



Γεωπονικό Πανεπιστήμιο Αθηνών
Σχολή Εφαρμοσμένων Οικονομικών και Κοινωνικών Επιστημών
Τμήμα Αγροτικής Οικονομίας και Ανάπτυξης

**ΜΗ ΓΡΑΜΜΙΚΕΣ ΕΠΙΔΡΑΣΕΙΣ ΣΕ ΓΡΑΜΜΙΚΑ ΚΑΙ ΜΗ-ΓΡΑΜΜΙΚΑ
ΥΠΟΔΕΙΓΜΑΤΑ ΠΙΘΑΝΟΤΗΤΩΝ: ΜΙΑ ΕΜΠΕΙΡΙΚΗ ΕΦΑΡΜΟΓΗ ΣΤΗΝ
ΣΥΜΠΕΡΙΦΟΡΑ ΚΑΤΑΝΑΛΩΤΗ ΓΥΡΩ ΑΠΟ ΤΑ ΤΡΟΦΙΜΑ**

ΠΤΥΧΙΑΚΗ ΜΕΛΕΤΗ

ΚΟΡΑΚΙΤΗ ΜΑΡΙΑ

ΕΞΕΤΑΣΤΙΚΗ ΕΠΙΤΡΟΠΗ

Δριχούτης Ανδρέας, Αναπληρωτής Καθηγητής (επιβλέπων)

Κλωνάρης Ευστάθιος, Καθηγητής

Μαλέσιος Χρσοβαλάντης, Επίκουρος Καθηγητής

Αθήνα, Ιούνιος 2021

ΠΕΡΙΛΗΨΗ

Στα εφαρμοσμένα οικονομικά είναι συνηθισμένο για τους ερευνητές να ενσωματώνουν σε διχοτομικά υποδείγματα πιθανότητας όρους αλληλεπίδρασης όταν θέλουν να λάβουν υπόψη την έννοια της μη-γραμμικότητας. Κάτι τέτοιο δεν είναι πάντα ορθό αφού παρατηρούνται μη γραμμικές επιδράσεις και χωρίς την ενσωμάτωση όρων αλληλεπίδρασης σε ένα υπόδειγμα. Είναι δυνατό κάποιες φορές, η αλόγιστη χρήση όρων αλληλεπίδρασης να οδηγεί σε εσφαλμένα συμπεράσματα για τις οριακές επιδράσεις των μεταβλητών. Ο σκοπός της παρούσας πτυχιακής μελέτης είναι να διερευνήσει τις μη-γραμμικές επιδράσεις σε γραμμικά και μη-γραμμικά υποδείγματα. Ακόμη, η πτυχιακή έχει στόχο να μελετήσει τις επιδράσεις των όρων αλληλεπίδρασης και να διερευνήσει κατά πόσο αυτές οδηγούν σε ορθά συμπεράσματα για την επίδραση των ανεξάρτητων μεταβλητών επάνω στην εξαρτημένη μεταβλητή. Προκειμένου να μελετηθεί αυτό, χρησιμοποιήθηκαν δευτερογενή δεδομένα από την Εθνική Έρευνα για την Υγεία και τη Διατροφή (National Health and Nutrition Examination Survey-NHANES), που αφορούν δεδομένα διατροφής και υγείας. Τα συμπεράσματα από την ανάλυση έδειξαν πως στο 20,2% των περιπτώσεων, εάν εισάγουμε όρο αλληλεπίδρασης στο υπόδειγμα μας θα καταλήξουμε σε λανθασμένα ερμηνεία των επιδράσεων των μεταβλητών. Το ποσοστό αυτό, υποδεικνύει πως είναι σημαντικό για έναν ερευνητή να αναλογιστεί καλύτερα κατά πόσο πρέπει να εισάγει όρους αλληλεπίδρασης σε ένα υπόδειγμα διακριτής απόκρισης προκειμένου να συμπεριλάβει μη-γραμμικές επιδράσεις των μεταβλητών. Πέρα του ότι δείχνουμε ότι μη-γραμμικές επιδράσεις υφίστανται ακόμη και σε υποδείγματα που δεν περιλαμβάνουν όρους αλληλεπίδρασης, είναι πολύ πιθανό να καταλήξει ο ερευνητής, σε λανθασμένες εκτιμήσεις και ερμηνεία της επίδρασης των ανεξάρτητων μεταβλητών.

Λέξεις Κλειδιά: Διχοτομικό υπόδειγμα πιθανότητας, όροι αλληλεπίδρασης, οριακές επιδράσεις, probit, logit

ABSTRACT

In applied economics it is common for researchers to incorporate interaction terms into binary probability models when they want to allow for non-linear effects. This is not always correct since non-linear effects are observed without the inclusion of interaction terms into a model. It is therefore possible that the reckless use of interaction terms will lead to erroneous conclusions about the marginal effects of the variables. The purpose of this thesis, is to investigate the non-linear effects on linear and non-linear models. Furthermore, the thesis aims to study the effects of interaction terms and to investigate whether they lead to correct conclusions about the effect of independent variables on the dependent variable. To study our research question we use secondary data from the National Health and Nutrition Examination Survey (NHANES) regarding label use. Our findings show that in 20.2% of the cases, if we introduce an interaction term in our model we will end up with a misinterpretation of the effects of the variables. This percentage indicates that it is important for a researcher to better consider whether they should introduce interaction terms into a discrete response model in order to allow for non-linear effects of variables. In addition to showing that nonlinear effects exist even in models with no interaction terms, it is very likely that the researcher will come to erroneous estimates and interpretations of the effect of the independent variables.

Key Words: Binary model of probability, interaction terms, marginal effects, probit, logit

