσμένη γραμμή παλινδρόμησης.

Η ίδια η προσαρμοσμένη γραμμή παλινδρόμησης περιγράφεται από την εξίσωση:

$$\widehat{\mathbf{Y}} = \mathbf{b_0} + \mathbf{b_1 X}$$

όπου $\widehat{Y}$ είναι η τιμή της Y που βρίσκεται στην προσαρμοσμένη γραμμή παλινδρόμησης για ένα δοθέν X.

Το μοντέλο της απλής γραμμικής παλινδρόμησης βασίζεται σε τρεις υποθέσεις:

1. Η σχέση των X και Y είναι μια σχέση ευθείας γραμμής.
2. Οι τιμές της ανεξάρτητης μεταβλητής X θεωρούνται σταθερές (όχι τυχαίες). Η μόνη τυχαιότητα στις τιμές της Y προέρχεται από τον όρο του σφάλματος *ε*.
3. Τα σφάλματα *ε* είναι κανονικά κατανεμημένα με μέση τιμή 0 (μηδέν) και σταθερή διακύμανση σ^2. Τα σφάλματα είναι ασυσχέτιστα μεταξύ τους σε διαδοχικές παρατηρήσεις.

Για τον προσδιορισμό της παραπάνω εξίσωσης και την εκτίμηση των παραμέτρων b_0 και b_1 η συχνότερα χρησιμοποιούμενη μέθοδος είναι η μέθοδος των ελαχίστων τετραγώνων.

Σκοπός της μεθόδου αυτής είναι να βρεθεί η γραμμή που ελαχιστοποιεί τα σφάλματα. Επειδή τα σφάλματα βρίσκονται εκατέρωθεν της γραμμής παλινδρόμησης πρέπει να βρεθεί η γραμμή που ελαχιστοποιεί το άθροισμα των τετραγώνων των σφαλμάτων.

Το άθροισμα των τετραγώνων των σφαλμάτων (SSE) δίνεται από την παρακάτω σχέση:

$$\mathrm{SSE} = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

Οι τιμές b_0 και b_1 που ελαχιστοποιούν το άθροισμα αυτό δίνονται από τις σχέσεις:

$$\boxed{b_1 = \frac{SS_{xy}}{SS_x} \quad \text{και} \quad b_0 = \bar{y} - b_1\bar{x}}$$

Όπου $SS_x = \sum(x - \bar{x})^2$ και $SS_{xy} = \sum(x - \bar{x})(y - \bar{y})$

Για τις τιμές αυτές των εκτιμητών b_0 και b_1ορίζονται οι αντίστοιχες τυπικές αποκλίσεις:

$$s(b_0) = \frac{s\sqrt{\sum x^2}}{\sqrt{n\,SS_x}} \quad \text{και} \quad s(b_1) = \frac{s}{\sqrt{SS_x}},$$

όπου $s = \sqrt{\frac{SSE}{n-2}}$.

Μια άλλη σημαντική έννοια είναι αυτή της συσχέτισης. Μετράται με το συντελεστή συσχέτισης **ρ** και εκφράζει το βαθμό της γραμμικής σχέσης μεταξύ των δύο μεταβλητών X και Y.

Ο συντελεστής συσχέτισης **ρ** παίρνει τιμές από -1 έως 1, με την τιμή μηδέν να υποδηλώνει πως δεν υπάρχει γραμμική συσχέτιση, ενώ οι τιμές -1 και 1 σημαίνουν τέλεια γραμμική συσχέτιση, αρνητική και θετική, αντίστοιχα. Όλες οι ενδιάμεσες τιμές φανερώνουν σε τι βαθμό η σχέση μεταξύ δύο μεταβλητών προσεγγίζει τη γραμμική.

Ο συντελεστής συσχέτισης **ρ** αφορά έναν πληθυσμό, επομένως στην πράξη χρησιμοποιείται η εκτίμηση του **ρ** που ονομάζεται δειγματικός συντλεστής συσχέτισης και προκύπτει από δεδομένα ενός δείγματος του πληθυσμού.

Ο δειγματικός συντελεστής συσχέτισης δίνεται από τη σχέση:

$$r = \frac{SS_{XY}}{\sqrt{SS_X\, SS_Y}}$$

Για τις τιμές του r ισχύει ό,τι και για το **ρ.**

Επιπλέον, το r δύναται να χρησιμοποιηθεί για τον έλεγχο υποθέσεων ως προς το ρ και συγκεκριμένα ως προς την ύπαρξη συσχέτισης μεταξύ δύο μεταβλητών X και Y. Ο έλεγχος υποθέσεων διατυπώνεται ως εξής:

$$H_0 : \rho = 0$$

$$H_1 : \rho \neq 0$$

Η στατιστική συνάρτηση ελέγχου τότε είναι:

$$t_{(n-2)} = \frac{r}{\sqrt{(1 - r^2)/(n - 2)}}$$

όπου n: το μέγεθος του δείγματος.

Αφού λοιπόν υπολογιστεί η τιμή $t_{(n-2)}$, συγκρίνοντάς την με την ευρεθείσα από τους αντίστοιχους πίνακες τιμή $t_{(n-2)}$, για καθορισμένο επίπεδο σημαντικότητας **α**, ελέγχουμε την παραπάνω μηδενική υπόθεση.

- εάν $t_{(n-2)} > |t_\alpha|$ τότε η μηδενική υπόθεση δεν μπορεί να απορριφθεί
- εάν $t_{(n-2)} < |t_\alpha|$ η μηδενική υπόθεση απορρίπτεται και οι δύο μεταβλητές συσχετίζονται γραμμικά

Χρησιμοποιώντας την ίδια στατιστική συνάρτηση ελέγχου για τον έλεγχο H_0: $\rho \geq 0$, H_1: $\rho < 0$, μπορούμε να ελέγξουμε την ύπαρξη θετικής συσχέτισης

Το αντίστοιχο με το συντελεστή συσχέτισης μέτρο που προσδιορίζει το πόσο κοντά στα δεδομένα προσαρμόζεται η γραμμή παλινδρόμησης είναι ο συντελεστής προσδιορισμού που ισούται με το τετράγωνο του δειγματικού συντελεστή συσχέτισης.

Ο συντελεστής προσδιορισμού r^2 είναι δηλαδή ένα μέτρο της ισχύος της εξίσωσης παλινδρόμησης. Παίρνει τιμές από 0 έως 1 και μετρά το πόσο καλά προσαρμόζεται η γραμμή παλινδρόμησης στα δεδομένα. Μπορούμε να πούμε επίσης ότι η τιμή του r^2 μας δείχνει το ποσοστό της μεταβλητότητας των δεδομένων που ερμηνεύεται από τη γραμμή παλινδρόμησης. Για παράδειγμα, όταν $r^2 = 0,7$ τότε η παλινδρόμηση ερμηνεύει το 70% της μεταβλητότητας των δεδομένων.

Για τον υπολογισμός του συντελεστή προσδιορισμού υπολογίζονται πρώτα το άθροισμα τετραγώνων για το σφάλμα (SSE) και το άθροισμα τετραγώνων για την παλινδρόμηση (SSR) από τις σχέσεις:

$$\mathrm{SSE} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$$\mathrm{SSR} = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$$

Προκύπτει έτσι το συνολικό άθροισμα τετραγώνων $\mathrm{SST} = \mathrm{SSE} + \mathrm{SSR}$

Το SSR ονομάζεται και ερμηνευμένη μεταβλητότητα, δηλαδή η μεταβλητότητα της Y που ερμηνεύεται από τη σχέση της με το X. Το SSE αντιστοίχως είναι η ανερμήνευτη μεταβλητότητα λόγω της ύπαρξης σφαλμάτων.

Ο συντελεστής προσδιορισμού δίνεται από την παρακάτω σχέση:

$$r^2 = \frac{\mathrm{SSR}}{\mathrm{SST}} = 1 - \frac{\mathrm{SSE}}{\mathrm{SST}}$$

Έχοντας υπολογίσει τη γραμμή παλινδρόμησης, μια από τις δυνατότητες που μας δίνονται είναι αυτή της πρόβλεψης**.** Πρόβλεψη είναι η εκτίμηση ενός αποτελέσματος που αναμένεται να

προκύψει σε μία άλλη κατάσταση από αυτήν για την οποία διαθέτουμε δεδομένα. Είναι δηλαδή η εκτίμηση των τιμών της εξαρτημένης μεταβλητής για διαφορετικές τιμές της ανεξάρτητης μεταβλητής με χρήση της εξίσωσης παλινδρόμησης $\hat{Y} = b_0 + b_1 X$, που για αυτό το σκοπό καλείται και εξίσωση πρόβλεψης. Για να γίνει αυτό, αντικαθιστούμε στην εξίσωση παλινδρόμησης την τιμή της ανεξάρτητης μεταβλητής X για την οποία θέλουμε να προβλέψουμε την τιμή της εξαρτημένης μεταβλητής Y που θα προκύψει.

Αυτές οι προβλέψεις δεν είναι ακριβείς και εξαρτώνται από το σφάλμα που οφείλεται στη μεταβλητότητα των σημείων γύρω από τη γραμμή παλινδρόμησης. Για το λόγο αυτό ορίζουμε ένα (1-α)100% διάστημα πρόβλεψης για την Y:

$$\hat{y} \pm t_{\alpha/2,(n-2)}\, s \sqrt{1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{SS_x}}$$

Πολλαπλή Παλινδρόμηση

Επέκταση της απλής γραμμικής παλινδρόμησης για περισσότερες από δύο μεταβλητές είναι η πολλαπλή γραμμική παλινδρόμηση. Σκοπός της πολλαπλής παλινδρόμησης είναι να περιγράψει τη σχέση μεταξύ της εξαρτημένης μεταβλητής Y και των k ανεξάρτητων μεταβλητών X_1, X_2, … , X_k. Το μοντέλο αυτό έχει τη μορφή:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \varepsilon,$$

όπου β_0: μία σταθερά, β_1, β_2,…,β_k: οι συντελεστές παλινδρόμησης που περιγράφουν την επίδραση των ανεξάρτητων μεταβλητών, ε: σφάλμα.

Οι συντελεστές παλινδρόμησης (αλλιώς: παράμετροι κλίσης) προσδιορίζουν τη μεταβολή της αναμενόμενης τιμής της Y σε κάθε μοναδιαία μεταβολή της X_i όταν όλες οι άλλες μεταβλητές διατηρούνται σταθερές.

Το μοντέλο αυτό βασίζεται στις εξής υποθέσεις όσον αφορά τα σφάλματα ε_i:

1. Για κάθε παρατήρηση ο όρος του σφάλματος ε έχει κανονική κατανομή, μέση τιμή μηδέν και τυπική απόκλιση σ και είναι ανεξάρτητος από τους όρους του σφάλματος των άλλων παρατηρήσεων.
2. Οι μεταβλητές X_j θεωρούνται σταθερές ποσότητες και ανεξάρτητες του όρου σφάλματος ε.

Αντιστοίχως με την απλή γραμμική παλινδρόμηση, οι εκτιμήσεις από τα δεδομένα του δείγματος των β_0, β_1, β_2,…,β_k συμβολίζονται με b_0, b_1, b_2,…b_k, αντίστοιχα. Έτσι προκύπτει η εκτιμώμενη εξίσωση παλινδρόμησης:

$$\hat{\mathbf{Y}} = \mathbf{b_0} + \mathbf{b_1 X_1} + \mathbf{b_2 X_2} + \ldots + \mathbf{b_k X_k} ,$$

όπου b_0, b_1, b_2,…b_k :οι συντελεστές μερικής παλινδρόμησης.

Με τον ίδιο τρόπο όπως και στην απλή παλινδρόμηση δύνανται να γίνουν προβλέψεις της Y για τις τιμές των X_j που θέτουμε. Ο υπολογισμός του διαστήματος πρόβλεψης στην περίπτωση της πολλαπλής παλινδρόμησης είναι αρκετά περίπλοκος και για την εξαγωγή του χρησιμοποιείται υπολογιστής.

2.4 ΑΝΑΛΥΣΗ ΔΙΑΚΥΜΑΝΣΗΣ (ANOVA)

Η ανάλυση διακύμανσης είναι μια μέθοδος πειραματικού σχεδιασμού που έχει ως σκοπό την ανίχνευση διαφορών στους μέσους πολλών πληθυσμών, προσδιορίζοντας τις διακυμάνσεις στα δειγματικά δεδομένα. Τα δειγματικά δεδομένα περιέχουν ένα μεγάλο αριθμό πηγών διακύμανσης. Με την ανάλυση διακύμανσης καθορίζονται οι πηγές διακύμανσης και τα ποσοστά της διακύμανσης που αποδίδονται σε κάθε μία από αυτές. Στην πράξη, η Ανάλυση Διακύμανσης χρησιμοποιείται για να επιβεβαιώσει ή όχι αν υπάρχει στατιστικά σημαντική διαφορά μεταξύ δύο ή περισσότερων πληθυσμών.

Στην ανάλυση διακύμανσης έχουμε τον εξής έλεγχο υπόθεσης:

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \ldots = \mu_k$$

H_1: τουλάχιστον δύο μέσοι διαφέρουν

Όταν ισχύει η μηδενική υπόθεση, η στατιστική συνάρτηση ελέγχου της ανάλυσης διακύμανσης ακολουθεί την F κατανομή. Από την τιμή του στατιστικού F για δεδομένο επίπεδο σημαντικότητας αποφασίζεται αν οι μέσοι των πληθυσμών είναι ίσοι.

Για να χρησιμοποιηθεί η ανάλυση διακύμανσης για τον παραπάνω έλεγχο πρέπει να ικανοποιούνται δύο υποθέσεις:

1. Για κάθε έναν από τους πληθυσμούς τα δείγματα που συλλέγονται πρέπει να είναι τυχαία και ανεξάρτητα
2. Οι k πληθυσμοί είναι κανονικά κατανεμημένοι με ίσες διακυμάνσεις σ^2.

Αν οι πληθυσμοί δεν ακολουθούν ακριβώς την κανονική κατανομή αλλά την προσεγγίζουν, η ανάλυση διακύμανσης συνεχίζει να δίνει ικανοποιητικά αποτελέσματα.

Όταν λοιπόν ισχύει η μηδενική υπόθεση η στατιστική συνάρτηση ελέγχου της ανάλυσης διακύμανσης ακολουθεί την κατανομή F.

Ο υπολογισμός της τιμής της στατιστικής συνάρτησης ελέγχου F είναι πολύπλοκος και δεν κρίνεται σκόπιμη η επεξήγησή του, καθώς πλέον η τιμή αυτή εξάγεται εύκολα με τη χρήση Η/Υ (excel, SPSS κ.λπ.). Συγκεκριμένα εξάγεται ο παρακάτω πίνακας ANOVA:

Πηγή μεταβλητότητας	**Άθροισμα τετραγώνων**	**Βαθμοί ελευθερίας**	**Μέσο τετράγωνο**	**F**
Μεταξύ των δειγμάτων	SSTR	k-1	MSTR=SSTR/(k-1)	**F = MSTR/MSE**
Εντός των δειγμάτων	SSE	n-k	MSE = SSE/(n-k)	
Σύνολο	SST = SSTR + SSE	n-1		

Όπου:

SSTR: το άθροισμα τετραγώνων των αποκλίσεων των μέσων των δειγμάτων από το γενικό μέσο

SSE: το άθροισμα τετραγώνων των αποκλίσεων των παρατηρήσεων κάθε δείγματος από το μέσο τους (σφάλματα)

k: ο αριθμός των πληθυσμών

n: το μέγεθος του συνολικού δείγματος

k-1: βαθμοί ελευθερίας αριθμητή

n-k: βαθμοί ελευθερίας παρονομαστή

MSTR: μέσο του αθροίσματος τετραγωνικών αποκλίσεων μεταξύ των δειγμάτων

MSE: μέσο του αθροίσματος τετραγωνικών αποκλίσεων εντός των δειγμάτων (μέσο τετραγωνικό σφάλμα)

Αφού λοιπόν υπολογιστεί/εξαχθεί η τιμή F, συγκρίνοντάς την με την ευρεθείσα από τους αντίστοιχους πίνακες τιμή $F_{(k-1,\ n-k)}$, για καθορισμένο επίπεδο σημαντικότητας **α**, ελέγχουμε την παραπάνω μηδενική υπόθεση. **Εάν η τιμή F που υπολογίσαμε είναι μικρότερη από την αντίστοιχη τιμή του πίνακα $F_{(k-1,\ n-k)}$τότε η μηδενική υπόθεση δεν μπορεί να απορριφθεί, δεν μπορούμε να ισχυριστούμε δηλαδή ότι οι μέσοι των πληθυσμών διαφέρουν. Ενώ εαν η τιμή F είναι μεγαλύτερη από την αντίστοιχη τιμή του πίνακα $F_{(k-1,\ n-k)}$ τότε η μηδενική υπόθεση απορρίπτεται, δηλαδή υπάρχει τουλάχιστον ένας μέσος πληθυσμού που διαφέρει από τους υπόλοιπους.**

Να σημειωθεί ότι για την εύρεση της τιμής σύγκρισης στους πίνακες χρησιμοποιούνται, για δεδομένο επίπεδο σημαντικότητας α, οι βαθμοί ελευθερίας του αριθμητή (k-1) και οι βαθμοί ελευθερίας του παρονομαστή (n-k).

Έλεγχος της μηδενικής υπόθεσης μπορεί να γίνει και απλούστερα, χωρίς τη χρήση πινάκων, με την εξαγωγή της τιμής p, που όπως προαναφέρθηκε είναι ένα είδος αξιολόγησης της αξιοπιστίας της μηδενικής υπόθεσης H_0. Συγκεκριμένα, όταν p>α η μηδενική υπόθεση δεν μπορεί να απορριφθεί και όταν p<α η μηδενική υπόθεση απορρίπτεται.