ΔΗΛΩΣΗ ΕΡΓΟΥ

Ο/Η κάτωθι υπογεγραμμένος/η φοιτήτρια, **Κορακίτη Μαρία**, δηλώνω ρητά ότι η παρούσα Πτυχιακή Εργασία με τίτλο «**Μη γραμμικές επιδράσεις σε γραμμικά και μη-γραμμικά υποδείγματα πιθανοτήτων: μια εμπειρική εφαρμογή στην συμπεριφορά καταναλωτή γύρω από τα τρόφιμα**» καθώς και τα ηλεκτρονικά αρχεία και πηγαίοι κώδικες που αναπτύχθηκαν ή τροποποιήθηκαν στα πλαίσια αυτής της εργασίας και αναφέρονται ρητώς μέσα στο κείμενο που συνοδεύουν, και η οποία έχει εκπονηθεί στο Τμήμα Αγροτικής Οικονομίας και Ανάπτυξης του Γεωπονικού Πανεπιστημίου Αθηνών, υπό την επίβλεψη του μέλους ΔΕΠ του Τμήματος κ. **Ανδρέα Δριχούτη**, αποτελεί αποκλειστικά δικό μου, μη υποβοηθούμενο πόνημα, δεν προσβάλλει κάθε μορφής πνευματικά δικαιώματα τρίτων και δεν είναι προϊόν μερικής ή ολικής αντιγραφής. Τα σημεία όπου έχουν χρησιμοποιηθεί ιδέες, κείμενο, αρχεία ή / και πηγές άλλων συγγραφέων, αναφέρονται ευδιάκριτα στο κείμενο με την κατάλληλη παραπομπή και η σχετική αναφορά περιλαμβάνεται στο τμήμα των βιβλιογραφικών αναφορών με πλήρη περιγραφή. Η πτυχιακή εργασία αυτή υποβάλλεται σε μερική εκπλήρωση των απαιτήσεων για την απόκτηση Πτυχίου από το τμήμα Αγροτικής Οικονομίας και Ανάπτυξης, του Γεωπονικού Πανεπιστημίου Αθηνών. Δεν έχει υποβληθεί ποτέ πριν για οιοδήποτε λόγο ή για εξέταση σε οποιοδήποτε άλλο πανεπιστήμιο ή εκπαιδευτικό ίδρυμα της χώρας ή του εξωτερικού. Η πτυχιακή εργασία αποτελεί προϊόν συνεργασίας του/της φοιτητή/τρια και του επιβλέποντος της εκπόνησής της. Τα φυσικά αυτά πρόσωπα έχουν και τα πνευματικά δικαιώματα στη δημοσίευση των αποτελεσμάτων της πτυχιακή εργασίας σε επιστημονικά περιοδικά και συνέδρια. Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και μόνο..

Κορακίτη Μαρία

17/06/2021



ΕΥΧΑΡΙΣΤΙΕΣ

Θα ήθελα να εκφράσω τις ευχαριστίες μου στον επιβλέποντα καθηγητή μου Δρ. Ανδρέα Δριχούτη για την εμπιστοσύνη που μου έδειξε και την υπομονή του κατά τη διάρκεια υλοποίησης της πτυχιακής μου εργασίας. Ακόμη, χάρις την πολύτιμη βοήθεια και καθοδήγηση του κατάφερα να επιλύσω διάφορα θέματα και προβλήματα που προέκυπταν όλο αυτό το διάστημα.

Θα ήθελα να ευχαριστήσω επιπλέον τον κ. Ευστάθιο Κλωνάρη και τον κ. Μαλέσιο Χρυσοβαλάντη για την προθυμία τους να συνδράμουν στην εξέταση της πτυχιακής μου εργασίας ως μέλη της εξεταστικής επιτροπής.

Τέλος, θα ήθελα να ευχαριστήσω την οικογένειά μου, που με στηρίζει σε κάθε μου βήμα.

Πίνακας περιεχομένων

1.	Υπόδειγμα Πιθανότητας.....	10
1.1	Το Γραμμικό Υπόδειγμα Πιθανότητας (Linear Probability Model)	10
1.2	Προβλήματα με το Γραμμικό Υπόδειγμα Πιθανότητας.....	11
1.3	Το υπόδειγμα της λανθάνουσας μεταβλητής για διχοτομικές εξαρτημένες μεταβλητές.....	12
2.	Συνάρτηση Πιθανοφάνειας.....	16
2.1	Η έννοια της συνάρτησης πιθανοφάνειας.....	16
2.2	Μέθοδος της Μεγίστης Πιθανοφάνειας	16
2.3	Η συνάρτηση πιθανοφάνειας για το διχοτομικό υπόδειγμα πιθανότητας	17
2.4	Ερμηνεία: η επίδραση των παραμέτρων.....	17
2.5	Το τεστ του λόγου της πιθανοφάνειας (Likelihood Ratio Test- LR test).....	19
3.	Οριακές Επιδράσεις (Marginal Effects) και Όροι Αλληλεπίδρασης (Interaction Effects) 20	
3.1	Οριακές Επιδράσεις (Marginal Effects)	20
3.2	Διακριτή Αλλαγή (Discrete Change).....	21
3.3	Όροι αλληλεπίδρασης (Interaction Effects)	22
3.4	Οριακές επιδράσεις των όρων αλληλεπίδρασης.....	23
3.4.1	OLS υποδείγματα παλινδρόμησης.....	23
3.4.2	Υποδείγματα παλινδρόμησης probit και logit	23
4.	Η έννοια της διατροφικής ετικέτας.....	25
4.1	Παράγοντες που επιδρούν στην χρήση της διατροφικής ετικέτας	25
5.	Σκοπός της μελέτης	28
	Περιγραφή της Έρευνας.....	28
5.1	Περιγραφή των δεδομένων	28
5.2	Περιγραφή των μεταβλητών του υποδείγματος και περιγραφική στατιστική.....	30
6.	Εξειδίκευση του Υποδείγματος- Το Διχοτομικό Υπόδειγμα Πιθανότητας.....	36
7.	Ερευνητικά Ερωτήματα, Μεθοδολογία Έρευνας και Αποτελέσματα.....	40
7.1	Όροι Αλληλεπίδρασης και Σύγκριση Υποδειγμάτων	40
7.2	Οριακές Επιδράσεις σε Γραμμικά και Μη-γραμμικά Υποδείγματα.....	44
7.3	Όροι αλληλεπίδρασης και εσφαλμένα συμπεράσματα	58
8.	Συμπεράσματα	61
9.	Βιβλιογραφία.....	63

Εικόνα 1: Το γραμμικό Υπόδειγμα Πιθανότητας	10
Εικόνα 2: Η κατανομή της γ^* δεδομένου της X σε ένα διχοτομικό υπόδειγμα παλινδρόμησης	13
Εικόνα 3: Κανονική και λογιστική κατανομή	15
Εικόνα 4: Οι επιδράσεις της αλλαγής της κλίσης και του σταθερού όρου σε ένα απλό διχοτομικό υπόδειγμα παλινδρόμησης	18
Εικόνα 5: Ιστόγραμμα της μεταβλητής Age	32
Εικόνα 6: Ιστόγραμμα της μεταβλητής Educ	33
Εικόνα 7: Ιστόγραμμα της μεταβλητής Income σε σχέση με την μεταβλητή Readlab	33
Εικόνα 8: Ραβδόγραμμα της μεταβλητής Nutrition	34
Εικόνα 9: Ραβδόγραμμα της μεταβλητής Price	34
Εικόνα 10: Ιστόγραμμα της μεταβλητής Money	35
Εικόνα 11: Οριακές επιδράσεις του φύλου καθώς μεταβάλλεται η ηλικία για το υπόδειγμα χωρίς αλληλεπίδραση	46
Εικόνα 12: Οριακές επιδράσεις του φύλου καθώς μεταβάλλεται η ηλικία για το υπόδειγμα με αλληλεπίδραση	46
Εικόνα 13: Οριακές επιδράσεις της ηλικίας καθώς μεταβάλλεται το εκπαιδευτικό επίπεδο για το υπόδειγμα χωρίς αλληλεπίδραση	49
Εικόνα 14: Οριακές επιδράσεις της ηλικίας καθώς μεταβάλλεται το εκπαιδευτικό επίπεδο για το υπόδειγμα με αλληλεπίδραση	49
Εικόνα 15: Οριακές επιδράσεις της εκπαίδευσης καθώς μεταβάλλεται η συχνότητα αγοράς τροφίμων για το υπόδειγμα χωρίς αλληλεπίδραση	52
Εικόνα 16: Οριακές επιδράσεις της εκπαίδευσης καθώς μεταβάλλεται η συχνότητα αγοράς τροφίμων για το υπόδειγμα με αλληλεπίδραση	52
Εικόνα 17: Οριακές επιδράσεις του εάν κάποιος ακολουθεί μια συγκεκριμένη διαίτα (diet) καθώς μεταβάλλεται η σημασία της διατροφικής αξίας (nutrition) για το υπόδειγμα χωρίς αλληλεπίδραση	55
Εικόνα 18: Οριακές επιδράσεις του εάν κάποιος ακολουθεί μια συγκεκριμένη διαίτα (diet) καθώς μεταβάλλεται η σημασία της διατροφικής αξίας (nutrition) για το υπόδειγμα με αλληλεπίδραση	56