Στην περίπτωση απόρριψης της μηδενικής υπόθεσης, όταν δηλαδή υπάρχουν ενδείξεις ότι τουλάχιστον ένας από τους πληθυσμιακούς μέσους διαφέρει από τους υπόλοιπους, υπάρχει η δυνατότητα περαιτέρω ανάλυσης με χρήση του κριτηρίου Tukey για να διαπιστωθεί ποιοι από τους πληθυσμιακούς μέσους διαφέρουν σημαντικά.

Συνοπτικά, για τη διενέργεια του ελέγχου Tukey υπολογίζονται οι απόλυτες διαφορές των δειγματικών μέσων ανά δύο. Π.χ. για k=3 υπολογίζονται οι τιμές $|\bar{\chi}_1 - \bar{\chi}_2|$, $|\bar{\chi}_1 - \bar{\chi}_3|$, $|\bar{\chi}_2 - \bar{\chi}_3|$, όπου χ_1, χ_2, χ_3: οι δειγματικοί μέσοι Στη συνέχεια κάθε μία από τις διαφορές αυτές συγκρίνεται με το κριτήριο Tukey $T = q_a \frac{\sqrt{MSE}}{\sqrt{n_i}}$. Η τιμή q_α βρίσκεται από πίνακες για δεδομένο α και για βαθμούς ελευθερίας k και n-k.

Κάθε διαφορά μεγαλύτερη από την τιμή T υποδεικνύει πως οι αντίστοιχοι πληθυσμιακοί μέσοι διαφέρουν σημαντικά, για συγκεκριμένο α.

ΚΕΦΑΛΑΙΟ 3: ΔΕΔΟΜΕΝΑ και ΠΛΑΙΣΙΟ ΑΝΑΛΥΣΗΣ

3.1 ΔΕΔΟΜΕΝΑ

Τα δεδομένα που διαθέτουμε χρησιμοποιήθηκαν από τον Montgomery (1997) για τη μελέτη των τιμολογιακών στρατηγικών σε micro-marketing επίπεδο. Το micro-marketing αναφέρεται στην προσαρμογή των μεταβλητών του μίγματος μάρκετινγκ σε επίπεδο καταστήματος. Ο Montgomery μελέτησε, χρησιμοποιώντας μια σειρά από δεδομένα, το πώς οι τιμές των χυμών, αντί να ακολουθούν μια ομοιόμορφη τιμολογιακή πολιτική, μπορούν να προσαρμοστούν σε κάθε κατάστημα ξεχωριστά με τρόπο επικερδή.

Τα δεδομένα αυτά καλύπτουν τις πωλήσεις χυμών πορτοκαλιού σε 83 καταστήματα της ευρύτερης περιοχής του Σικάγο για διάρκεια 121 εβδομάδων και αφορούν σε 3 διαφορετικά εμπορικά σήματα (Dominicks, MinuteMaid, Tropicana) συσκευασίας 2Lt (64oz). Τα δεδομένα είναι διατεταγμένα σε στήλες που παρουσιάζουν τις πωλήσεις (με τη μορφή λογαρίθμου), το εμπορικό σήμα (εφεξής: brand), την τιμή, την παρουσία/απουσία διαφήμισης καθώς και κάποια δημογραφικά χαρακτηριστικά του καταστήματος.

STORE	Αριθμός καταστήματος
BRAND	Εμπορικό σήμα (Dominicks, MinuteMaid, Tropicana)
WEEK	Εβδομάδα
LOGMOVE	Λογαριθμος των πωληθέντων χυμών 64oz (τεμάχια)
PRICE	Τιμή των χυμών 64oz
FEAT	Χαρακτηριστικό της διαφήμισης
AGE60	Ποσοστό ατόμων άνω των 60 ετών
EDUC	Ποσοστό πτυχιούχων ατόμων
ETHNIC	Ποσοστό έγχρωμων/ισπανόφωνων
INCOME	Λογάριθμος του μέσου εισοδήματος
HHLARGE	Ποσοστό νοικοκυριών 5 ή περισσότερων ατόμων
WORKWOM	Ποσοστό εργαζόμενων γυναικών πλήρους απασχόλησης
HVAL150	Ποσοστό νοικοκυριών με περιουσιακά στοιχεία > $150.000

3.2 ΠΛΑΙΣΙΟ ΑΝΑΛΥΣΗΣ

Αρχικά αποτυπώνονται οι κατανομές των πωλήσεων των τριών brands με χρήση ιστογραμμάτων και θηκογραμμάτων. Στη συνέχεια μελετώνται:

- Η επίδραση της τιμής στις πωλήσεις για τα τρία brands: η τιμή είναι ένας από τους κύριους παράγοντες που επηρεάζουν τη συμπεριφορά των αγοραστών ως προς την ποσότητα προϊόντος που αγοράζουν. Σε πολλές περιπτώσεις μάλιστα η τιμή είναι η πιο σημαντική μεταβλητή στην αγοραστική απόφαση αφού μπορεί να λειτουργήσει είτε ως κίνητρο είτε ως ανασταλτικός παράγοντας.
- Η επίδραση της διαφήμισης στις πωλήσεις για τα τρία brands: η διαφήμιση είναι μία από τις δραστηριότητες που συμβάλλουν στη δημιουργία προϋποθέσεων πειθούς του καταναλωτή και συντελεί επομένως στην τόνωση της ζήτησης και, μακροχρόνια, στην αύξηση των κερδών.
- Οι συσχετίσεις των δημογραφικών παραγόντων: οι δημογραφικοί παράγοντες παίζουν μείζονα ρόλο κατά τη διαμόρφωση των στρατηγικών μάρκετινγκ (τμηματοποίηση της αγοράς, τοποθέτηση, μίγμα μάρκετινγκ κ.λπ.). Σκοπός είναι να εντοπιστούν οι πιο ισχυρές συσχετίσεις και οι παράγοντες αυτοί να αναλυθούν περαιτέρω για κάθε brand ξεχωριστά σε σχέση με τις πωλήσεις.
- Η ταυτόχρονη επίδραση τιμής, διαφήμισης, brand και δημογραφικών παραγόντων.

Μετά από όλα τα παραπάνω επιχειρείται η δημιουργία μοντέλου που θα περιγράφει ικανοποιητικά τις πωλήσεις φυσικών χυμών και θα επιτρέπει την πρόβλεψη πωλήσεων, δραστηριότητα που αποτελεί σημείο εκκίνησης για ένα μεγάλο εύρος επιχειρησιακών διεργασιών.

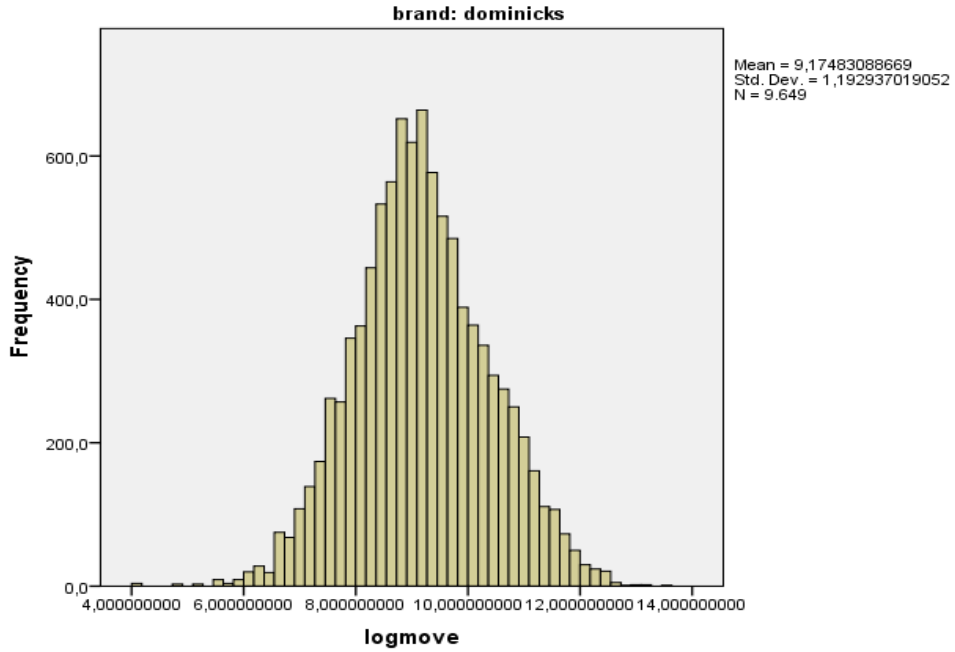
ΚΕΦΑΛΑΙΟ 4: ΣΤΑΤΙΣΤΙΚΗ ΑΝΑΛΥΣΗ

4.1 ΚΑΤΑΝΟΜΕΣ ΔΕΔΟΜΕΝΩΝ

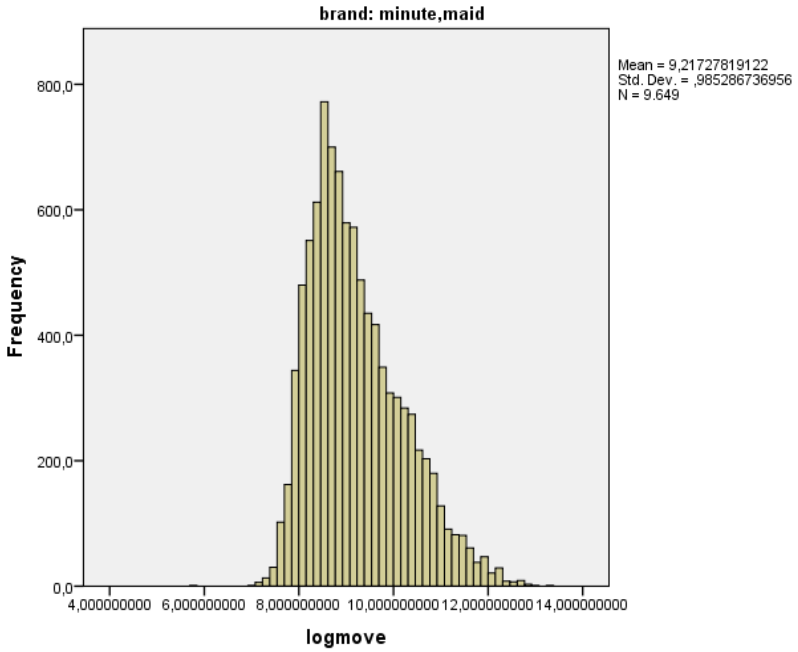
Στα Διαγράμματα 1, 2 και 3 παρουσιάζονται τα ιστογράμματα της κατανομής των πωλήσεων ανά brand και στο Διάγραμμα 4 τα θηκογράμματα (boxplots) της κατανομής των πωλήσεων ανά brand.

Οι κατανομές δεν φαίνεται να διαφέρουν ως προς την θέση (η διάμεσος και για τις τρεις κατανομές είναι περίπου ίδια) ενώ υπάρχουν πιθανές έκτροπες και έκτροπες παρατηρήσεις (πάνω και κάτω) και για τα τρία brands.

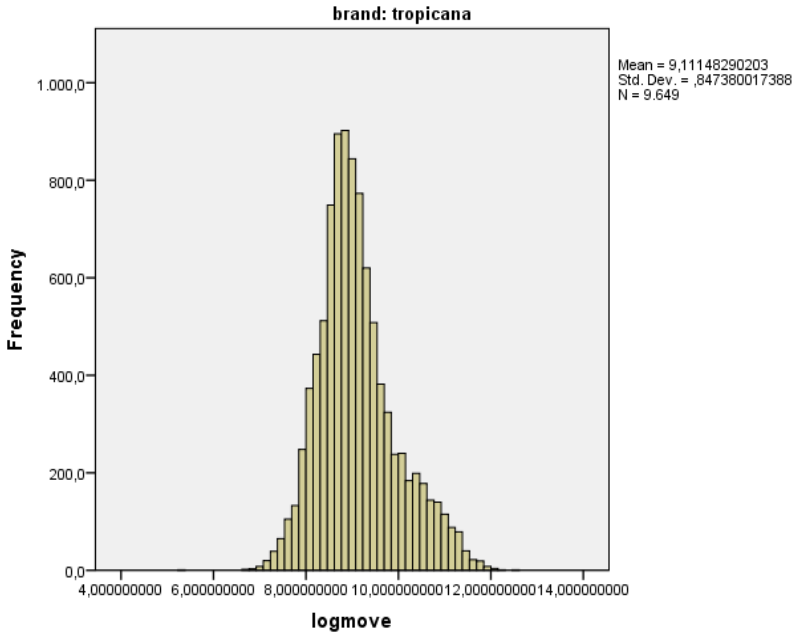
Για την Dominicks η κατανομή των πωλήσεων φαίνεται να προσεγγίζει σε ικανοποιητικό βαθμό την κανονική με μεταβλητότητα μεγαλύτερη από τις υπόλοιπες ενώ οι κατανομές για την MinuteMaid και την Tropicana παρουσιάζουν δεξιά ασυμμετρία.



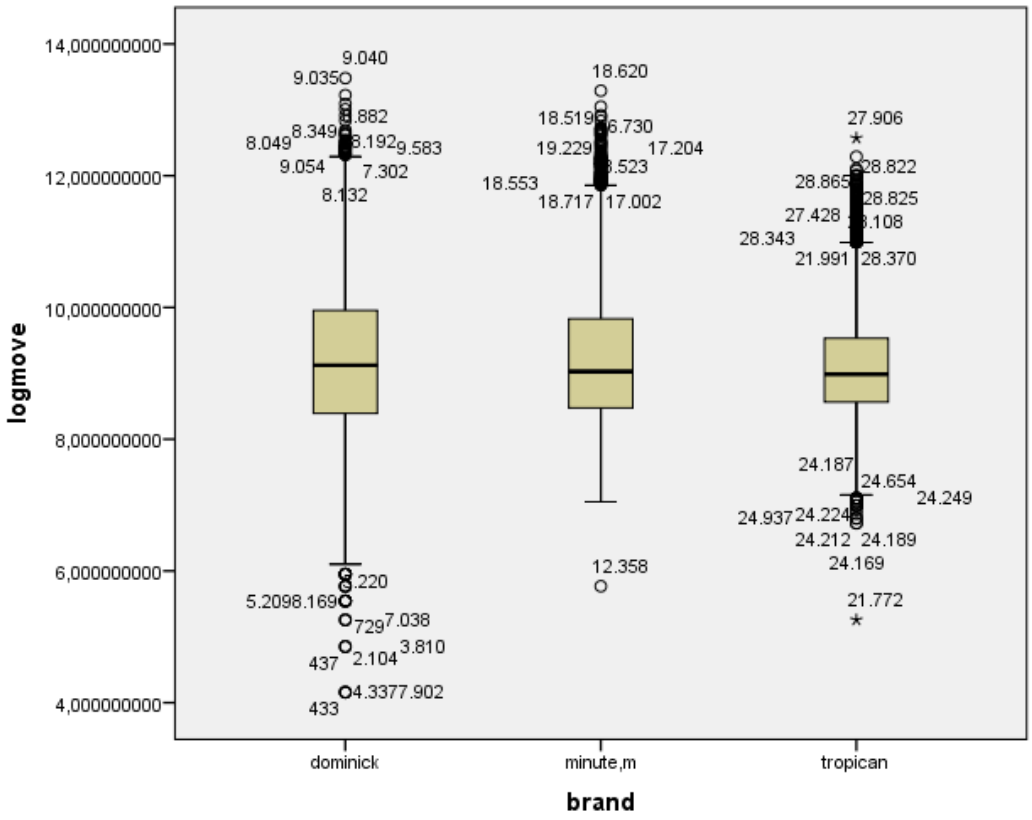
Διάγραμμα 1:Κατανομή των πωλήσεων για το brand Dominicks



Διάγραμμα 2: Κατανομή των πωλήσεων για το brand MinuteMaid



Διάγραμμα 3: Κατανομή των πωλήσεων για το brand Tropicana



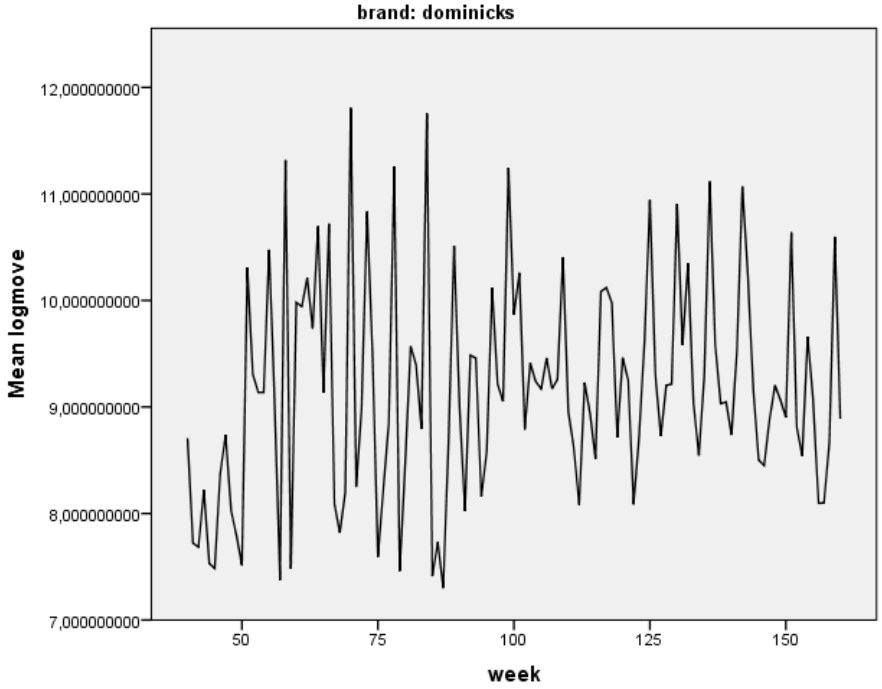
Διάγραμμα 4: Θηκογράμματα (boxplots) της κατανομής των πωλήσεων ανά brand

4.2 ΕΠΙΔΡΑΣΗ ΤΙΜΗΣ ΣΤΙΣ ΠΩΛΗΣΕΙΣ

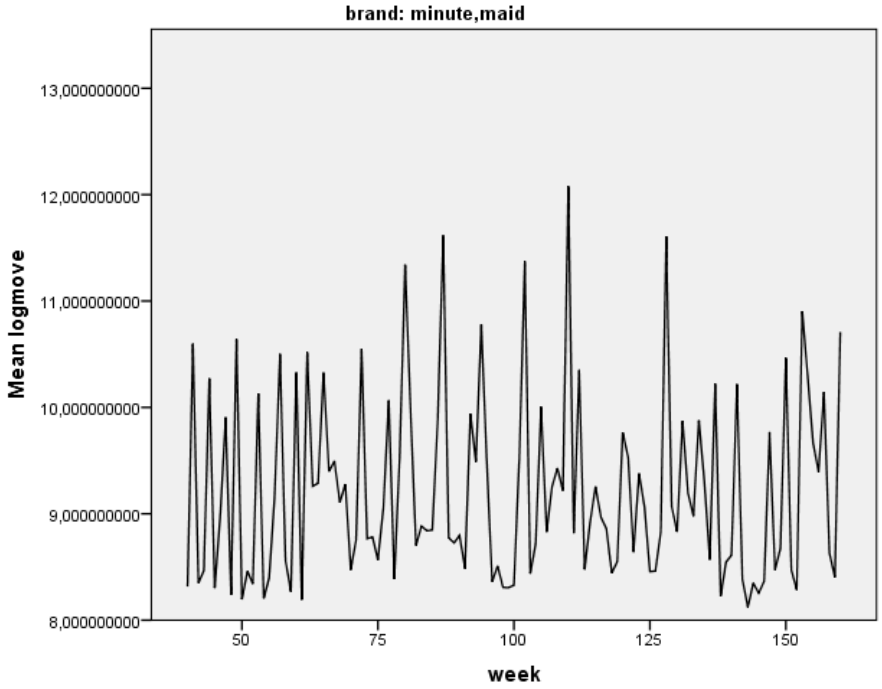
Στη συνέχεια παρουσιάζονται τα διαγράμματα πωλήσεων - χρόνου και τιμής – χρόνου ώστε να συγκριθεί η τάση τους και να μελετηθεί η συσχέτισή τους.

Πωλήσεις - Χρόνος

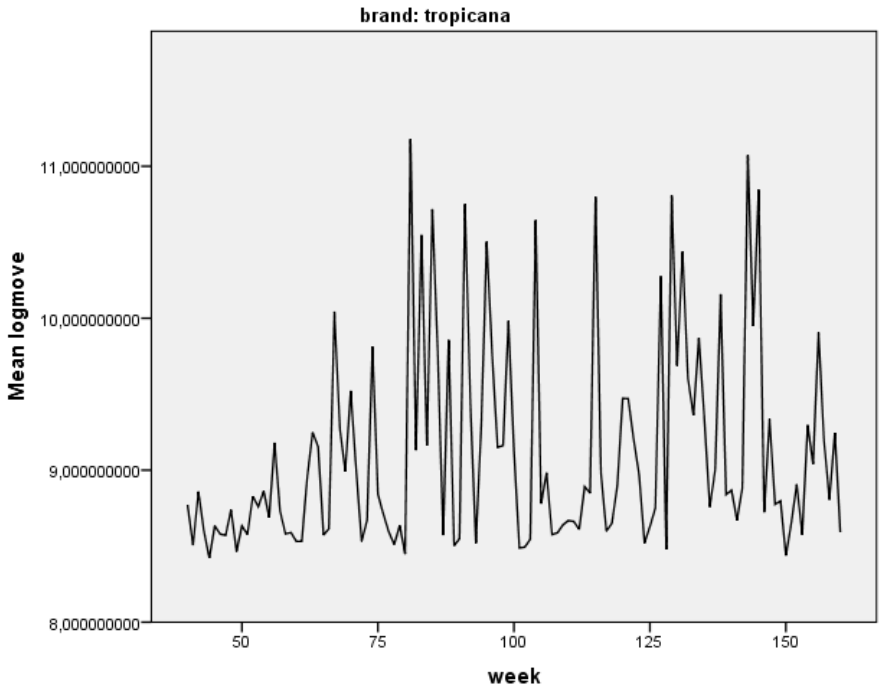
Στα Διαγράμματα 5, 6 και 7 παρουσιάζονται οι πωλήσεις ανά εβδομάδα για τα brands Dominicks, MinuteMaid και Tropicana, αντίστοιχα. Παρατηρούμε ότι για την Dominicks υπάρχει μια σχετικά σταθερή τάση στις πωλήσεις, ενώ παρατηρείται μειωμένη μεταβλητότητα μετά την 100η εβδομάδα. Αναφορικά με την MinuteMaid, επίσης παρατηρείται μικρή αύξηση κατά το «μεσαίο» διάστημα με αντίστοιχη αύξηση της μεταβλητότητας. Τέλος, για την Tropicana παρατηρούμε μια αρχικά αυξανόμενη τάση στις πωλήσεις, η οποία όμως στη συνέχεια μειώνεται, με μια αντίστοιχη μεταβολή και της μεταβλητότητας.



Διάγραμμα 5: Διάγραμμα χρονολογικής σειράς των πωλήσεων (logmove) ανά εβδομάδα για το εμπορικό σήμα Dominicks



Διάγραμμα 6: Διάγραμμα χρονολογικής σειράς των πωλήσεων (logmove) ανά εβδομάδα για το brand MinuteMaid



Διάγραμμα 7: Διάγραμμα χρονολογικής σειράς των πωλήσεων (logmove) ανά εβδομάδα για το brand Tropicana

Τιμές - Χρόνος

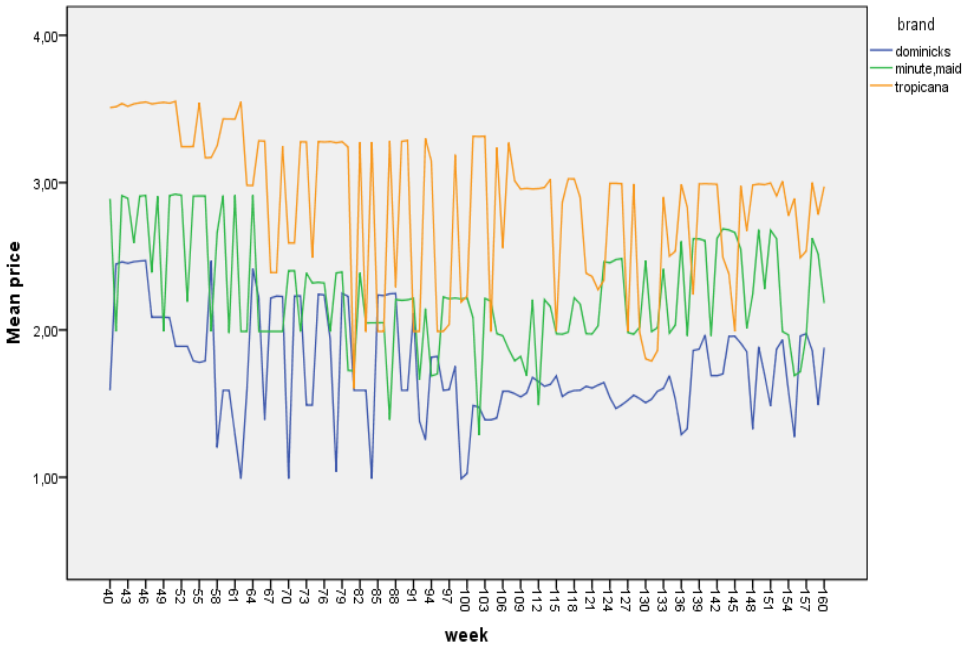
Συνολικά, η τιμή των χυμών διαφέρει στατιστικά σημαντικά μεταξύ των τριών διαφορετικών brand (Πίνακας 1) και μάλιστα διαχρονικά (Διάγραμμα 8).

ANOVA

price

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	6236,305	2	3118,152	15249,710	,000
Within Groups	5918,263	28944	,204		
Total	12154,568	28946			

Πίνακας 1: one-way ANOVA για τις τιμές των τριών brands



Διάγραμμα 8: Διάγραμμα χρονολογικής σειράς της τιμής σε δολάρια (price) ανά εβδομάδα και ανά brand

Συγκρίνοντας το Διάγραμμα 8 με τα Διαγράμματα 5, 6 και 7 παρατηρούμε ότι η τάση είναι αντίστροφη και η μεταβλητότητα παρόμοια. Δηλαδή:

- Dominicks: πτώση των τιμών μέχρι την εβδομάδα 55, σταθερή τάση των τιμών στη συνέχεια και μειωμένη μεταβλητότητα από την εβδομάδα 100 και μετά. Αντίστοιχα παρατηρούμε και στο Διάγραμμα 5 ότι μέχρι την εβδομάδα 55 έχουμε αύξηση των πωλήσεων που ακολουθείται από σταθερότητα και μειωμένη μεταβλητότητα.
- MinuteMaid: Μικρή πτώση των τιμών κατά το ενδιάμεσο διάστημα (εβδομάδες 70 - 110) με αντίστοιχα μικρή αύξηση των πωλήσεων κατά το ίδιο περίπου διάστημα στο Διάγραμμα 6.
- Tropicana: Μείωση των τιμών, σταθερή τάση με αυξημένη μεταβλητότητα στη συνέχεια και μικρή αύξηση κατά το τελευταίο διάστημα. Αντίστοιχα στο Διάγραμμα 7 παρατηρούμε αύξηση των πωλήσεων το αντίστοιχο διάστημα, σταθερή τάση με αυξημένη μεταβλητότητα στη συνέχεια και μικρή μείωση των πωλήσεων κατά το τελευταίο διάστημα.

Τιμές – Πωλήσεις

Στα Διαγράμματα 9, 10 και 11 παρουσιάζεται η πορεία των πωλήσεων σε σχέση με την τιμή, για τα τρία brands χωριστά. Για το Dominicks παρατηρούμε ότι οι πωλήσεις φαίνεται να μειώνονται σταθερά όσο αυξάνει η τιμή. Για το MinutesMaid, οι πωλήσεις φαίνεται να μειώνονται όσο αυξάνει η τιμή ως περίπου την τιμή $2.20 και στη συνέχεια να σταθεροποιούνται. Τέλος, για το Tropicana, οι πωλήσεις φαίνεται να μειώνονται όσο αυξάνει η τιμή ως την τιμή $3.20 και στη συνέχεια να σταθεροποιούνται. Έτσι λοιπόν, επιβεβαιώνεται η αρνητική συσχέτιση της τιμής με τις πωλήσεις και για τα τρία brands, αν και η επιρροή αυτή της τιμής δε φαίνεται να είναι ακριβώς η ίδια μεταξύ τους. Να σημειωθεί ότι και για τα τρία brands το p-value είναι 0.000, δηλαδή μικρότερο από κάθε λογικό επίπεδο σημαντικότητας, συνεπώς **υπάρχει στατιστικά σημαντική συσχέτιση**.

Model Summary and Parameter Estimates[a]

Dependent Variable: logmove

Equation	Model Summary					Parameter Estimates	
	R Square	F	df1	df2	Sig.	Constant	b1
Linear	,388	6104,773	1	9647	,000	12,516	-1,925

The independent variable is price.[a]

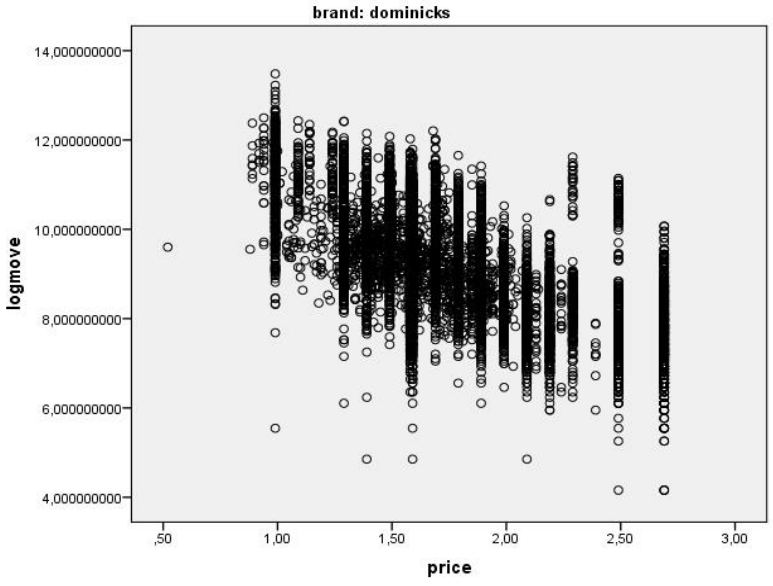
a. brand = dominicks

Correlations[a]

		logmove	price
logmove	Pearson Correlation	1	-,623**
	Sig. (2-tailed)		,000
	N	9649	9649
price	Pearson Correlation	-,623**	1
	Sig. (2-tailed)	,000	
	N	9649	9649

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

a. brand = dominicks



Διάγραμμα 9: Διάγραμμα διασποράς (scatterplot) ανάμεσα στις μεταβλητές πωλήσεις (logmove) και τιμή (price) για το brand Dominicks

Model Summary and Parameter Estimates[a]

Dependent Variable: logmove

Equation	Model Summary					Parameter Estimates	
	R Square	F	df1	df2	Sig.	Constant	b1
Linear	,352	5247,291	1	9647	,000	12,457	-1,446

The independent variable is price.[a]

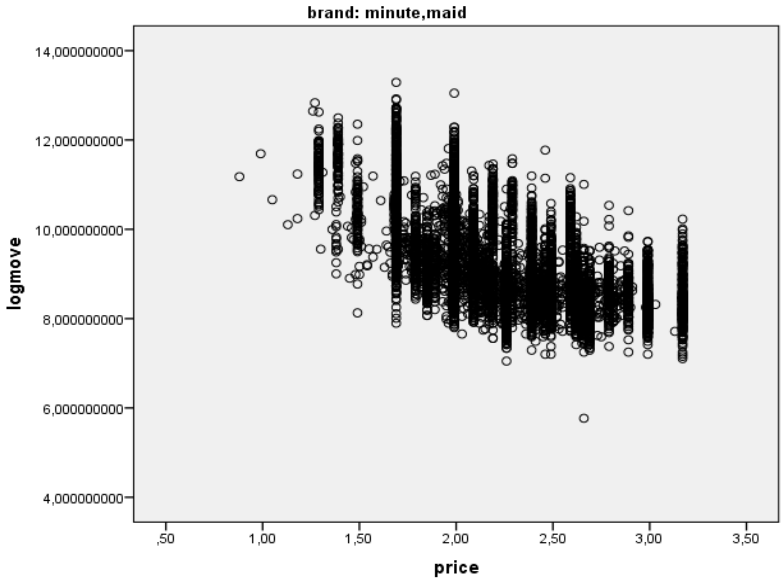
a. brand = minute,maid

Correlations[a]

		logmove	price
logmove	Pearson Correlation	1	-,594**
	Sig. (2-tailed)		,000
	N	9649	9649
price	Pearson Correlation	-,594**	1
	Sig. (2-tailed)	,000	
	N	9649	9649

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

a. brand = minute,maid



Διάγραμμα 10: Διάγραμμα διασποράς (scatterplot) ανάμεσα στις μεταβλητές πωλήσεις (logmove) και τιμή (price) για το brand MinuteMaid

Model Summary and Parameter Estimates[a]

Dependent Variable: logmove

Equation	Model Summary					Parameter Estimates	
	R Square	F	df1	df2	Sig.	Constant	b1
Linear	,400	6431,370	1	9647	,000	11,916	-,977

The independent variable is price.[a]

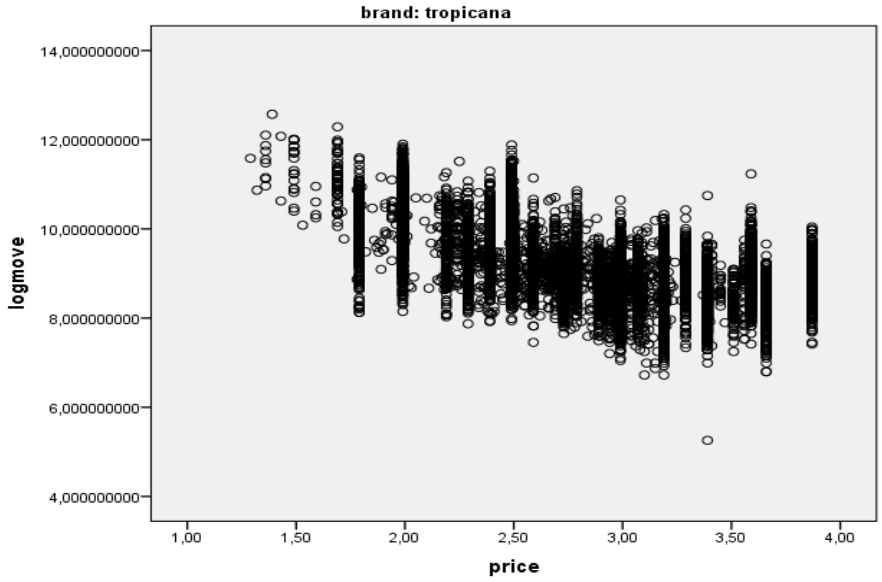
a. brand = tropicana

Correlations[a]

		logmove	price
logmove	Pearson Correlation	1	-,632**
	Sig. (2-tailed)		,000
	N	9649	9649
price	Pearson Correlation	-,632**	1
	Sig. (2-tailed)	,000	
	N	9649	9649