Πίνακας 1: Μέτρα περιγραφικής στατιστικής για τις μεταβλητές του υποδείγματος	31
Πίνακας 2: Αποτελέσματα παλινδρόμησης της μεταβλητής <i>Readlab</i>	36
Πίνακας 3: Έλεγχος λόγου πιθανοφάνειας για την αλληλεπίδραση της μεταβλητής <i>Gender</i> και <i>Income</i>	42
Πίνακας 4: Κριτήρια πληροφόρησης AIC για την αλληλεπίδραση της μεταβλητής <i>Gender</i> και <i>Income</i>	42
Πίνακας 5: Έλεγχος λόγου πιθανοφάνειας για την αλληλεπίδραση της μεταβλητής <i>Age</i> και <i>Educ</i>	43
Πίνακας 6: Κριτήρια πληροφόρησης AIC για την αλληλεπίδραση της μεταβλητής <i>Age</i> και <i>Educ</i>	43
Πίνακας 7: Σύγκριση υποδειγμάτων με και χωρίς αλληλεπίδραση	43
Πίνακας 8: Αποτελέσματα <i>t-test</i> για την μεταβλητή <i>Age</i> και <i>Gender</i> χωρίς αλληλεπίδραση	47
Πίνακας 9: Αποτελέσματα <i>t-test</i> για την μεταβλητή <i>Age</i> και <i>Gender</i> με αλληλεπίδραση	47
Πίνακας 10: Αποτελέσματα <i>t-test</i> για την μεταβλητή <i>Age</i> και <i>Educ</i> χωρίς αλληλεπίδραση ..	50
Πίνακας 11: Αποτελέσματα <i>t-test</i> για την μεταβλητή <i>Age</i> και <i>Educ</i> με αλληλεπίδραση	50
Πίνακας 12: Αποτελέσματα <i>t-test</i> για την μεταβλητή <i>Educ</i> και <i>Purchase</i> με αλληλεπίδραση	53
Πίνακας 13: Αποτελέσματα <i>t-test</i> για την μεταβλητή <i>Educ</i> και <i>Purchase</i> χωρίς αλληλεπίδραση	54
Πίνακας 14: Αποτελέσματα <i>t-test</i> για την μεταβλητή <i>Diet</i> και <i>Nutrition</i> χωρίς αλληλεπίδραση	56
Πίνακας 15: Αποτελέσματα <i>t-test</i> για την μεταβλητή <i>Diet</i> και <i>Nutrition</i> χωρίς αλληλεπίδραση	57
Πίνακας 16: Σύγκριση οριακών επιδράσεων μεταξύ των υποδειγμάτων με και χωρίς αλληλεπίδραση	57
Πίνακας 17: Περιπτώσεις εσφαλμένων συμπερασμάτων όταν χρησιμοποιούμε υπόδειγμα με όρο αλληλεπίδρασης.....	60

Εισαγωγή

Στα εφαρμοσμένα οικονομικά είναι συνηθισμένο για τους ερευνητές να ενσωματώνουν σε διχοτομικά υποδείγματα πιθανότητας όρους αλληλεπίδρασης όταν θέλουν να λάβουν υπόψη την έννοια της μη-γραμμικότητας. Αυτό σημαίνει ότι ενσωματώνουν μη-γραμμικές επιδράσεις στα υποδείγματα με στόχο οι ανεξάρτητες μεταβλητές να επιδρούν στην εξαρτημένη με μη-γραμμικό τρόπο. Παρόλο αυτά, κάτι τέτοιο δεν είναι πάντα ορθό εφόσον παρατηρούνται μη γραμμικές επιδράσεις και χωρίς την ενσωμάτωση όρων αλληλεπίδρασης σε ένα υπόδειγμα. Είναι δυνατό κάποιες φορές, η αλόγιστη χρήση όρων αλληλεπίδρασης να οδηγεί σε εσφαλμένα συμπεράσματα για τις οριακές επιδράσεις των μεταβλητών.

Σκοπός της παρούσας πτυχιακής μελέτης είναι να διερευνήσει τις μη-γραμμικές επιδράσεις σε γραμμικά και μη-γραμμικά υποδείγματα. Ακόμη, η πτυχιακή έχει στόχο να μελετήσει τις επιδράσεις των όρων αλληλεπίδρασης και να διερευνήσει κατά πόσο αυτές οδηγούν σε ορθά συμπεράσματα για την επίδραση των ανεξάρτητων μεταβλητών επάνω στην εξαρτημένη. Προκειμένου να μελετηθεί αυτό, χρησιμοποιήθηκαν δευτερογενή δεδομένα από την Εθνική Έρευνα για την Υγεία και τη Διατροφή (National Health and Nutrition Examination Survey-NHANES), που αφορούν δεδομένα διατροφής και υγείας.

Προκειμένου να γίνουν κατανοητά τα ερευνητικά ερωτήματα της πτυχιακής μελέτης και να λάβουμε μια ολοκληρωμένη θεώρηση του θέματος, στο πρώτο μέρος της εργασίας παρατίθενται η θεωρία που θα χρησιμοποιήσουμε στην συνέχεια. Στο 1^ο κεφάλαιο αναφερόμαστε στα Υποδείγματα Πιθανότητας, γραμμικά και μη-γραμμικά. Στο 2^ο κεφάλαιο εστιάζουμε στην έννοια της Πιθανοφάνειας και του ελέγχου του λόγου της Πιθανοφάνειας, ενώ στο 3^ο κεφάλαιο εξηγούμε τις έννοιες των οριακών επιδράσεων και των όρων αλληλεπίδρασης.