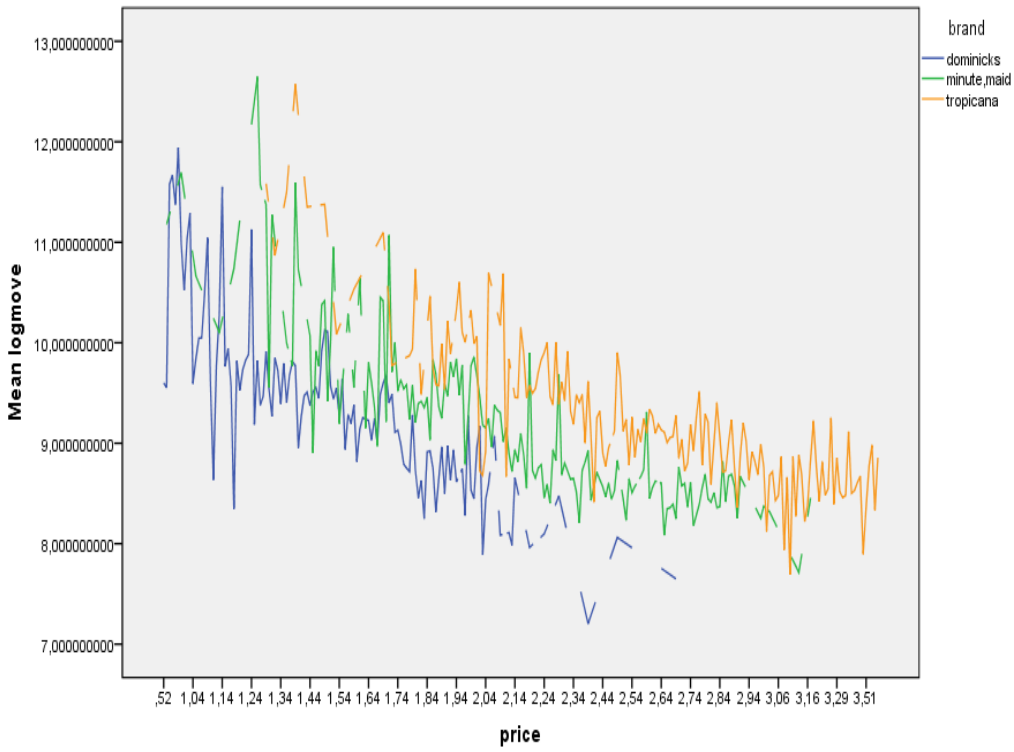
**. Correlation is significant at the 0.01 level (2-tailed).

a. brand = tropicana



Διάγραμμα 11: Διάγραμμα διασποράς (scatterplot) ανάμεσα στις μεταβλητές πωλήσεις (logmove) και τιμή (price) για το brand Tropicana

Συγκεντρωτικά στο Διάγραμμα 12 φαίνεται ότι, ανεξαρτήτως τιμής, οι πωλήσεις των χυμών «Tropicana» είναι οι υψηλότερες ενώ ακολουθούν οι χυμοί «MinuteMaid» και τέλος οι «Dominicks». Επίσης, το brand Tropicana έχει ελάχιστη τιμή μεγαλύτερη σε σχέση με τα άλλα δύο brands και ταυτόχρονα ανέρχεται σε μέγιστη τιμή πώλησης μεγαλύτερη από τα τρία brands, εφόσον το διάγραμμα τιμής-πωλήσεων ξεκινά δεξιότερα. Όσο η τιμή αυξάνει παρατηρείται πτώση στις πωλήσεις και για τα τρία brands, ενώ η πτώση αυτή, στο διάστημα όπου μπορεί να πραγματοποιηθεί σύγκριση, φαίνεται να γίνεται με τον ίδιο σχετικά ρυθμό για τα τρία brands, καθώς η τιμή τους αυξάνει.



Διάγραμμα 12: Διάγραμμα της σχέσης πωλήσεων (logmove) τιμής σε δολάρια (price) ανά brand

4.3 ΕΠΙΔΡΑΣΗ ΔΙΑΦΗΜΙΣΗΣ ΣΤΙΣ ΠΩΛΗΣΕΙΣ

Στον Πίνακα 2 παρουσιάζονται τα αποτελέσματα της σύγκρισης των πωλήσεων ανάλογα με το αν τα προϊόντα διαφημίζονταν ή όχι, συνολικά (ανεξαρτήτως brand). Κατ' αρχάς, υπάρχει στατιστικά σημαντική διαφορά στις πωλήσεις, ανάλογα με το αν υπήρχε ή όχι διαφήμιση ($p<0.001$). Βρέθηκε ότι:

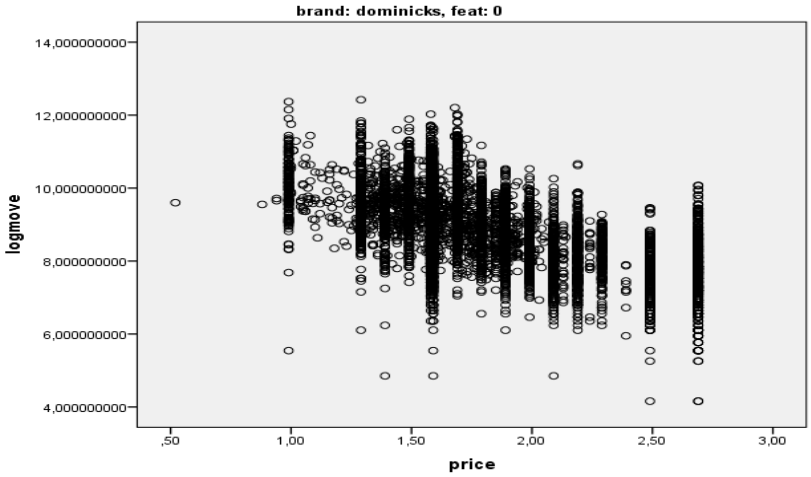
- οι πωλήσεις ήταν υψηλότερες όταν υπήρχε διαφήμιση. Από 7.000τεμ./εβδομάδα αυξάνονταν περίπου στις 25.500τεμ./εβδομάδα
- η τιμή πώλησης ήταν κατά μέσο όρο χαμηλότερη όταν υπήρχε διαφήμιση, επομένως χρήζει περαιτέρω διερεύνησης σε τι βαθμό οι πωλήσεις επηρεάζονται από την ύπαρξη διαφήμισης και σε τι βαθμό από τη μεταβολή των τιμών.

Total	ΧΩΡΙΣ διαφήμιση (feat=0)		ΜΕ διαφήμιση (feat=1)		
	Μέσος όρος	Τυπική απόκλιση	Μέσος όρος	Τυπική απόκλιση	p-value
logmove	8.86	0.81	10.15	1.01	<0.001
price	2.39	0.66	1.94	0.48	<0.001

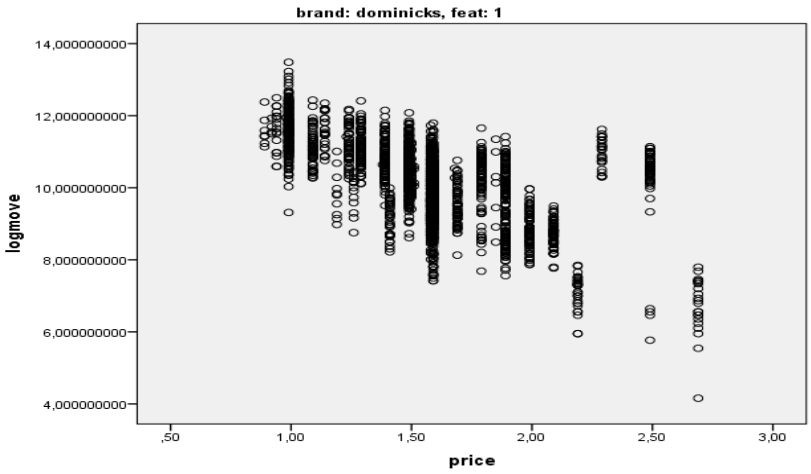
Πίνακας 2: Σύγκριση (t-test) των πωλήσεων (logmove) και της τιμής σε δολάρια (price) όταν υπάρχει και όταν δεν υπάρχει διαφήμιση

Πιο αναλυτικά, στα Διαγράμματα 13-18 παρουσιάζεται η πορεία των πωλήσεων σε σχέση με την τιμή, ανά brand, ενώ υπάρχει και ενώ απουσιάζει η διαφήμιση. Οι πωλήσεις με διαφήμιση είναι γενικά υψηλότερες. Η τάση είναι παρόμοια για τα τρία brands. Βέβαια, τα σημεία με διαφήμιση είναι λιγότερα από τα σημεία χωρίς διαφήμιση.

Στους Πίνακες 3-5 παρουσιάζονται τα αποτελέσματα της σύγκρισης των πωλήσεων με ή χωρίς διαφήμιση ανά brand. Από τα τρία brand, η διαφήμιση φαίνεται να επηρεάζει περισσότερο τις πωλήσεις του MinuteMaid και λιγότερο του Tropicana.



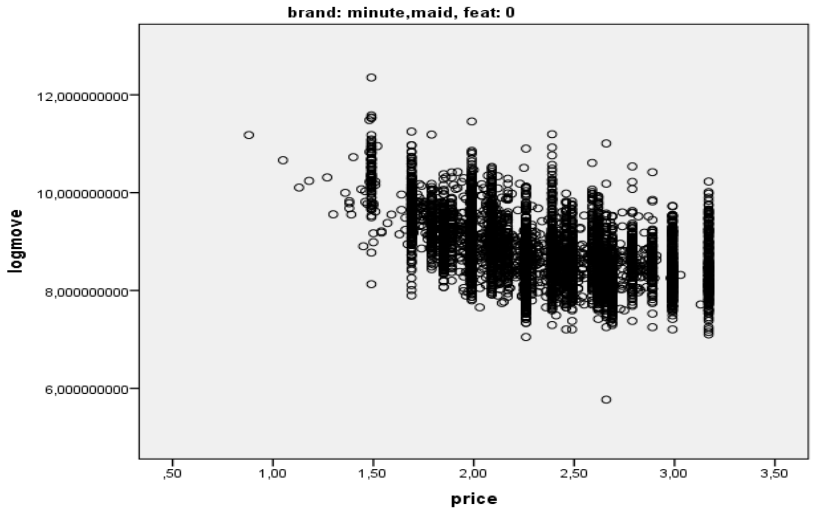
Διάγραμμα 13: Διάγραμμα διασποράς (scatterplot) ανάμεσα στις μεταβλητές πωλήσεις (logmove) και τιμή (price) για το brand Dominicks, ΧΩΡΙΣ διαφήμιση



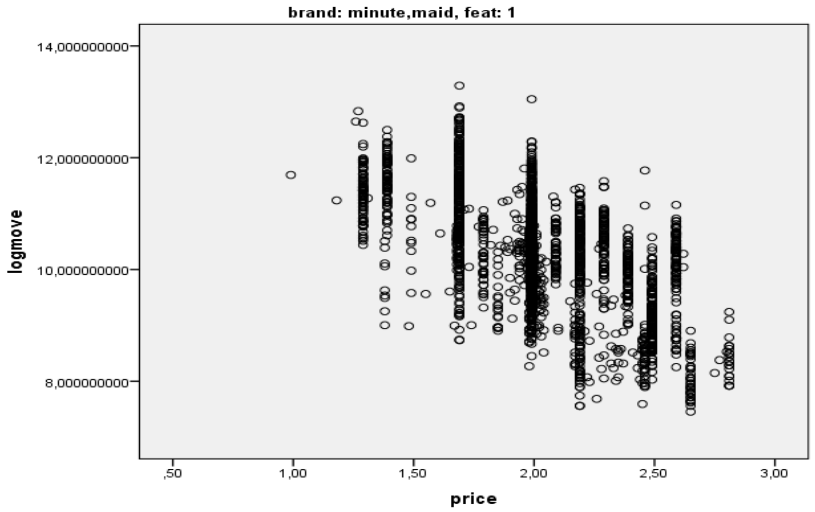
Διάγραμμα 14: Διάγραμμα διασποράς (scatterplot) ανάμεσα στις μεταβλητές πωλήσεις (logmove) και τιμή (price) για το brand Dominicks, ΜΕ διαφήμιση

Dominicks	ΧΩΡΙΣ διαφήμιση (feat=0)		ΜΕ διαφήμιση (feat=1)		
	Μέσος όρος	Τυπική απόκλιση	Μέσος όρος	Τυπική απόκλιση	p-value
logmove	8.84	1.03	10.13	1.13	<0.001
price	1.79	0.38	1.56	0.34	<0.001

Πίνακας 3: Σύγκριση (t-test) των πωλήσεων (logmove) και της τιμής σε δολάρια (price) όταν υπάρχει και όταν δεν υπάρχει διαφήμιση για το brand Dominicks



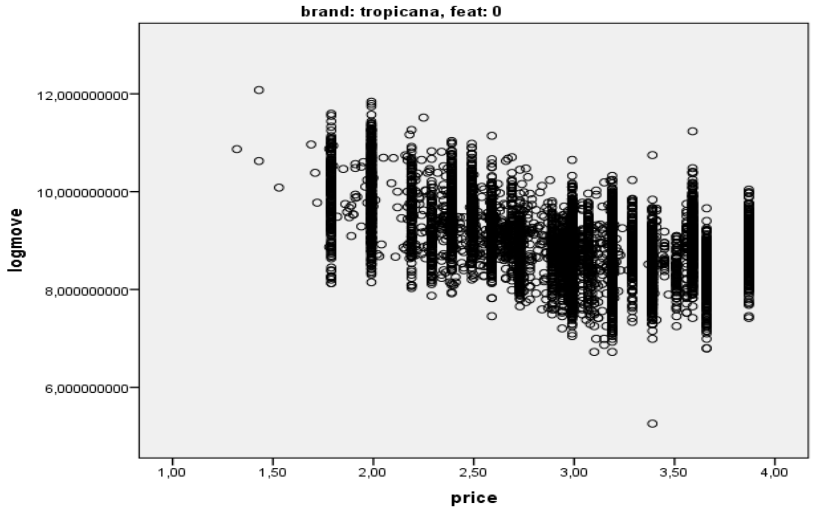
Διάγραμμα 15:Διάγραμμα διασποράς (scatterplot) ανάμεσα στις μεταβλητές πωλήσεις (logmove) και τιμή (price) για το brand MinuteMaid, ΧΩΡΙΣ διαφήμιση



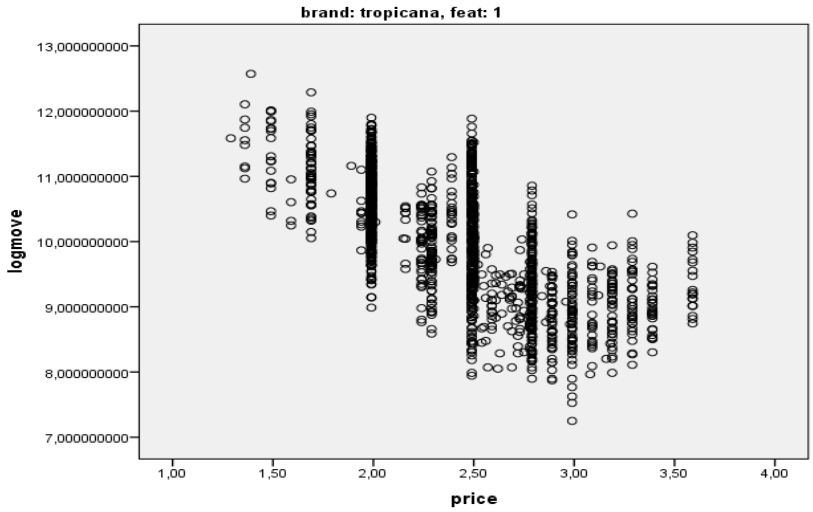
Διάγραμμα 16:Διάγραμμα διασποράς (scatterplot) ανάμεσα στις μεταβλητές πωλήσεις (logmove) και τιμή (price) για το brand MinuteMaid, ΜΕ διαφήμιση

MinuteMaid	ΧΩΡΙΣ διαφήμιση (feat=0)		ΜΕ διαφήμιση (feat=1)		
	Μέσος όρος	Τυπική απόκλιση	Μέσος όρος	Τυπική απόκλιση	p-value
logmove	8.80	0.64	10.24	0.93	<0.001
price	2.33	0.41	2.02	0.30	<0.001

Πίνακας 4: Σύγκριση (t-test) των πωλήσεων (logmove) και της τιμής σε δολάρια (price) όταν υπάρχει και όταν δεν υπάρχει διαφήμιση για το brand MinuteMaid



Διάγραμμα 17: Διάγραμμα διασποράς (scatterplot) ανάμεσα στις μεταβλητές πωλήσεις (logmove) και τιμή (price) για το brand Tropicana, ΧΩΡΙΣ διαφήμιση



Διάγραμμα 18: Διάγραμμα διασποράς (scatterplot) ανάμεσα στις μεταβλητές πωλήσεις (logmove) και τιμή (price) για το brand Tropicana, ΜΕ διαφήμιση

Tropicana	ΧΩΡΙΣ διαφήμιση (feat=0)		ΜΕ διαφήμιση (feat=1)		
	Μέσος όρος	Τυπική απόκλιση	Μέσος όρος	Τυπική απόκλιση	p-value
logmove	8.93	0.70	10.01	0.93	<0.001
price	2.97	0.51	2.39	0.46	<0.001

Πίνακας 5: Σύγκριση (t-test) των πωλήσεων (logmove) και της τιμής σε δολάρια (price) όταν υπάρχει και όταν δεν υπάρχει διαφήμιση για το brand Tropicana

4.4 ΣΥΣΧΕΤΙΣΕΙΣ ΔΗΜΟΓΡΑΦΙΚΩΝ ΠΑΡΑΓΟΝΤΩΝ

Στον Πίνακα 6 παρουσιάζονται όλες οι ανά δύο συσχετίσεις, μεταξύ των βασικών δημογραφικών χαρακτηριστικών και της τιμής των χυμών (ανεξαρτήτως brand). Βρέθηκε ότι το εισόδημα (ο λογάριθμος του μέσου εισοδήματος) και το ποσοστό των έγχρωμων/ισπανόφωνων παρουσιάζουν έντονη αρνητική στατιστικά σημαντική συσχέτιση ($r=0.72$, $p<0.001$). Δηλαδή, όπου το ποσοστό των έγχρωμων/ισπανόφωνων είναι πιο υψηλό, το εισόδημα του πληθυσμού εκεί παρουσιάζεται μειωμένο. Το ποσοστό των πτυχιούχων και το ποσοστό των νοικοκυριών με περιουσιακά στοιχεία άνω των $150.000 παρουσιάζουν έντονη θετική στατιστικά σημαντική συσχέτιση ($r=0.887$, $p<0.001$). Δηλαδή, υψηλότερο ποσοστό πτυχιούχων αντιστοιχεί σε υψηλότερο ποσοστό νοικοκυριών με περιουσιακά στοιχεία άνω των $150.000.