Στο 4^ο κεφάλαιο γίνεται μια σύντομη αναφορά στην έννοια της διατροφικής επισήμανσης και αναλύονται οι προσδιοριστικοί παράγοντες που παρακινούν ένα άτομο να διαβάσει τις ετικέτες των τροφίμων που αγοράζει, μερικούς από τους οποίους θα συμπεριλάβουμε στο υπόδειγμα παλινδρόμησης που θα μελετήσουμε.

Εν συνεχεία, στο 5^ο και 6^ο κεφάλαιο προχωράμε στην οικονομετρική ανάλυση όπου εξετάζουμε την εξειδίκευση του υποδείγματος μας. Στο 5^ο κεφάλαιο περιγράφουμε την έρευνα από την οποία προήλθαν τα δευτερογενή δεδομένα, τα ίδια τα δεδομένα και τις μεταβλητές που χρησιμοποιήσαμε, παραθέτοντας μέτρα και διαγράμματα περιγραφικής στατιστικής. Επίσης, στο 6^ο κεφάλαιο εστιάζουμε στην ερμηνεία των αποτελεσμάτων της παλινδρόμησης.

Τέλος, στο 7^ο κεφάλαιο παρουσιάζονται και αναλύονται τα αποτελέσματα της έρευνας που αφορούν σε τρία ερευνητικά ερωτήματα που μελετήσαμε. Στο πρώτο μέρος ασχοληθήκαμε με την διερεύνηση του κατά πόσο ένα υπόδειγμα με όρο αλληλεπίδρασης προσαρμόζεται καλύτερα στα δεδομένα σε σύγκριση με ένα υπόδειγμα χωρίς όρο αλληλεπίδρασης. Στο δεύτερο μέρος διερευνήσαμε το κατά πόσο οδηγούμαστε σε διαφορετικά συμπεράσματα για τον τρόπο που επιδρούν οι ανεξάρτητες μεταβλητές στην εξαρτημένη όταν

συμπεριλαμβάνουμε στο υπόδειγμα μας όρο αλληλεπίδρασης σε σχέση με ένα υπόδειγμα που δεν περιλαμβάνει αλληλεπίδραση. Στο τρίτο και τελευταίο μέρος, καταλήξαμε στις περιπτώσεις όπου οδηγούμαστε σε εσφαλμένο συμπέρασμα εάν χρησιμοποιήσουμε σε ένα υπόδειγμα όρο αλληλεπίδρασης.

Στο 8^ο κεφάλαιο αποτυπώνονται τα συμπεράσματα της ερευνητικής μας μελέτης καθώς και οι δυνατότητες αξιοποίησης αυτής για περαιτέρω έρευνα.

1. Υπόδειγμα Πιθανότητας

1.1 Το Γραμμικό Υπόδειγμα Πιθανότητας (Linear Probability Model)

Το γραμμικό υπόδειγμα πιθανότητας (ΓΥΠ) είναι ένα υπόδειγμα παλινδρόμησης όπου η εξαρτημένη μεταβλητή είναι διχοτομική και οι εκτιμήσεις των παραμέτρων υπολογίζονται με τη μέθοδο ελαχίστων τετραγώνων. Το υπόδειγμα περιγράφεται από την εξίσωση:

$$y_i = \mathbf{X}_i\boldsymbol{\beta} + \varepsilon_i \quad (1)$$

Όπου \mathbf{X}_i είναι το διάνυσμα τιμών της i -νιοστής παρατήρησης, $\boldsymbol{\beta}$ είναι το διάνυσμα των εκτιμώμενων παραμέτρων και ε_i είναι ο όρος σφάλματος. Η εξαρτημένη μεταβλητή παίρνει τιμή $y_i=1$ όταν το γεγονός πραγματοποιείται και $y_i=0$ όταν το γεγονός δεν πραγματοποιείται.

Για παράδειγμα $y=1$ αν μια γυναίκα συμμετέχει στην αγορά εργασίας και $y=0$ αν δεν συμμετέχει. Αν το υπόδειγμα μας περιέχει μια ανεξάρτητη μεταβλητή μπορεί να γραφεί ως:

$$y_i = \alpha + \beta X_i + \varepsilon_i \quad (2)$$

Η προσδοκώμενη τιμή της y δεδομένου της X , $E(y|X)=a+bX$, απεικονίζεται από την συμπαγή κόκκινη γραμμή στο Σχήμα 1. Οι παρατηρήσεις φαίνονται σαν κύκλοι στις τιμές $y=0$ και $y=1$.

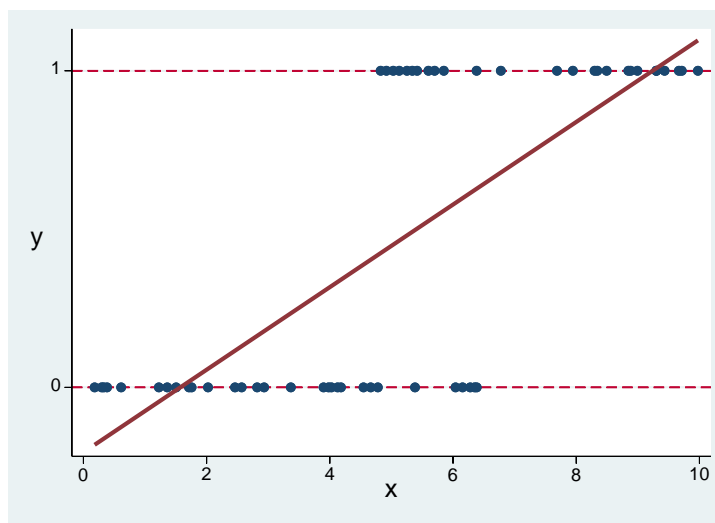
Για να καταλάβουμε το ΓΥΠ, ας αναλογιστούμε την ερμηνεία της $E(y|X)$. Όταν η y είναι μια διχοτομική μεταβλητή, η προσδοκώμενη τιμή της y είναι η πιθανότητα το γεγονός να συμβεί:

$$E(y_i) = [1 \times \Pr(y_i = 1)] + [0 \times \Pr(y_i = 0)] = \Pr(y_i = 1) \quad (3)$$

Και για το υπόδειγμα παλινδρόμησης παίρνουμε δεσμευμένες πιθανότητες:

$$E(y_i|x_i) = [1 \times \Pr(y_i = 1|x_i)] + [0 \times \Pr(y_i = 0|x_i)] = \Pr(y_i = 1|x_i) \quad (4)$$

Εικόνα 1: Το γραμμικό Υπόδειγμα Πιθανότητας



Πηγή:
(Long,
1997)

Ως εκ τούτου, η προσδοκώμενη τιμή της y δεδομένου της X είναι η πιθανότητα να συμβεί το ενδεχόμενο $y=1$ δεδομένου της X . Οπότε μπορούμε να ξαναγράψουμε το ΓΥΠ ως:

$$\Pr(y_i = 1|X_i) = a + \beta X_i \quad (5)$$

Άρα έχουμε: $\frac{\partial Pr}{\partial X} = \beta$

Δηλαδή, για μια μεταβολή της X κατά μία μονάδα, το β εκφράζει την αλλαγή στην πιθανότητα να συμβεί το γεγονός, διατηρώντας τις υπόλοιπες μεταβλητές σταθερές. Εφόσον, πρόκειται για ένα γραμμικό υπόδειγμα μια αλλαγή στην X κατά μία μονάδα έχει σταθερή επίδραση στην πιθανότητα (Long, 1997).

1.2 Προβλήματα με το Γραμμικό Υπόδειγμα Πιθανότητας

Ενώ η ερμηνεία των συντελεστών δεν διαφέρει από την συνήθη ερμηνεία όταν χρησιμοποιούμε τη μέθοδο των ελαχίστων τετραγώνων για μια διχοτομική εξαρτημένη μεταβλητή, αρκετές υπόθεσες του γραμμικού υποδείγματος αναγκαστικά παραβιάζονται.

A. Ετεροσκεδαστικότητα: Αν μια τυχαία διχοτομική μεταβλητή έχει μέσο μ , τότε η διακύμανση της είναι $\mu(1-\mu)$. Εφόσον, η προσδοκώμενη τιμή της y δεδομένου των X είναι $X\beta$, η δεσμευμένη διακύμανση της y δεδομένων των X δίνεται ως:

$$Var(y|X) = \Pr(y = 1|X)[1 - \Pr(y = 1|X)] = X\beta(1 - X\beta) \quad (6)$$

Η Εξίσωση 6 υποδηλώνει ότι η διακύμανση των σφαλμάτων εξαρτάται από τα X και δεν είναι σταθερή. Εφόσον, το ΓΥΠ είναι ετεροσκεδαστικό, οι εκτιμητές OLS των β είναι μη-αποτελεσματικοί και τα τυπικά σφάλματα είναι μεροληπτικά, και οδηγούν σε λανθασμένα στατιστικά τεστ. Παρ' όλα αυτά, το πρόβλημα μπορεί να διορθωθεί χρησιμοποιώντας την διόρθωση των Huber (1967) και White (1980,1982) για τον εκτιμητή της διακύμανσης, γνωστός και ως εκτιμητής sandwich.

B. Κανονικότητα: Ας σκεφτούμε μια συγκεκριμένη τιμή της X , έστω X^* . Στο σχήμα 1.1, $E(y|X^*)$ απεικονίζεται από ένα σημείο στην γραμμή παλινδρόμησης. Το τυπικό σφάλμα ϵ αφορά στην απόσταση της παρατηρούμενης τιμής από την $E(y|X)$. Εφόσον η y μπορεί να λάβει μόνο τις τιμές 0 και 1, τα σφάλματα μπορούν να ισούνται είτε με $\epsilon_1=1 - E(y|X^*)$ είτε με $\epsilon_0=0 - E(y|X^*)$. Προφανώς, από αυτό φαίνεται ότι τα σφάλματα δεν μπορούν να κατανέμονται κανονικά

Γ. Μη-λογικές προβλέψεις: Το Γραμμικό Υπόδειγμα Πιθανότητας προβλέπει τιμές της y που είτε είναι αρνητικές ή ξεπερνούν την μονάδα. Δεδομένου της ερμηνείας της $E(y|X)$ ως $Pr(y=1|X)$, αυτό οδηγεί σε μη λογικές προβλέψεις των πιθανοτήτων. Αυτό σημαίνει ότι η γραμμική παλινδρόμηση θα μπορούσε να οδηγήσει κάποιον να προβλέψει ότι η πιθανότητα ενός ατόμου να διαπράξει ένα έγκλημα στη διάρκεια της ζωής του είναι 1.2 ή -0.4.

Δ) Συναρτησιακή μορφή: Εφόσον το υπόδειγμα είναι γραμμικό, μια μοναδιαία αύξηση στην X οδηγεί σε μια σταθερή αλλαγή β στην πιθανότητα ενός γεγονότος, διατηρώντας τις υπόλοιπες μεταβλητές σταθερές. Η αύξηση της πιθανότητας είναι η ίδια, ασχέτως της τιμής της X . Σε πολλές εφαρμογές κάτι τέτοιο δεν είναι ρεαλιστικό. Για παράδειγμα, με το ΓΥΠ κάθε αύξηση της μεταβλητής X κατά μία μονάδα μειώνει την πιθανότητα κατά 0.3 κατά αναλογία μια αύξηση της μεταβλητής X κατά 4 μονάδες θα πρέπει να μειώνει την πιθανότητα 1.2, πράγμα το οποίο φυσικά δεν έχει νόημα καθώς οι πιθανότητες ορίζονται στο διάστημα 0-1. Πιο ρεαλιστικά, κάθε επιπλέον μονάδα αύξησης της X θα πρέπει να έχει μια μειούμενη επίδραση στην πιθανότητα. Ενώ λοιπόν, η πρώτη μονάδα θα μειώνει την πιθανότητα κατά 0.3, η δεύτερη μονάδα μπορεί να την μειώνει μόνο κατά 0.2 κλπ. Αυτό υποδεικνύει ότι το υπόδειγμα πρέπει να είναι μη-γραμμικό ως προς την επίδραση των μεταβλητών στην πιθανότητα. Γενικά, όταν η ερμηνευόμενη μεταβλητή είναι πιθανότητα, είναι συχνά λογικό ότι οι επιδράσεις των ανεξάρτητων μεταβλητών θα βαίνουν μειούμενες καθώς η πιθανότητα μεταβαίνει από 0 σε 1 και αντίστροφα.

Ένα διχοτομικό υπόδειγμα θα πρέπει να έχει μια σχέση τύπου-S μεταξύ των ανεξάρτητων μεταβλητών και της πιθανότητας πραγματοποίησης ενός συμβάντος, κάτι που υπογραμμίζει το πρόβλημα συναρτησιακής μορφής των ΓΥΠ (Long, 1997).

1.3 Το υπόδειγμα της λανθάνουσας μεταβλητής για διχοτομικές εξαρτημένες μεταβλητές

Όπως και στο γραμμικό υπόδειγμα πιθανότητας, έχουμε μια παρατηρούμενη διχοτομική μεταβλητή y . Υποθέτουμε πως υπάρχει μια μη παρατηρούμενη μεταβλητή y^* που λαμβάνει τιμές από $-\infty$ έως $+\infty$ η οποία δημιουργεί την παρατηρούμενη y . Οι παρατηρήσεις εκείνες που έχουν τιμή μεγαλύτερη από y^* παρατηρούνται ως $y=1$, ενώ οι παρατηρήσεις εκείνες που έχουν τιμή μικρότερη από y^* παρατηρούνται ως $y=0$.