Correlations	age60	educ	ethnic	income	hhlarge	workwom
AGE60	1					
EDUC	**-0,310** (p<0.001)	1				
ETHNIC	**-0,094** (p<0.001)	**-0,340** (p<0.001)	1			
INCOME	**-0,153** (p<0.001)	**0,663** (p<0.001)	**-0,720** **(p<0.001)**	1		
HHLARGE	**-0,323** (p<0.001)	**-0,390** (p<0.001)	**0,253** (p<0.001)	**-0,082** (p<0.001)	1	
WORKWOM	**-0,629** (p<0.001)	**0,560** (p<0.001)	**-0,288** (p<0.001)	**0,399** (p<0.001)	**-0,283** (p<0.001)	1
HVAL150	**-0,009** (p<0.001)	**0,887** **(p<0.001)**	**-0,421** (p<0.001)	**0,639** (p<0.001)	**-0,480** (p<0.001)	**0,452** (p<0.001)

Πίνακας 6: Συσχετίσεις δημογραφικών παραγόντων

4.4.1 Σχέση πωλήσεων - πτυχιούχων

Στα Διαγράμματα 19-21 παρουσιάζονται τα ευρήματα αναφορικά με την επίδραση του ποσοστού των πτυχιούχων στις πωλήσεις χυμών, ανά brand, και στο Διάγραμμα 22 απεικονίζονται συγκριτικά. Συγκεκριμένα:

- Διάγραμμα 19: το Dominicks παρουσιάζει μείωση των πωλήσεων όσο αυξάνεται το ποσοστό των πτυχιούχων

Correlations[a]

		logmove	EDUC
logmove	Pearson Correlation	1	-,170**
	Sig. (2-tailed)		,000
	N	9649	9649
EDUC	Pearson Correlation	-,170**	1
	Sig. (2-tailed)	,000	
	N	9649	9649

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

a. brand = dominicks

Model Summary and Parameter Estimates[a]

Dependent Variable: logmove

Equation	Model Summary					Parameter Estimates	
	R Square	F	df1	df2	Sig.	Constant	b1
Linear	,029	286,334	1	9647	,000	9,590	-1,842

The independent variable is EDUC.[a]

a. brand = dominicks

logmove =9,590 -1,842educ

- Διάγραμμα 20: οι πωλήσεις του MinuteMaid δε φαίνεται να επηρεάζονται από το ποσοστό των πτυχιούχων

Correlations[a]

		logmove	EDUC
logmove	Pearson Correlation	1	,031**
	Sig. (2-tailed)		,002
	N	9649	9649
EDUC	Pearson Correlation	,031**	1
	Sig. (2-tailed)	,002	
	N	9649	9649

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

a. brand = minute,maid

Model Summary and Parameter Estimates[a]

Dependent Variable: logmove

Equation	Model Summary					Parameter Estimates	
	R Square	F	df1	df2	Sig.	Constant	b1
Linear	,001	9,456	1	9647	,002	9,154	,280

The independent variable is EDUC.[a]

a. brand = minute,maid

logmove = 9,154 + 0,280educ

- Διάγραμμα 21: το Tropicana παρουσιάζει αύξηση των πωλήσεων όσο το ποσοστό των πτυχιούχων αυξάνεται

Correlations[a]

		logmove	EDUC
logmove	Pearson Correlation	1	,219**
	Sig. (2-tailed)		,000
	N	9649	9649
EDUC	Pearson Correlation	,219**	1
	Sig. (2-tailed)	,000	
	N	9649	9649

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

a. brand = tropicana

Model Summary and Parameter Estimates[a]

Dependent Variable: logmove

Equation	Model Summary					Parameter Estimates	
	R Square	F	df1	df2	Sig.	Constant	b1
Linear	,048	486,254	1	9647	,000	8,731	1,688

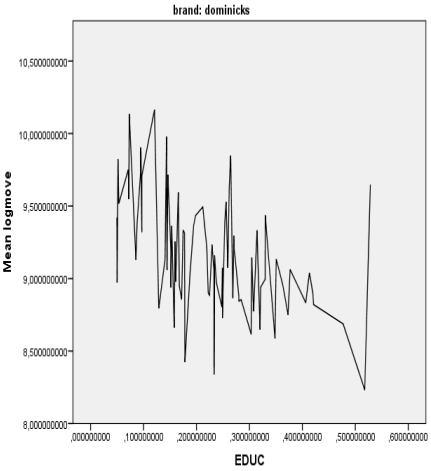
The independent variable is EDUC.[a]

a. brand = tropicana

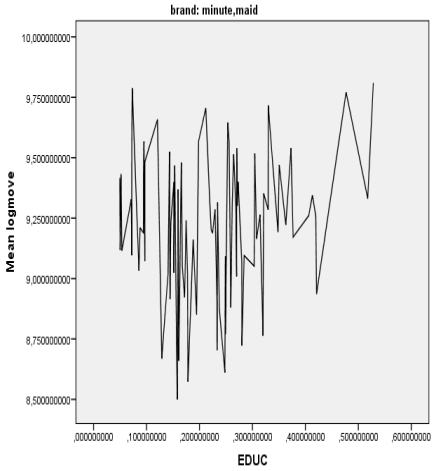
logmove = 8,731 + 1,688educ

Και στις τρείς περιπτώσεις το p-value μας δείχνει ότι η όποια συσχέτιση είναι στατιστικά σημαντική.

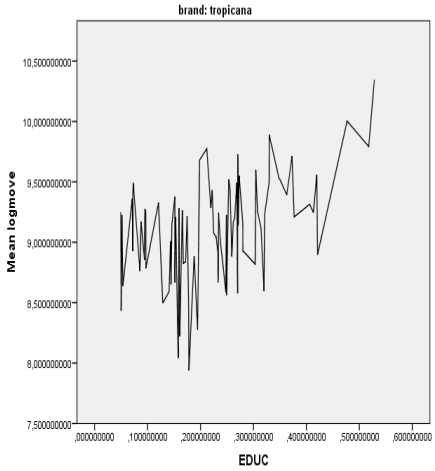
- Διάγραμμα 22: Συνολικά βλέπουμε ότι στα χαμηλότερα ποσοστά πτυχιούχων η Dominicks έρχεται πρώτη και η Tropicana τρίτη σε πωλήσεις, ενώ στα υψηλότερα η εικόνα αντιστρέφεται με την Tropicana πρώτη και την Dominicks τρίτη στις πωλήσεις



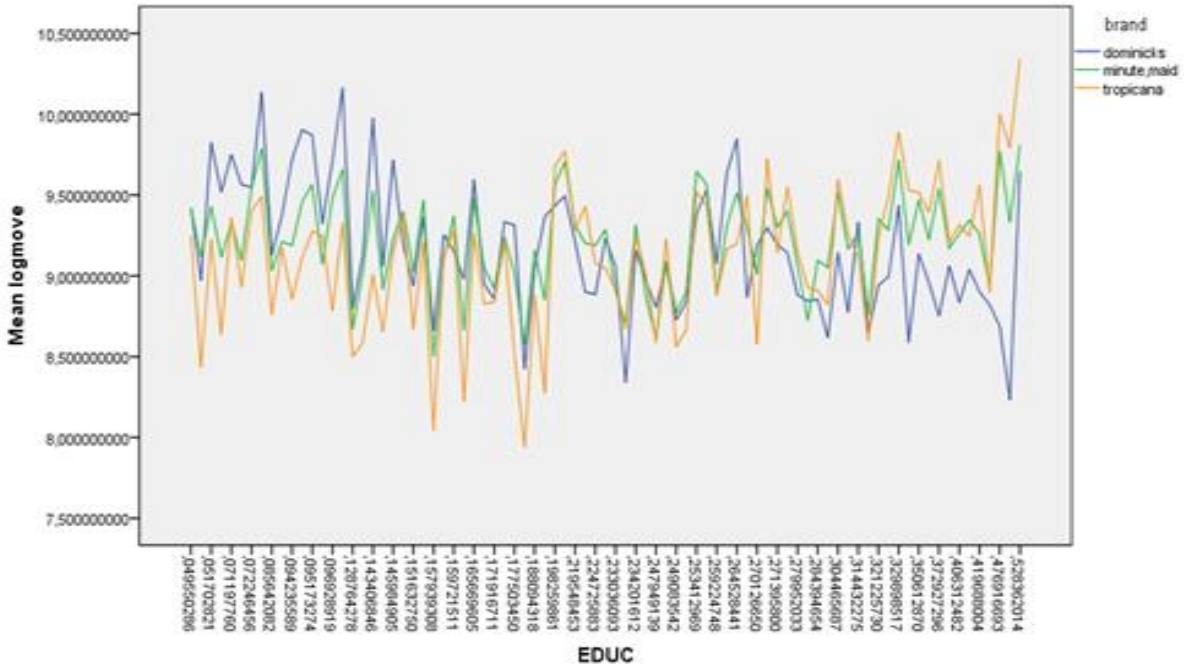
Διάγραμμα 19: Dominicks



Διάγραμμα 20: MinuteMaid



Διάγραμμα 21: Tropicana



Διάγραμμα 22: Διάγραμμα της σχέσης πωλήσεων (logmove) και του ποσοστού των πτυχιούχων (Educ) ανά brand

4.4.2 Σχέση πωλήσεων – φυλετικών ομάδων

Αναφορικά με την σχέση του ποσοστού των έγχρωμων/ισπανόφωνων, τα τρία brands φαίνεται να μη διαφοροποιούνται έντονα παρά μόνο όταν το ποσοστό αυτό υπερβαίνει το 25%, όπου οι πωλήσεις Dominicks παρουσιάζουν ελαφρώς ανοδική τάση (Διάγραμμα 23).

Correlations[a]

		logmove	ETHNIC
logmove	Pearson Correlation	1	,145**
	Sig. (2-tailed)		,000
	N	9649	9649
ETHNIC	Pearson Correlation	,145**	1
	Sig. (2-tailed)	,000	
	N	9649	9649

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

a. brand = dominicks

Model Summary and Parameter Estimates[a]

Dependent Variable: logmove

Equation	Model Summary					Parameter Estimates	
	R Square	F	df1	df2	Sig.	Constant	b1
Linear	,021	207,581	1	9647	,000	9,031	,923

The independent variable is ETHNIC.

a. brand = dominicks

logmove = 9,031 + 0,923ethnic

Correlations[a]

		logmove	ETHNIC
logmove	Pearson Correlation	1	,057**
	Sig. (2-tailed)		,000
	N	9649	9649
ETHNIC	Pearson Correlation	,057**	1
	Sig. (2-tailed)	,000	
	N	9649	9649

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

a. brand = minute,maid

Model Summary and Parameter Estimates[a]

Dependent Variable: logmove

Equation	Model Summary					Parameter Estimates	
	R Square	F	df1	df2	Sig.	Constant	b1
Linear	,003	31,970	1	9647	,000	9,170	,302

The independent variable is ETHNIC.

a. brand = minute,maid

logmove = 9,170 + 0,302ethnic

Correlations[a]

		logmove	ETHNIC
logmove	Pearson Correlation	1	-,055**
	Sig. (2-tailed)		,000
	N	9649	9649
ETHNIC	Pearson Correlation	-,055**	1
	Sig. (2-tailed)	,000	
	N	9649	9649

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

a. brand = tropicana

Model Summary and Parameter Estimates[a]

Dependent Variable: logmove

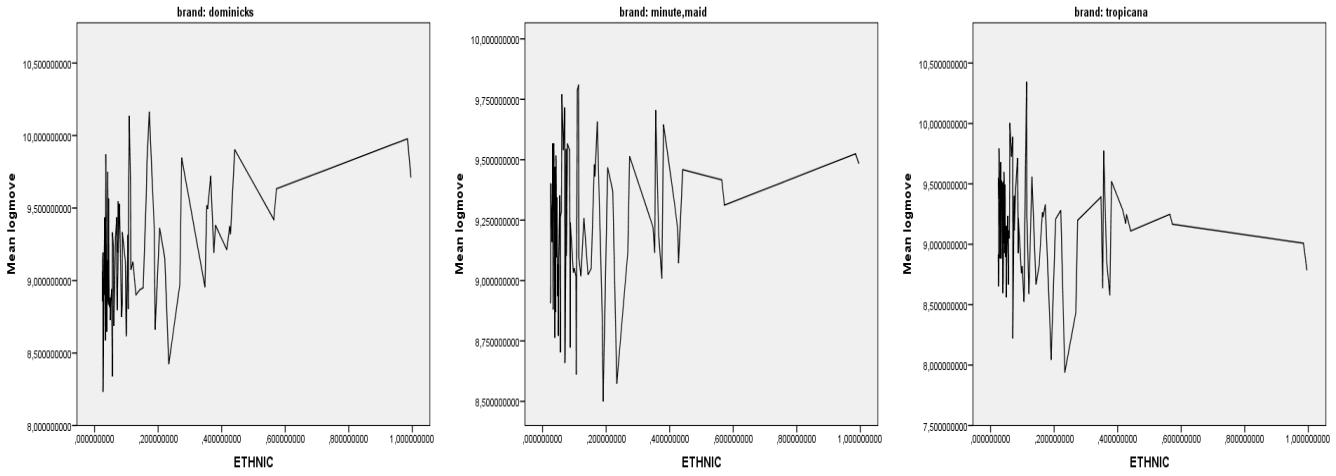
Equation	Model Summary					Parameter Estimates	
	R Square	F	df1	df2	Sig.	Constant	b1
Linear	,003	29,359	1	9647	,000	9,150	-,249

The independent variable is ETHNIC.

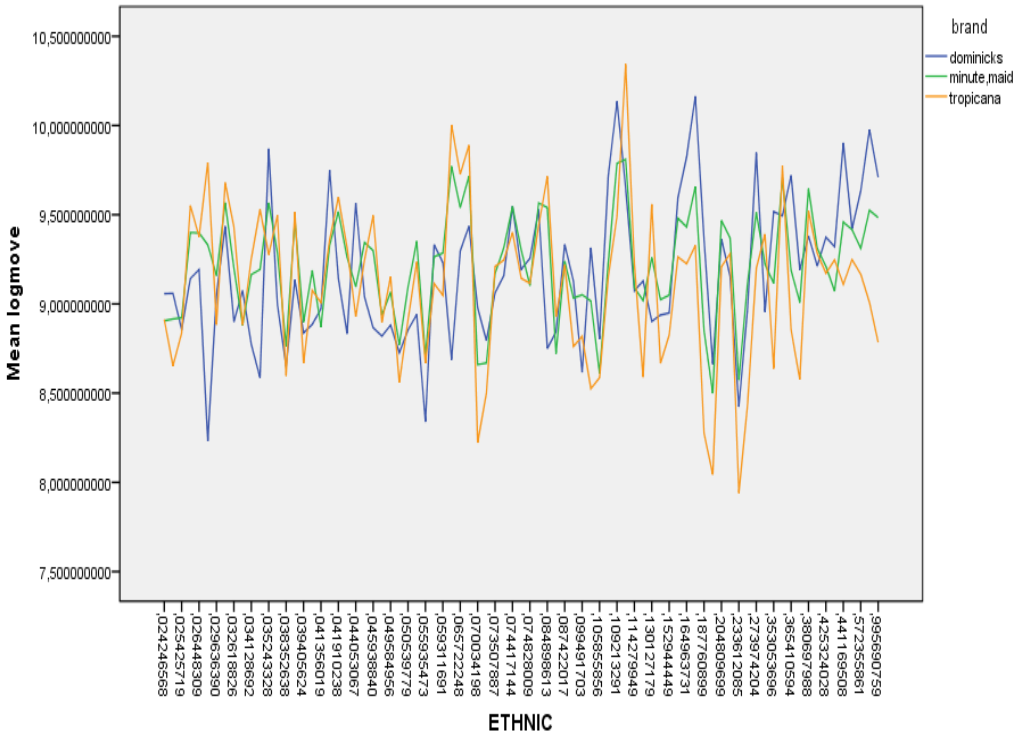
a. brand = tropicana

logmove = 9,150 - 0,249ethnic

Και στις τρείς περιπτώσεις το p-value μας δείχνει ότι οι συσχετίσεις είναι στατιστικά σημαντικές.



Dominicks MinuteMaid Tropicana



Διάγραμμα 23: Διάγραμμα της σχέσης πωλήσεων (logmove) και του ποσοστού έγχρωμων/ισπανόφωνων (ETHNIC) ανά brand

4.4.3 Σχέση πωλήσεων - εισοδήματος

Τέλος, αναφορικά με την επίδραση του συνολικού εισοδήματος, τα τρία brands κινούνται χωρίς μεγάλες διαφοροποιήσεις μεταξύ τους. Η Dominicks παρουσιάζει συνεχώς πτωτική πορεία στις πωλήσεις της όσο αυξάνεται το εισόδημα, ενώ η MinuteMaid και η Tropicana παρουσιάζουν ανοδική πορεία μόνο από την τιμή 10.8 και πάνω.

Correlations[a]

		logmove	INCOME
logmove	Pearson Correlation	1	-,162**
	Sig. (2-tailed)		,000
	N	9649	9649
INCOME	Pearson Correlation	-,162**	1
	Sig. (2-tailed)	,000	
	N	9649	9649

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

a. brand = dominicks

Model Summary and Parameter Estimates[a]

Dependent Variable: logmove

Equation	Model Summary					Parameter Estimates	
	R Square	F	df1	df2	Sig.	Constant	b1
Linear	,026	258,940	1	9647	,000	16,428	-,683

The independent variable is INCOME.[a]

a. brand = dominicks

logmove = 16,428 - 0,683income

Correlations[a]

		logmove	INCOME
logmove	Pearson Correlation	1	-,019
	Sig. (2-tailed)		,067
	N	9649	9649
INCOME	Pearson Correlation	-,019	1
	Sig. (2-tailed)	,067	
	N	9649	9649

a. brand = minute,maid

Model Summary and Parameter Estimates[a]

Dependent Variable: logmove

Equation	Model Summary					Parameter Estimates	
	R Square	F	df1	df2	Sig.	Constant	b1
Linear	,000	3,362	1	9647	,067	9,909	-,065

The independent variable is INCOME.[a]

a. brand = minute,maid

logmove = 9,909 - 0,065income

Correlations[a]

		logmove	INCOME
logmove	Pearson Correlation	1	,095**
	Sig. (2-tailed)		,000
	N	9649	9649
INCOME	Pearson Correlation	,095**	1
	Sig. (2-tailed)	,000	
	N	9649	9649

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

a. brand = tropicana

Model Summary and Parameter Estimates[a]

Dependent Variable: logmove

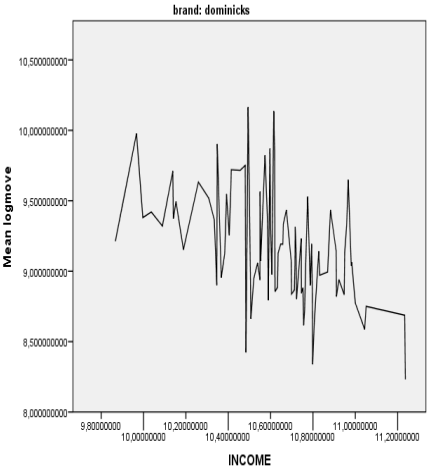
Equation	Model Summary					Parameter Estimates	
	R Square	F	df1	df2	Sig.	Constant	b1
Linear	,009	87,710	1	9647	,000	6,087	,285

The independent variable is INCOME.[a]

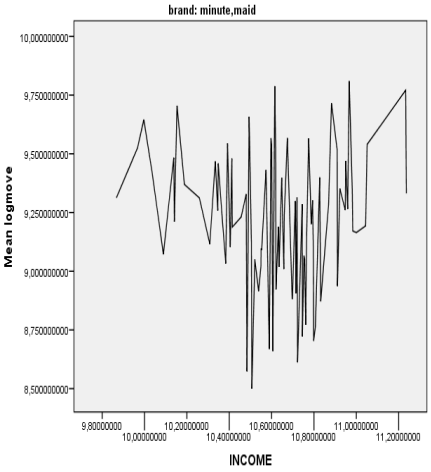
a. brand = tropicana

logmove = 6,087 - 0,285income

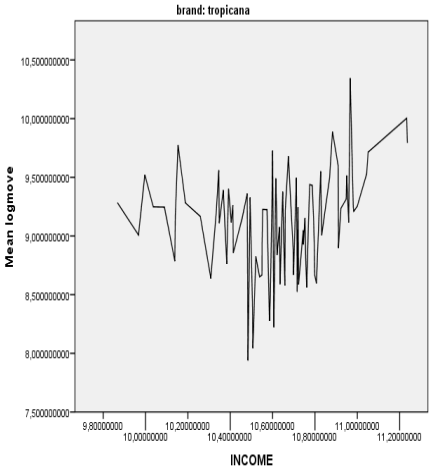
Εδώ βλέπουμε ότι για την περίπτωση του MinuteMaid η συσχέτιση που υπολογίστηκε για 95% σιγουριά δεν είναι στατιστικά σημαντική (για τα άλλα δύο brands είναι).



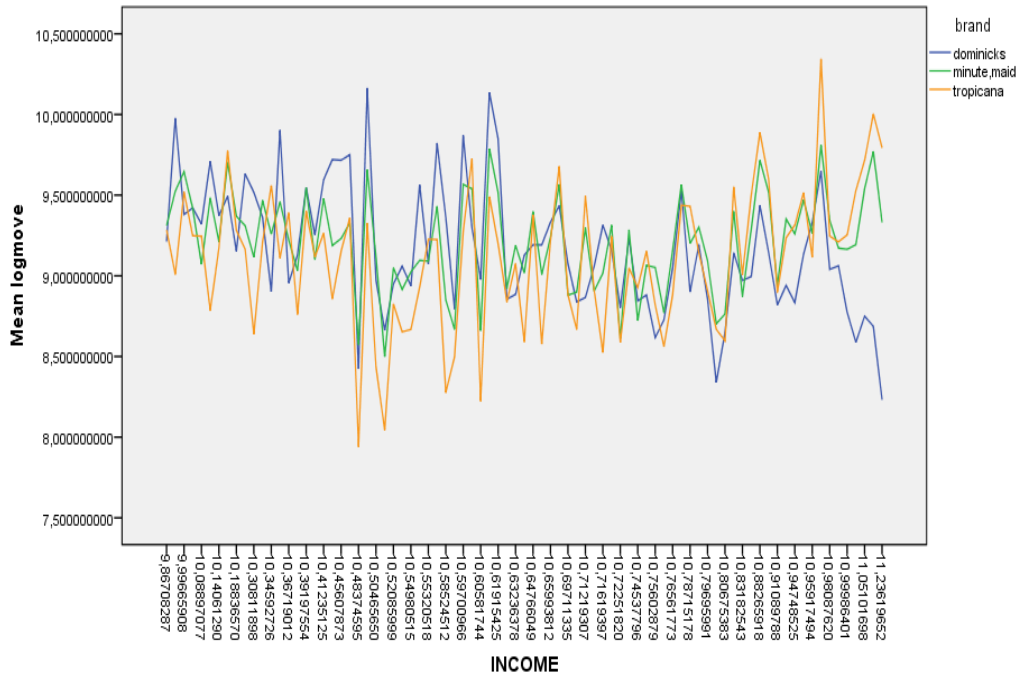
Dominicks



MinuteMaid



Tropicana



Διάγραμμα 24: Διάγραμμα της σχέσης πωλήσεων (logmove) και του λογάριθμου του διάμεσου εισοδήματος (INCOME) ανά εμπορικό σήμα

Ο τρόπος επίδρασης των τριών αυτών χαρακτηριστικών (ποσοστό πτυχιούχων, ποσοστό έγχρωμων/ισπανόφωνων και εισόδημα) στις πωλήσεις, ανάλογα με το brand, θα γίνει σαφέστερος με την ανάλυση παλινδρόμησης της επόμενης ενότητας.

4.5 ΑΝΑΛΥΣΗ ΠΑΛΙΝΔΡΟΜΗΣΗΣ

Λαμβάνοντας υπόψη όλα τα παραπάνω, στη συνέχεια θα διερευνηθεί η επίδραση διάφορων πιθανών παραγόντων στις πωλήσεις χυμών και τέλος, θα γίνει προσπάθεια εύρεσης του βέλτιστου δυνατού μοντέλου που να ερμηνεύει τις διαφοροποιήσεις στις πωλήσεις των χυμών. Συνεπώς, ακολουθεί μια προσπάθεια ανάδειξης των πιθανών παραγόντων που επιδρούν ταυτόχρονα στις πωλήσεις των χυμών των τριών brands που χρησιμοποιήθηκαν στην παρούσα μελέτη.

Αρχικά, στον Πίνακα 7, παρουσιάζονται τα αποτελέσματα της ανάλυσης γραμμικής παλινδρόμησης με εξαρτημένη μεταβλητή τις συνολικές πωλήσεις χυμών (σε λογαριθμική κλίμακα) και ως ανεξάρτητη την τιμή. Βρέθηκε ότι για κάθε μία μονάδα αύξηση στην τιμή πώλησης, αναμένουμε μείωση κατά 0.680 μονάδες στις πωλήσεις (σε λογαριθμική κλίμακα), με $p<0.001$. Βέβαια, η μεταβλητότητα της τιμής πώλησης ερμηνεύει μόνη της το 18.7% της μεταβλητότητας των πωλήσεων χυμών (σε λογαριθμική κλίμακα).

Coefficients

	Unstandardized Coefficients		Standardized Coefficients		
	B	Std. Error	Beta	t	Sig.
price	-,680	,008	-,432	-81,490	,000
(Constant)	10,719	,020		541,738	,000

Model Summary

R	R Square	Adjusted R Square	Std. Error of the Estimate
,432	,187	,187	,919

The independent variable is price.

Πίνακας 7: Ανάλυση γραμμικής παλινδρόμησης με εξαρτημένη μεταβλητή το λογάριθμο των πωλήσεων και ανεξάρτητη την τιμή σε δολάρια (price)

$$\boldsymbol{logmove = 10,719 - 0,680price}$$

Στον Πίνακα 8 παρουσιάζονται όλα τα μονοπαραγοντικά μοντέλα γραμμικής παλινδρόμησης, με εξαρτημένη μεταβλητή τις πωλήσεις σε λογαριθμική κλίμακα και ανεξάρτητες τους

δημογραφικούς παράγοντες. Τα αντίστοιχα γραφήματα παρατίθενται στο Παράρτημα (Διάγραμμα 1.11-1.20). Παρατηρήθηκε ότι οι περισσότεροι δημογραφικοί παράγοντες παρουσιάζουν στατιστικά σημαντική γραμμική σχέση με τις πωλήσεις. Βέβαια, φαίνεται ότι η επίδραση των παραγόντων αυτών στις πωλήσεις είναι ισχνή, ενώ ταυτόχρονα κανένας από αυτούς του παράγοντες δεν επιτυγχάνει να ερμηνεύσει από μόνος του πάνω από το 0.8% της συνολικής μεταβλητότητας του λογάριθμου των πωλήσεων. Για παράδειγμα, βρέθηκε ότι για αύξηση κατά 100% του ποσοστού των ηλικιωμένων άνω των 60 ετών, αναμένεται μέση αύξηση του λογάριθμου των πωλήσεων κατά 1.52 μονάδες, ενώ η μεταβλητή αυτή ερμηνεύει μόλις το 0.8% της συνολικής μεταβλητότητας του λογάριθμου των πωλήσεων.

	β	**Std. Error**	**R^2**	**p-value**
AGE60	1.517*	0.096	0.008	<0.001
EDUC	0.042*	0.054	<0.001	0.439
ETHNIC	0.325*	0.032	0.004	<0.001
INCOME	-0.154	0.021	0.002	<0.001
HHLARGE	-1.935*	0.198	0.003	<0.001
WORKWOM	-1.559*	0.113	0.006	<0.001
HVAL150	0.075*	0.025	<0.001	0.003

* για αύξηση 100% του ποσοστού

Πίνακας 8: Ανάλυση γραμμικής παλινδρόμησης με εξαρτημένη μεταβλητή το λογάριθμο των πωλήσεων και ανεξάρτητες τους δημογραφικούς παράγοντες

Προκειμένου να διασαφηνιστεί η πιθανή επίδραση κάποιων σημαντικών παραγόντων στις πωλήσεις, στον Πίνακα 9 παρουσιάζονται τα αποτελέσματα της ανάλυσης πολλαπλής γραμμικής παλινδρόμησης, με εξαρτημένη μεταβλητή τις πωλήσεις σε λογαριθμική κλίμακα. Ως ανεξάρτητοι πιθανοί παράγοντες επιλέχθηκαν η τιμή σε δολάρια (price), ο χρόνος σε εβδομάδες (week), η ύπαρξη διαφήμισης (feat) και το brand του προϊόντος. Όλοι οι παράγοντες βρέθηκε να επιδρούν σε στατιστικά σημαντικό βαθμό στις πωλήσεις, ενώ το μοντέλο επιτυγχάνει να ερμηνεύσει συνολικά το 48.6% (R^2=0.486) της συνολικής μεταβλητότητας του λογάριθμου των πωλήσεων.

Για αύξηση κατά 1 δολάριο στην τιμή του προϊόντος αναμένεται στατιστικά σημαντική μείωση κατά περίπου 1.1 μονάδες ($p<0.001$) στο λογάριθμο των πωλήσεων, για τις άλλες παραμέτρους σταθερές. Επίσης, για τις άλλες παραμέτρους σταθερές, αν το προϊόν διαφημίζεται αναμένεται να παρατηρηθεί στατιστικά σημαντική αύξηση στις πωλήσεις, η οποία κατά μέσο όρο ανέρχεται περίπου στη μία μονάδα σε λογαριθμική κλίμακα. Τέλος, η Tropicana αναμένεται να έχει κατά μέσο όρο 1.261 μονάδες περισσότερες και η MinuteMaid 0.567 μονάδες περισσότερες πωλήσεις (σε λογαριθμική κλίμακα) από την Dominicks, για τις άλλες μεταβλητές σταθερές.

	β	Std. Error	t	p-value
(Constant)	11.014	0.028	400.339	<0.001
price	-1.094	0.011	-103.561	<0.001
week	-0.002	<0.001	-13.384	<0.001
feat*	0.916	0.011	84.910	<0.001
Minute/Maid**	0.567	0.012	47.770	<0.001
Tropicana**	1.261	0.016	80.133	<0.001

*κατηγορία αναφοράς: απουσία διαφήμισης

** κατηγορία αναφοράς: «Dominicks»

Πίνακας 9: Πολλαπλή γραμμική παλινδρόμηση με εξαρτημένη μεταβλητή το λογάριθμο των πωλήσεων και ανεξάρτητες βασικούς παράγοντες (τιμή, χρόνος, διαφήμιση, brand)

Τέλος, εφόσον φάνηκε ότι οι παραπάνω παράγοντες πραγματικά επιτυγχάνουν να ερμηνεύσουν ένα σημαντικό ποσοστό της μεταβλητότητας των πωλήσεων, ελέγχθηκε κατά πόσο η συν-αξιολόγηση και των δημογραφικών στοιχείων μπορούν να βελτιώσουν την ερμηνευτική ικανότητα του μοντέλου. Στον Πίνακα 10 παρουσιάζονται τα ευρήματα της πολλαπλής γραμμικής παλινδρόμησης, με εξαρτημένη μεταβλητή τις πωλήσεις σε λογαριθμική κλίμακα και ανεξάρτητες την τιμή σε δολάρια (price), τον χρόνο σε εβδομάδες (week), την ύπαρξη διαφήμισης (feat), το brand του προϊόντος και τα δημογραφικά χαρακτηριστικά.

Μετά από έλεγχο για την ύπαρξη πολυσυγγραμμικότητας, εξαιρέθηκε αρχικά από την ανάλυση η μεταβλητή EDUC, η οποία αντιστοιχεί στο ποσοστό των πτυχιούχων (VIF=8.3), και στη συνέχεια η μεταβλητή INCOME, η οποία αντιστοιχεί στο λογάριθμο του διάμεσου εισοδήματος (VIF=4.5). Το τελικό μοντέλο επιτυγχάνει να ερμηνεύσει το 53.1% της συνολικής μεταβλητότητας του λογάριθμου των πωλήσεων. Έτσι, αν και όλα τα δημογραφικά χαρακτηριστικά επιδρούν σε στατιστικά σημαντικό βαθμού στις πωλήσεις, δεν βελτιώνουν πολύ την ερμηνευτική ικανότητα του μοντέλου με τους 5 αρχικούς παράγοντες, οι οποίοι παρουσιάστηκαν στον Πίνακα 9. Μάλιστα, ακόμη και αν δεν εξαιρεθούν οι δύο μεταβλητές που αποκλείστηκαν για να περιοριστεί η πολυσυγγραμμικότητα, η ερμηνευτική ικανότητα του μοντέλου δεν παρουσιάζει κάποια ουσιαστική μεταβολή (R^2=0.534).

Βρέθηκε ότι, σταθμίζοντας ως προς την τιμή, το brand, την ύπαρξη διαφήμισης και τον χρόνο, ο διπλασιασμός του ποσοστού των ηλικιωμένων άνω των 60 ετών αναμένεται να προκαλέσει μέση αύξηση κατά 1.2 μονάδες περίπου στο λογάριθμο των πωλήσεων ($p<0.001$), για τις άλλες μεταβλητές σταθερές. Από την άλλη, ο διπλασιασμός του ποσοστού των εργαζόμενων γυναικών, για τις άλλες μεταβλητές σταθερές αντιστοιχεί σε μέση μείωση στο λογάριθμο των πωλήσεων κατά 1.5 μονάδες ($p<0.001$).