Είναι σημαντικό να καταλάβουμε την έννοια της λανθάνουσας μεταβλητής. Ας σκεφτούμε το εάν διαβάσει ή όχι κάποιος τις διατροφικές πληροφορίες των τροφίμων που βρίσκονται στην ετικέτα ως την παρατηρούμενη μεταβλητή y . Η μεταβλητή y μπορεί να παρατηρηθεί μόνο σε δύο περιπτώσεις: στην περίπτωση όπου ένα άτομο διαβάσει ή δεν διαβάσει τις πληροφορίες σε μία ετικέτα. Ωστόσο, δεν διαβάζουν όλοι τις πληροφορίες στην ετικέτα με την ίδια βεβαιότητα. Ένα άτομο μπορεί να είναι πολύ κοντά στο να αποφασίσει να μην διαβάσει τις πληροφορίες ενώ ένα άλλο άτομο μπορεί να είναι πολύ σταθερό στην απόφαση του να τις διαβάσει. Και στις δύο περιπτώσεις, παρατηρούμε την $y=1$. Η ιδέα της λανθάνουσας μεταβλητής y^* είναι ότι υπάρχει μια υποβόσκουσα τάση για το να διαβάσει κάποιος τις ετικέτες των τροφίμων που δημιουργεί την παρατηρούμενη κατάσταση. Ενώ, δεν μπορούμε άμεσα να παρατηρήσουμε την y^* σε κάποιο βαθμό μια αλλαγή στην y^* οδηγεί σε αλλαγή σε αυτό που παρατηρούμε.. Όταν η τάση αυτή ξεπεράσει ένα όριο (threshold) τότε είναι πιθανό αυτό να οδηγήσει στο να διαβάσει κάποιος τις αναγραφόμενες πληροφορίες στις ετικέτες τροφίμων.

Η λανθάνουσα μεταβλητή y^* υποθέτουμε πως συνδέεται γραμμικά με τις παρατηρούμενες μεταβλητές X , μέσω του μοντέλου:

$$y_i^* = \mathbf{X}_i\boldsymbol{\beta} + \varepsilon_i \quad (7)$$

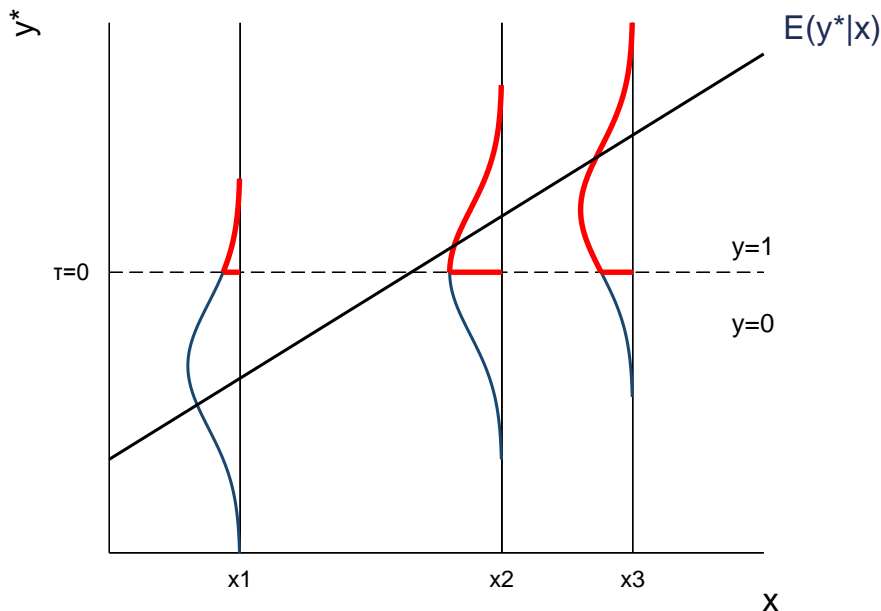
Η λανθάνουσα μεταβλητή y^* συνδέεται με την παρατηρούμενη διχοτομική μεταβλητή y μέσω του υποδείγματος:

$$y_i = \begin{cases} 1, & y_i^* > \tau \\ 0, & y_i^* \leq \tau \end{cases} \quad (8)$$

Όπου τ είναι το όριο ή το κρίσιμο σημείο. Αν $y_i^* \leq \tau$, τότε $y = 0$ ενώ αν η y_i^* ξεπεράσει το όριο ($y_i^* > \tau$) τότε $y=1$. Μια συνήθης υπόθεση είναι ότι $\tau = 0$.

Η σύνδεση μεταξύ της y^* και της παρατηρούμενης y παρουσιάζεται στην Εικόνα 2 για το υπόδειγμα $y^* = a + \beta X + \varepsilon$. Σε αυτό το σχήμα, η y^* απεικονίζεται στον κάθετο άξονα, με το όριο τ να απεικονίζεται από την οριζόντια διακεκομμένη γραμμή. Η κατανομή της y^* φαίνεται από τις συμμετρικές καμπύλες με σχήμα καμπάνας που θα πρέπει να θεωρούνται ότι αποτυπώνεται σε μια τρίτη διάσταση στο διάγραμμα. Όταν η y^* είναι μεγαλύτερη του ορίου τ παρατηρούμε την $y=1$, κάτι που υποδεικνύεται από την χρωματισμένη με κόκκινο περιοχή.

Εικόνα 2: Η κατανομή της y^* δεδομένου της X σε ένα διχοτομικό υπόδειγμα παλινδρόμησης



Πηγή: (Long, 1997)

Εφόσον η εξαρτημένη μεταβλητή δεν παρατηρείται, το υπόδειγμα δεν μπορεί να εκτιμηθεί με τη μέθοδο των ελαχίστων τετραγώνων. Αντ' αυτού χρησιμοποιούμε τη μέθοδο μέγιστης πιθανοφάνειας, που απαιτεί κάποιες υποθέσεις για την κατανομή των σφαλμάτων. Πιο συχνά, η επιλογή είναι ανάμεσα στην κανονική συνάρτηση αθροιστικής κατανομής με $var(\varepsilon/X)=1$ που οδηγεί στο υπόδειγμα probit και τη λογιστική συνάρτηση αθροιστικής κατανομής με $var(\varepsilon/X)=\pi^2/3$ που οδηγεί στο υπόδειγμα logit. Όπως και με τα γραμμικά μοντέλα παλινδρόμησης, υποθέτουμε ότι $E(\varepsilon/\mathbf{X})=0$.

Η ακριβής τιμή που υποθέτουμε για τη διακύμανση του όρου σφάλματος είναι αυθαίρετη αφού δεν μπορεί να επιβεβαιωθεί από τα δεδομένα. Οι συναρτήσεις πυκνότητας πιθανότητας και αθροιστικής κατανομής για την κανονική και την λογιστική κατανομή φαίνονται στο Σχήμα 1.3. Η κανονική κατανομή απεικονίζεται με την συνεχόμενη γραμμή. Όταν ο όρος σφάλματος κατανέμεται κανονικά με $E(\varepsilon/\mathbf{X})=0$ και $Var(\varepsilon/\mathbf{X})=1$ η συνάρτηση πυκνότητας πιθανότητας είναι:

$$\varphi(\varepsilon) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\varepsilon^2}{2}\right) \quad (9)$$

Και η συνάρτηση αθροιστικής κατανομής είναι:

$$\Phi(\varepsilon) = \int_{-\infty}^{\varepsilon} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt \quad (10)$$

Η συνάρτηση αθροιστικής κατανομής υποδηλώνει ότι η πιθανότητα μιας τυχαίας μεταβλητής είναι μικρότερη ή ίση για δεδομένη τιμή. Για παράδειγμα $\Phi(0)=Pr(\varepsilon \leq 0)=0.5$

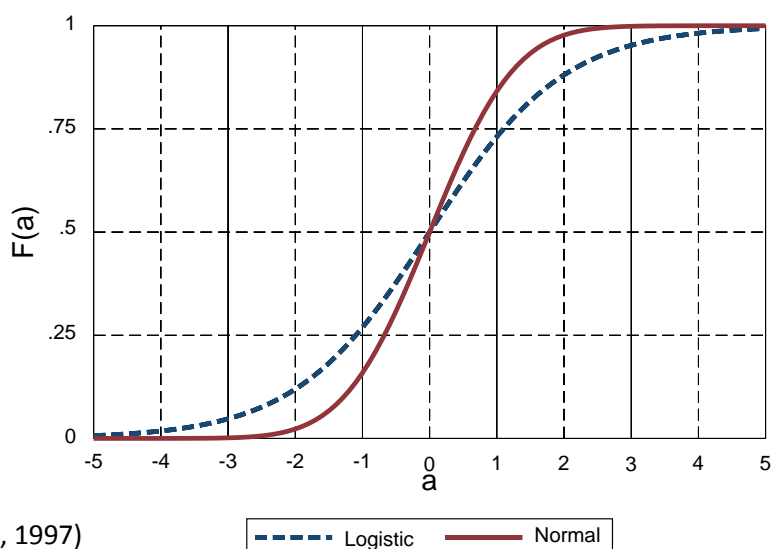
Στο υπόδειγμα logit, υποθέτουμε για τα σφάλματα ότι ακολουθούν μια τυπική λογιστική κατανομή με μέσο το 0 και διακύμανση $\pi^2/3$. Αυτή η μη συνηθισμένη διακύμανση έχει επιλεχθεί επειδή οδηγεί μια συγκεκριμένη σχέση για την συνάρτηση πυκνότητας πιθανότητας:

$$\lambda(\varepsilon) = \frac{\exp(\varepsilon)}{[1 + \exp(\varepsilon)]^2} \quad (11)$$

Και για τη συνάρτηση αθροιστικής κατανομής είναι:

$$\Lambda(\varepsilon) = \frac{\exp(\varepsilon)}{1 + \exp(\varepsilon)} \quad (12)$$

Εικόνα 3: Κανονική και λογιστική κατανομή



Πηγή: (Long, 1997)

Υποθέτοντας την κατανομή του όρου σφάλματος, είναι δυνατόν να υπολογίσουμε την πιθανότητα $Pr(y=1|\mathbf{X})$. Στην Εικόνα 2 οι τιμές για την $y=1$ φαίνονται από την κόκκινη σκιασμένη περιοχή της κατανομής του σφάλματος πάνω από το όριο τ .

Εφόσον $y=1$ όταν $y^* > 0$

$$Pr(y = 1 | \mathbf{X}) = Pr(y^* > 0 | \mathbf{X}) \quad (13)$$

Αντικαθιστώντας $y^* = \mathbf{X}\boldsymbol{\beta} + \varepsilon$, παίρνουμε:

$$Pr(y = 1 | \mathbf{X}) = Pr(\mathbf{X}\boldsymbol{\beta} + \varepsilon > 0 | \mathbf{X}) \quad (14)$$

Αφαιρώντας τον όρο $\mathbf{X}\boldsymbol{\beta}$ από κάθε πλευρά της ανισότητας;

$$Pr(y = 1 | \mathbf{X}) = Pr(\varepsilon > -\mathbf{X}\boldsymbol{\beta} | \mathbf{X}) \quad (15)$$

Εφόσον η κανονική και λογιστική κατανομή είναι και οι δύο συμμετρικές έχουμε:

$$Pr(y = 1 | \mathbf{X}) = Pr(\varepsilon \leq \mathbf{X}\boldsymbol{\beta} | \mathbf{X}) \quad (16)$$

Αυτή η Εξίσωση 16 αντιπροσωπεύει ουσιαστικά η συνάρτηση αθροιστικής κατανομής του όρου σφάλματος, οπότε:

$$Pr(y = 1 | \mathbf{X}) = F(\mathbf{X}\boldsymbol{\beta}) \quad (17)$$

Όπου η F είναι η κανονική συνάρτηση αθροιστική κατανομή Φ για το υπόδειγμα probit και η λογιστική συνάρτηση αθροιστικής κατανομής Λ για το υπόδειγμα logit (Long, 1997).

2. Συνάρτηση Πιθανοφάνειας

2.1 Η έννοια της συνάρτησης πιθανοφάνειας

Η γενική μέθοδος με την οποία εκτιμούμε τις παραμέτρους της παλινδρόμησης σε μη-γραμμικά υποδείγματα ονομάζεται *μέγιστη πιθανοφάνεια*. Η μέθοδος της μέγιστης πιθανοφάνειας αποδίδει τιμές στις άγνωστες παραμέτρους έτσι ώστε να μεγιστοποιείται η πιθανότητα να προκύψουν τα παρατηρούμενα δεδομένα. Προκειμένου να εφαρμόσουμε αυτή την μέθοδο, πρέπει πρώτα να δημιουργήσουμε μια συνάρτηση που ονομάζεται *συνάρτηση πιθανοφάνειας*. Η συνάρτηση αυτή εκφράζει την πιθανότητα να παρατηρήσουμε τα δεδομένα ως συνάρτηση των άγνωστων παραμέτρων. Οι *εκτιμητές μέγιστης πιθανοφάνειας* αυτών των παραμέτρων επιλέγονται έτσι ώστε να μεγιστοποιούν αυτή τη συνάρτηση (Collett, 2003).