	β	Std. Error	t	p-value
(Constant)	11.563	0.097	119.461	<0.001
price	-1.149	0.010	-111.678	<0.001
week	-0.002	<0.001	-15.806	<0.001
feat[a]	0.897	0.010	86.893	<0.001
Minute/Maid[b]	0.595	0.011	52.298	<0.001
Tropicana[b]	1.322	0.015	86.989	<0.001
AGE60	1.193[c]	0.126	9.502	<0.001
ETHNIC	0.903 [c]	0.032	28.106	<0.001
HHLARGE	-2.421 [c]	0.211	-11.482	<0.001
WORKWOM	-1.502 [c]	0.146	-10.261	<0.001
HVAL150	0.487 [c]	0.024	20.445	<0.001

[a]κατηγορία αναφοράς: απουσία διαφήμισης

[b] κατηγορία αναφοράς: «Dominics»

[c] για αύξηση 100% του ποσοστού

Πίνακας 10: Πολλαπλή γραμμική παλινδρόμηση με εξαρτημένη μεταβλητή το λογάριθμο των πωλήσεων και ανεξάρτητες βασικούς παράγοντες (τιμή, χρόνος, διαφήμιση) και δημογραφικά χαρακτηριστικά

Συνολικά, συγκρίνοντας τα αποτελέσματα του Πίνακα 10 με εκείνα των Πινάκων 8 και 9, παρατηρούμε ότι η εκτιμώμενη επίδραση κάθε ενός ξεχωριστά των δημογραφικών παραγόντων στις πωλήσεις δεν διαφοροποιήθηκε ιδιαίτερα όταν αξιολογήθηκαν όλοι μαζί και σταθμίζοντας το μοντέλο ως προς την τιμή, το χρόνο, τη διαφήμιση και το brand του προϊόντος. Έτσι, δεν φαίνεται να υπάρχει σημαντική συνεργατική επίδραση των παραγόντων που ελέγχθηκαν στις πωλήσεις. Όταν προστέθηκαν στο μοντέλο οι όροι αλληλεπίδρασης του brand με το ποσοστό έγχρωμων/ισπανόφωνων βρέθηκε ότι είναι στατιστικά σημαντικοί, αλλά δεν διαφοροποιούνται σημαντικά ούτε οι υπόλοιποι συντελεστές (β), ούτε η συνολική ερμηνευτική ικανότητα του μοντέλου (R^2=0.538). Έτσι, τα αποτελέσματα δεν παρουσιάζονται, αλλά παρατίθενται στο Παράρτημα 2.

Δυστυχώς, για τα δεδομένα που διαθέτουμε οι απόπειρες προβλέψεων που έγιναν δεν κρίνονται ερμηνεύσιμες λόγω των μεγάλων σφαλμάτων που προκύπτουν. Ενδεικτικά παρατίθεται ο παρακάτω πίνακας (από excel) από τον οποίο προκύπτει για δεδομένες τιμές των μεταβλητών X_j το διάστημα πρόβλεψης για τις πωλήσεις της dominicks [0,31 – 74,42]

Multiple Regression Results

	0	1	2	3	4	5	6	7	8	9	10
	Intercept	price	week	feat	MM	Tropicana	age60	ethnic	hhlarge	workwom	hval150
b	11,563	-1,149	-0,002	0,897	0,595	1,322	1,193	0,903	-2,421	-1,502	0,487
s(b)	0,097	0,01	1E-07	0,01	0,011	0,015	0,126	0,032	0,211	0,146	0,024
t	119,461	-111,678	-15,806	86,893	52,298	86,989	9,502	28,106	-11,482	-10,261	20,445
p-value	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000

ANOVA Table

Source	SS	df	MS	*F*	*p*-value
Regn.	6732,66	3	2244,22	3093,36	0,0000
Error	6997,4	9645	0,72549		
Total	13730,1	9648			

R^2 0,4904 Adjusted R^2 0,4902

Prediction Interval

Given X	price	week	feat	MM	Tropicana	age60	ethnic	hhlarge	workwom	hval150
	1,4	165	1	11,2	9	0,15	0,2	0,5	0,4	0,35

1-α	(1-α) CI for Y for given X		
95%	0,31319	+ or -	74,4211

1-α	(1-α) CI for E[Y | X]		
95%	0,31319	+ or -	74,4023

ΚΕΦΑΛΑΙΟ 5: ΣΥΜΠΕΡΑΣΜΑΤΑ

Στην εργασία αυτή μελετήθηκαν οι πωλήσεις φυσικών χυμών πορτοκαλιού και το πώς αυτές επηρεάζονται από παράγοντες όπως η τιμή, το εμπορικό σήμα, η διαφήμιση, τα δημογραφικά χαρακτηριστικά.

Τα δεδομένα που διαθέτουμε αφορούν τις πωλήσεις χυμών πορτοκαλιού συσκευασίας 2Lt (64oz) τριών διαφορετικών brands (Dominicks, MinuteMaid, Tropicana) σε 83 καταστήματα της ευρύτερης περιοχής του Σικάγο για διάρκεια 121 εβδομάδων

Έτσι λοιπόν είναι καταγεγραμμένα ανά εβδομάδα και ανά κατάστημα: οι πωλήσεις, η τιμή, η παρουσία/απουσία διαφήμισης, το ποσοστό ατόμων άνω των 60 ετών, το ποσοστό πτυχιούχων ατόμων, το ποσοστό έγχρωμου/ισπανόφωνου πληθυσμού, το μέσο εισόδημα, το ποσοστό νοικοκυριών 5 ή περισσότερων ατόμων, το ποσοστό εργαζόμενων γυναικών πλήρους απασχόλησης και το ποσοστό νοικοκυριών με περιουσιακά στοιχεία > $150.000

Από την ανάλυση αυτή προέκυψε ένας αριθμός συμπερασμάτων που συνοψίζονται παρακάτω:

Για τις τρεις διαφορετικές μάρκες:

- Η τιμή διαφέρει ανάλογα με το brand και το κατάστημα
- Η τιμή και ο χρόνος σχετίζονται μεταξύ τους ήπια αρνητικά: όσο περνά ο καιρός η τιμή εμφανίζεται ελαφρώς μειωμένη
- Η τιμή συσχετίζεται, όπως αναμενόταν, αρνητικά με τις πωλήσεις. Για το φθηνότερο brand (Dominicks) οι πωλήσεις μειώνονται συνεχώς όσο αυξάνεται η τιμή. Για τα δύο ακριβότερα brands παρατηρείται επίσης μείωση πωλήσεων μέχρι κάποια τιμή, αλλά από κάποια τιμή και έπειτα παύει να έχει σημασία η αύξηση στις τιμές και οι πωλήσεις είναι σταθερές ανεξαρτήτως τιμής.
- Γενικά, οι πωλήσεις των χυμών Tropicana είναι οι υψηλότερες ενώ ακολουθούν οι χυμοί MinuteMaid και τέλος οι Dominicks, παρόλο που οι χυμοί Tropicana είναι διαχρονικά ακριβότεροι και οι Dominicks φθηνότεροι.

- Οι πωλήσεις με διαφήμιση είναι υψηλότερες από όταν δεν υπάρχει διαφήμιση και για τα τρία brands.
- Το ποσοστό των πτυχιούχων επιδρά διαφορετικά στις πωλήσεις των τριών brands, με τις πωλήσεις του Tropicana να παρουσιάζουν θετική συσχέτιση με το εν λόγω ποσοστό, του Dominicks αρνητική συσχέτιση, ενώ οι πωλήσεις του MinuteMaid δε φαίνεται να επηρεάζονται.
- Το εισόδημα σαν δημογραφικός παράγοντας φαίνεται να επηρεάζει πτωτικά τις πωλήσεις του Dominicks, ενώ δε διαφοροποιεί ιδιαίτερα τις πωλήσεις των Tropicana και MinuteMaid, παρά μόνο στις υψηλότερες τιμές του.

Μεταξύ των δημογραφικών χαρακτηριστικών:

- Η εκπαίδευση (ποσοστό πτυχιούχων) και τα οικονομικά χαρακτηριστικά (εισόδημα, ποσοστό «ευκατάστατων» νοικοκυριών) σχετίζονται θετικά.
- Το ποσοστό έγχρωμων/ισπανόφωνων και το εισόδημα συσχετίζονται αρνητικά.

Πολλαπλή Παλινδρόμηση:

- Η τιμή, η ύπαρξη διαφήμισης και η επωνυμία του προϊόντος είναι οι σημαντικότεροι παράγοντες που επηρεάζουν τις πωλήσεις
- Τα δημογραφικά χαρακτηριστικά, αν και φαίνεται να επηρεάζουν στατιστικά σημαντικά τις πωλήσεις (τόσο κάθε χαρακτηριστικών χωριστά, όσο και όλα μαζί σταθμίζοντας ως προς τους 5 βασικούς παράγοντες), η επίδραση αυτή φαίνεται να είναι πολύ μικρή για να αξιοποιηθεί.
- Πέρα από την τιμή, τη διαφήμιση και την επωνυμία, ίσως υπάρχουν και άλλοι παράγοντες που πρέπει να ληφθούν υπόψιν ώστε να ερμηνευτεί το ύψος των πωλήσεων των χυμών.

Μια ανάλυση όπως της παρούσας εργασίας μπορεί να δώσει στην εκάστοτε επιχείρηση πώλησης φυσικών χυμών πληροφορίες για τους καταναλωτές και το περιβάλλον τους ώστε να προσδιοριστούν προβλήματα της στατηγικής μάρκετινγκ, να εντοπιστούν ευκαιρίες, να αναθεωρηθούν και να αξιολογηθούν προγράμματα μάρκετινγκ και να ληθφούν σχετικές αποφάσεις. Τα παραπάνω αποτελέσματα παρέχουν, δηλαδή, δεδομένα τα οποία δύνανται να ερμηνευθούν και:

- να απεικονίσουν την παρούσα κατάσταση, π.χ. ποιοι καταναλωτές αγοράζουν τι, ποιο είναι το μερίδιο αγοράς, τι εικόνα έχουν οι καταναλωτές για το κάθε brand κ.λπ. ή/και
- μετουσιωθούν σε στρατηγικές των επιχειρήσεων του κλάδου φυσικών χυμών, π.χ. τι είδους αλλαγές πρέπει να γίνουν στις διαφημιστικές καμπάνιες ώστε αυτές να καταστούν αποτελεσματικότερες, ποια πρέπει να είναι η τιμή του προϊόντος, ποιες τεχνικές προώθησης πρέπει να χρησιμοποιηθούν κ.λπ

Σε συνέχεια της παρούσας ανάλυσης και με σκοπό την καλύτερη σκιαγράφηση των πωλήσεων του κλάδου, περαιτέρω μελέτη μπορεί να πραγματοποιηθεί λαμβάνοντας υπόψη και εστιάζοντας και σε άλλους παράγοντες που πιθανόν σχετίζοναι με την κατανάλωση φυσικών χυμών και διαφροποιούν τις πωλήσεις τους. Τέτοιοι παράγοντες μπορεί να αφορούν το γενικότερο τρόπο ζωής (lifestyle) του καταναλωτή όπως οι διατροφικές του συνήθειες, η σχέση με αθλητικές δρατηριότητες, η εργασιακή κατάσταση, η οικογενιακή κατάσταση κ.λπ.

Έτσι, η παρούσα εργασία μπορεί να αποτελέσει μία βάση αναφοράς για πιο στοχευμένη έρευνα σχετικά με τον τομέα των φυσικών χυμών πορτοκαλιού.

ΒΙΒΛΙΟΓΡΑΦΙΑ

[1] **Γούναρης Σπύρος**, *Μαρκετινγκ Υπηρεσιών*, Rosili, 2003

[2] **Σταθακόπουλος Βλάσης**, *Μέθοδοι Έρευνας Αγοράς*, Α. Σταμούλης, 2001

[3] **Χαλικιάς Ιωάννης**, *Στατιστική - Μέθοδοι ανάλυσης για επιχειρηματικές αποφάσεις*, Rosili, 2001

[4] **Ψιλούτσικου Μαρίνα**, *Ποσοτικές Μέθοδοι ΙΙ – SPSS*, Οικονομικό Πανεπιστήμιο Αθηνών, 2003

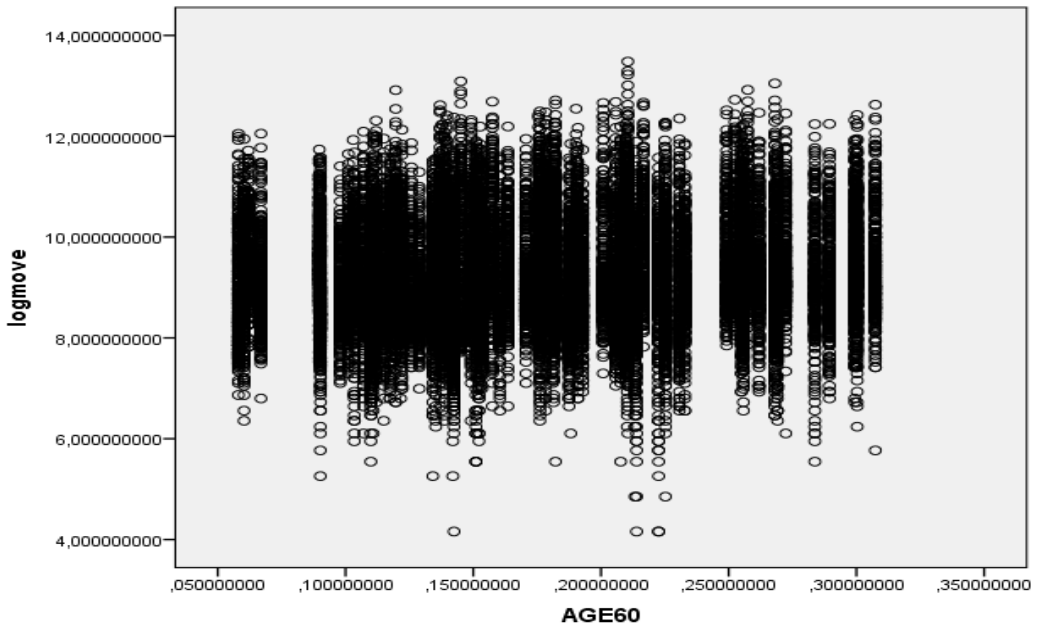
[5] **Aczel Amir, Sounderandian Jayavel,** *Στατιστική σκέψη στον κόσμο των επιχειρήσεων*, McGraw-Hill, 2009

[6] **Feleke Shiferaw, Kilmer Richard**, *Global competition for the Japanese fruit juice market: A Uniform Substitute Demand Analysis,* American Agricultural Economics Association Annual Meeting, 2007

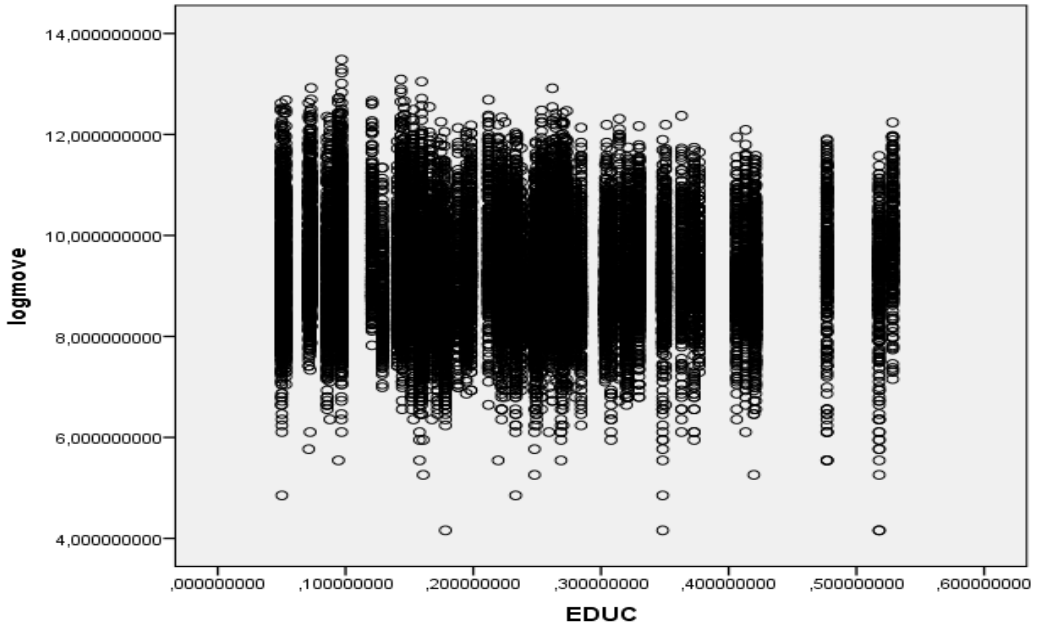
[7] **Ledolter Johannes**, *Data mining and business analytics with R*, Wiley, 2013

[8] **Montgomery Alan**, *Creating Micro-Marketing Pricing Strategies Using Supermarket Scanner Data*, Marketing Science, Vol. 16, No. 4, 1997

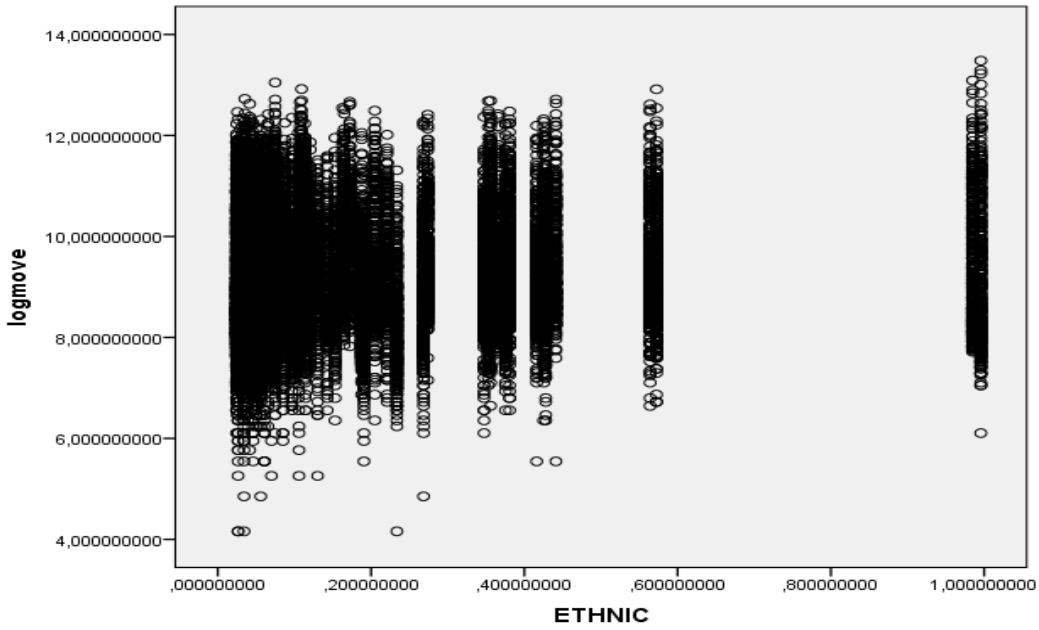
Παράρτημα 1



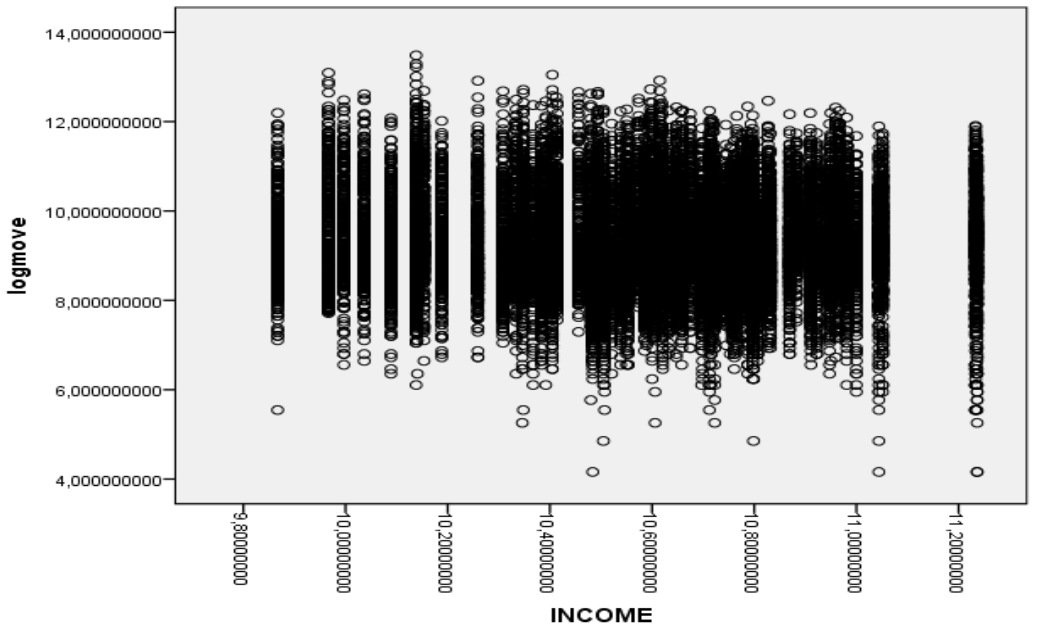
Διάγραμμα Π1.1



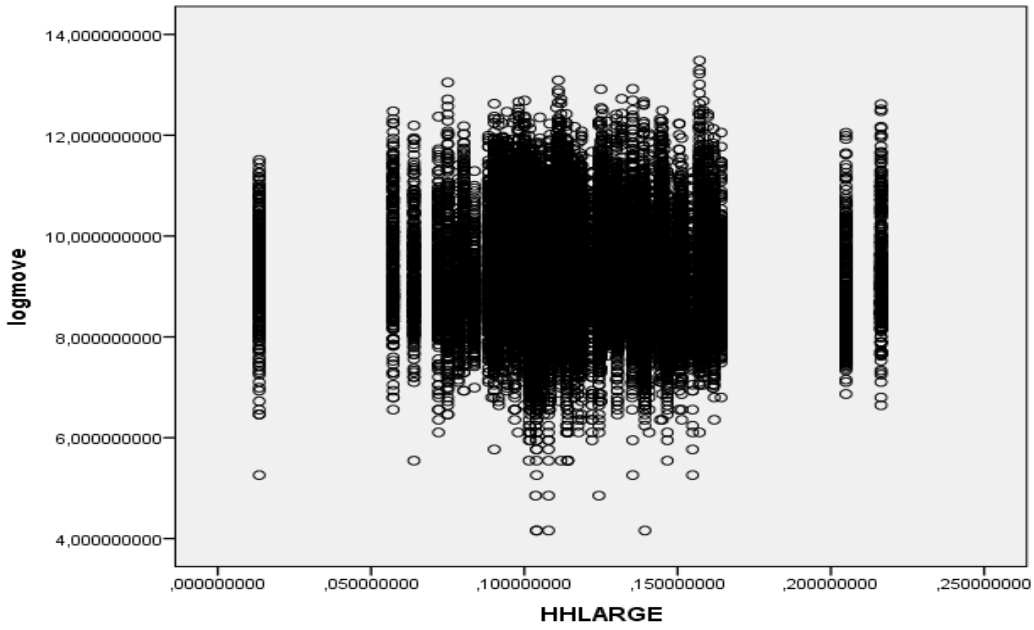
Διάγραμμα Π1.2



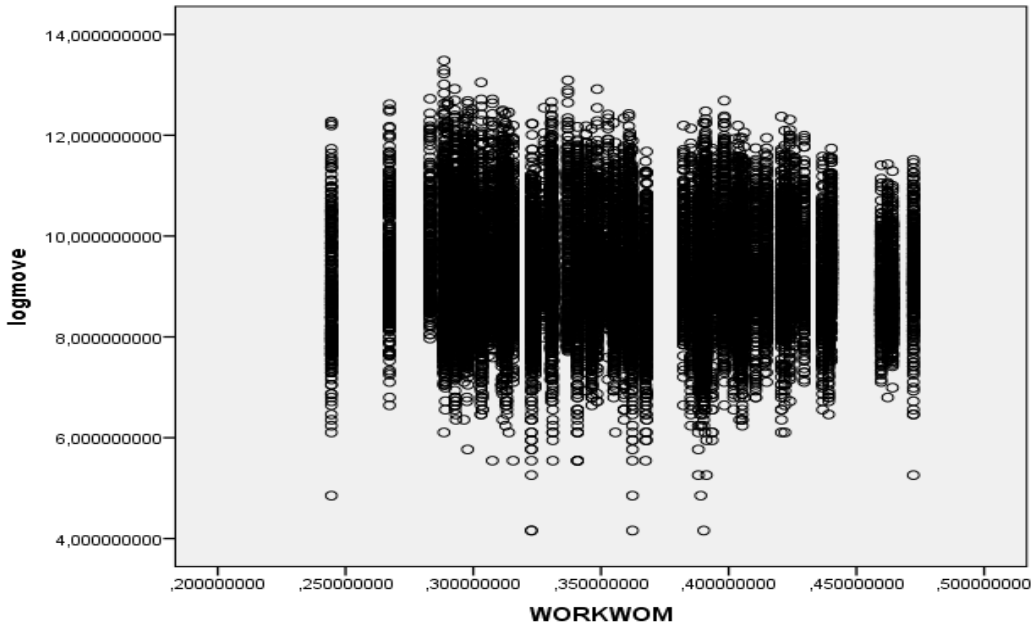
Διάγραμμα Π1.3



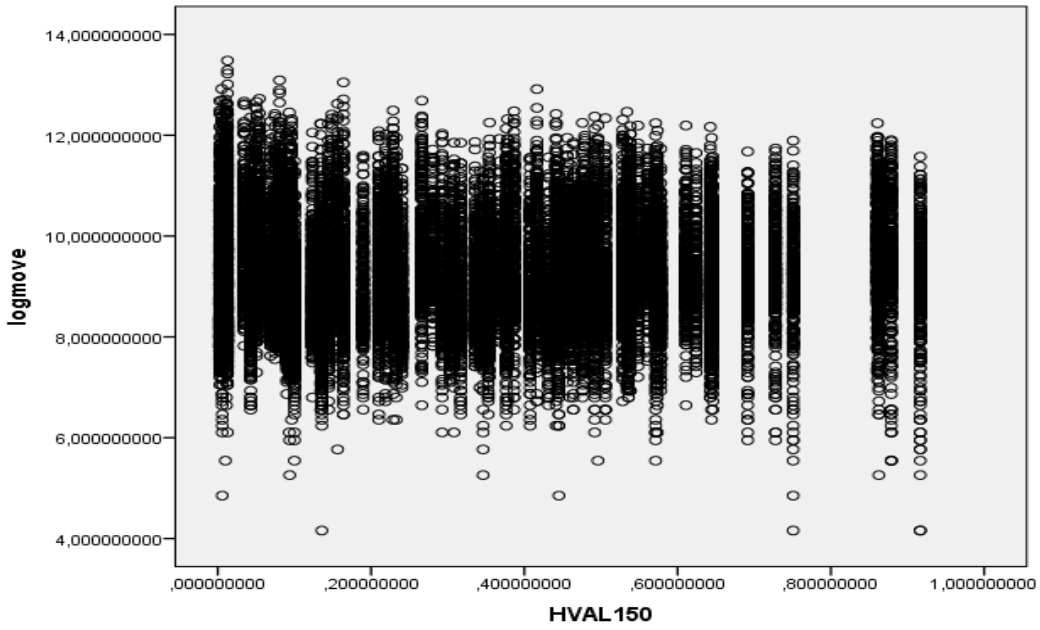
Διάγραμμα Π1.4



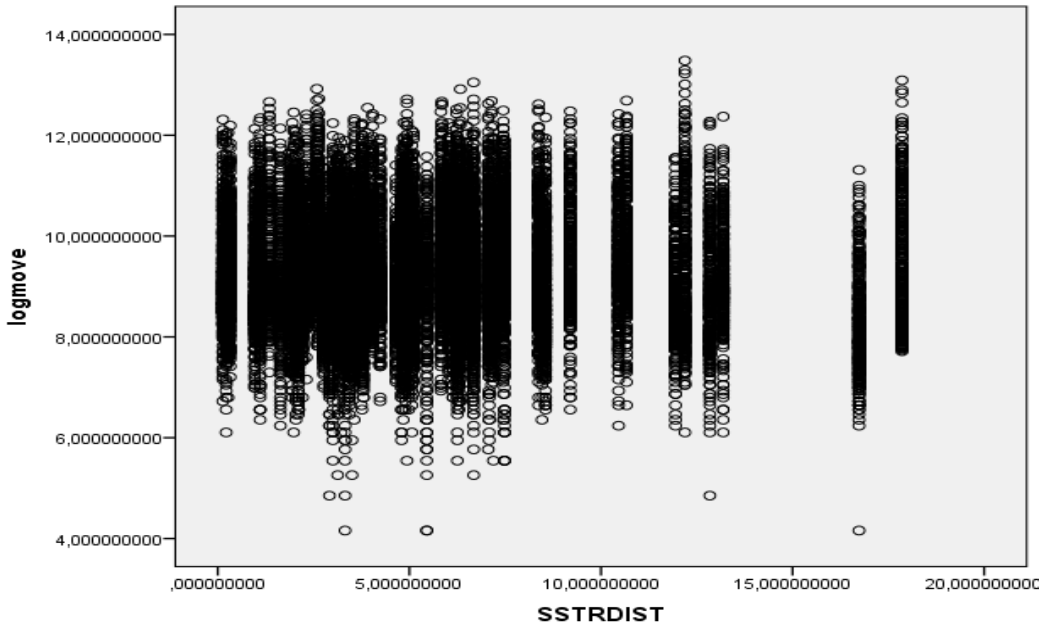
Διάγραμμα Π1.5



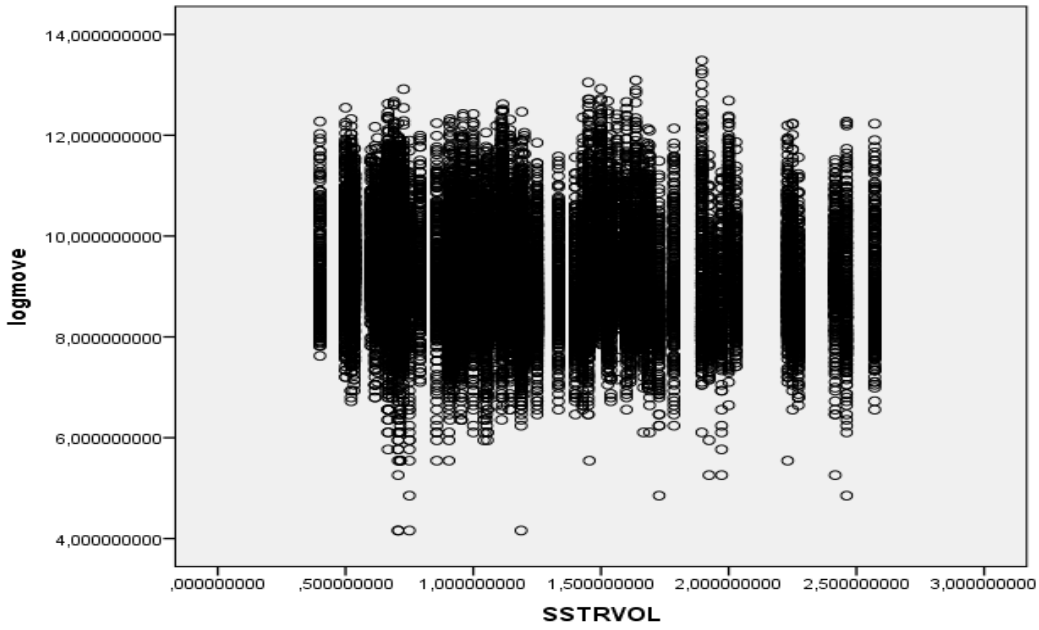
Διάγραμμα Π1.6



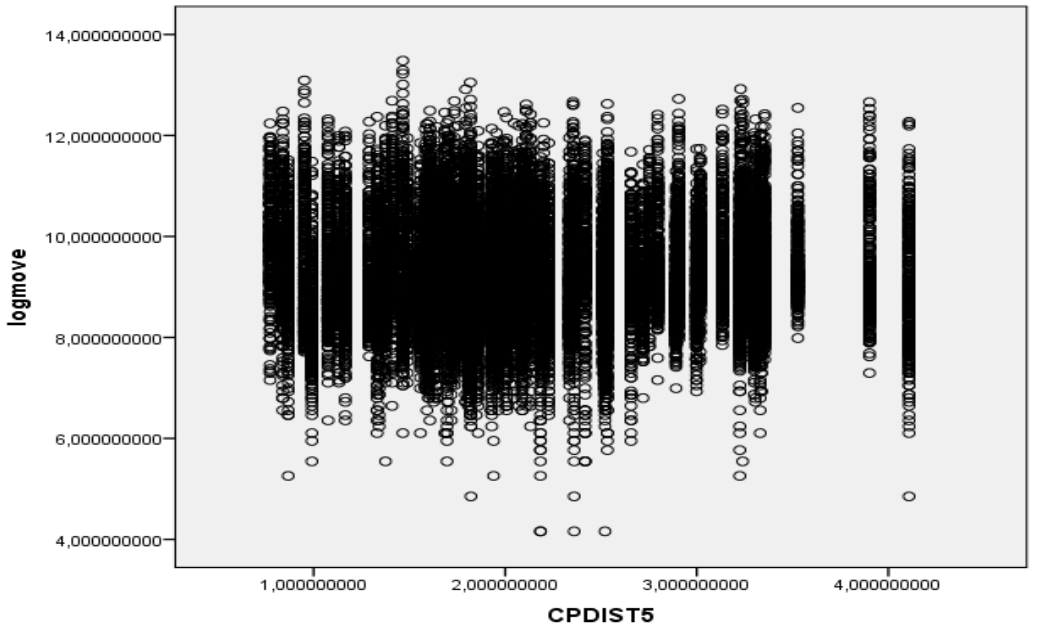
Διάγραμμα Π1.71



Διάγραμμα Π1.8



Διάγραμμα Π1.9



Διάγραμμα Π1.102

Παράρτημα 2

	β	τ.α.	t	p-value
(Constant)	11,466	,096	119,231	,000
price	-1,147	,010	-112,351	,000
week	-,002	,000	-15,873	,000
feat	,898	,010	87,664	,000
Minute/Maid	,697	,014	49,758	,000
Tropicana	1,495	,017	87,176	,000
AGE60	1,192	,125	9,567	,000
ETHNIC	1,499	,044	33,867	,000
HHLARGE	-2,422	,209	-11,573	,000
WORKWOM	-1,502	,145	-10,340	,000
HVAL150	,487	,024	20,587	,000
SSTRDIST	-,004	,001	-3,019	,003
SSTRVOL	-,145	,008	-17,600	,000
CPDIST5	,035	,006	5,699	,000
Minute/Maid *ethnic	-,662	,053	-12,454	,000
Tropicana *ethnic	-1,126	,053	-21,165	,000

Πίνακας Π2.1 Πολλαπλή γραμμική παλινδρόμηση με εξαρτημένη μεταβλητή το λογάριθμο των πωλήσεων και ανεξάρτητες βασικούς παράγοντες (τιμή, χρόνος, διαφήμιση) και δημογραφικά χαρακτηριστικά και αλληλεπιδράσεις