2.2 Μέθοδος της Μεγίστης Πιθανοφάνειας

Σύμφωνα με τον Greene (2002) η μέθοδος αυτή δίνει την εκτίμηση που έχει τη μέγιστη πιθανοφάνεια, δηλαδή δίνει την τιμή της παραμέτρου η οποία, μεταξύ όλων των δυνατών τιμών της παραμέτρου, είναι η πιο πιθανή με βάση το δείγμα. Η συνάρτηση πυκνότητας πιθανότητας για κάποια τιμή x που εξαρτάται από κάποια παράμετρο θ παίρνει την τιμή $f(X_i; \theta)$. Επειδή $\{X_1, \dots, X_n\}$ είναι ανεξάρτητες, η πιθανότητα να τις παρατηρήσουμε σε ένα τυχαίο δείγμα μεγέθους n δίνεται από τη συνάρτηση πιθανοφάνειας (**likelihood function**) ως προς θ :

$$L(X_1, \dots, X_n; \theta) = f(X_1; \theta) \cdots f(X_n; \theta) \quad (18)$$

Στα προβλήματα εκτίμησης, θεωρούμε τα $\{X_1, \dots, X_n\}$ δεδομένα και ενδιαφερόμαστε για τη θ . Αν λοιπόν $L(X_1, \dots, X_n; \theta_1) > L(X_1, \dots, X_n; \theta_2)$ για δύο τιμές θ_1 και θ_2 της θ , τότε η τιμή θ_1 είναι πιο αληθοφανής από τη θ_2 γιατί δίνει μεγαλύτερη πιθανότητα να παρατηρήσουμε τα $\{X_1, \dots, X_n\}$. Θέλουμε λοιπόν να βρούμε την «πιο αληθοφανή» τιμή της θ , δηλαδή την τιμή $\hat{\theta}$ που μεγιστοποιεί τη $L(X_1, \dots, X_n; \theta)$ ή καλύτερα (για ευκολότερους υπολογισμούς) τη $\log L(X_1, \dots, X_n; \theta)$. Άρα η εκτιμήτρια μέγιστης πιθανοφάνειας (maximum likelihood estimator) $\hat{\theta}$ βρίσκεται από τη σχέση:

$$\frac{\partial \log L(X_1, \dots, X_n; \theta)}{\partial \theta} = 0. \quad (19)$$

Αν θέλουμε να εκτιμήσουμε δύο ή περισσότερες παραμέτρους $\theta_1, \dots, \theta_m$, η συνάρτηση πιθανοφάνειας είναι $L(X_1, \dots, X_n; \theta_1, \dots, \theta_m)$ και οι εκτιμήτριες $\hat{\theta}_1, \dots, \hat{\theta}_m$, βρίσκονται λύνοντας το σύστημα των m εξισώσεων:

$$\frac{\partial \log L(X_1, \dots, X_n; \theta_1, \dots, \theta_m)}{\partial \theta_j} = 0 \text{ για } j = 1, \dots, m. \quad (20)$$

2.3 Η συνάρτηση πιθανοφάνειας για το διχοτομικό υπόδειγμα πιθανότητας

Για να κάνουμε πιο συγκεκριμένη της συνάρτηση πιθανοφάνειας, ας ορίσουμε σύμφωνα με τον Long (1997) ως p την πιθανότητα παρατήρησης της τιμής της y που παρατηρείται στην πραγματικότητα:

$$p_i = \begin{cases} \Pr(y_i = 1|\mathbf{X}_i) & \text{αν παρατηρείται } y_i = 1 \\ 1 - \Pr(y_i = 1|\mathbf{X}_i) & \text{αν παρατηρείται } y_i = 0 \end{cases} \quad (21)$$

Αν οι παρατηρήσεις είναι ανεξάρτητες, η συνάρτηση πιθανοφάνειας δίνεται από:

$$L(\boldsymbol{\beta}|\mathbf{y}, \mathbf{X}) = \prod_{i=1}^N p_i \quad (22)$$

Συνδυάζοντας τις εξισώσεις (21) και (22) προκύπτει:

$$L(\boldsymbol{\beta}|\mathbf{y}, \mathbf{X}) = \prod_{y=1} \Pr(y_i = 1|\mathbf{X}_i) * \prod_{y=0} [1 - \Pr(y_i = 1|\mathbf{X}_i)] \quad (23)$$

Όπου ο δείκτης του πολλαπλασιασμού υποδηλώνει το αποτέλεσμα λαμβάνεται μόνο στις περιπτώσεις όπου $y=1$ και $y=0$, αντίστοιχα.

Τα β 's ενσωματώνονται στην εξίσωση πιθανοφάνειας αντικαθιστώντας το δεξί μέρος της εξίσωσης 17 παίρνουμε:

$$L(\boldsymbol{\beta}|\mathbf{y}, \mathbf{X}) = \prod_{y=1} F(\mathbf{X}_i\boldsymbol{\beta}) * \prod_{y=0} [1 - F(\mathbf{X}_i\boldsymbol{\beta})] \quad (24)$$

Λογαριθμίζοντας και στις 2 πλευρές, καταλήγουμε στην εξίσωση του λογάριθμου της πιθανοφάνειας:

$$\ln L(\boldsymbol{\beta}|\mathbf{y}, \mathbf{X}) = \sum_{y=1} \ln F(\mathbf{X}_i\boldsymbol{\beta}) + \sum_{y=0} \ln [1 - F(\mathbf{X}_i\boldsymbol{\beta})] \quad (25)$$

2.4 Ερμηνεία: η επίδραση των παραμέτρων

Στην περίπτωση ενός διχοτομικού υποδείγματος παλινδρόμησης με μία μόνο ανεξάρτητη μεταβλητή X :

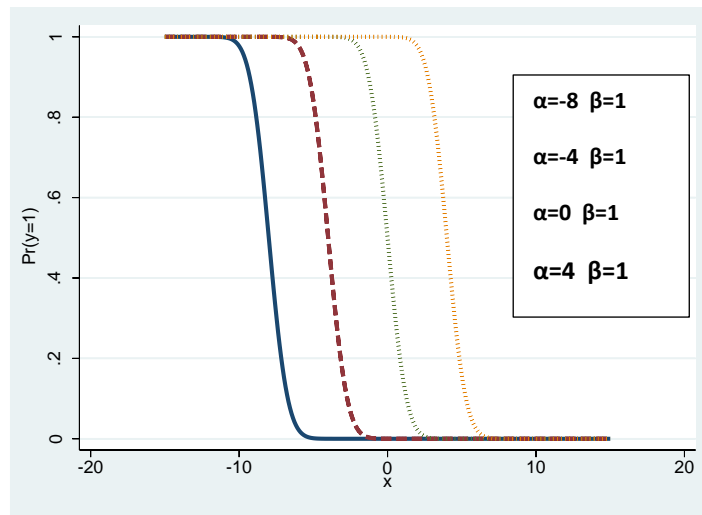
$$\Pr(y = 1|X) = F(\alpha + \beta X) \quad (26)$$

Η Εικόνα 4.A δείχνει την επίδραση του σταθερού όρου στην καμπύλη πιθανότητας. Όταν το $\alpha=0$, που φαίνεται στην ελαφριά διακεκομμένη γραμμή, η καμπύλη περνάει από το σημείο (0, 0.5). Καθώς το α μεγαλώνει, η καμπύλη μετατοπίζεται προς τα αριστερά ενώ; καθώς το α μικραίνει, η καμπύλη μετατοπίζεται προς τα δεξιά. Η εικόνα 4.B δείχνει τις επιδράσεις

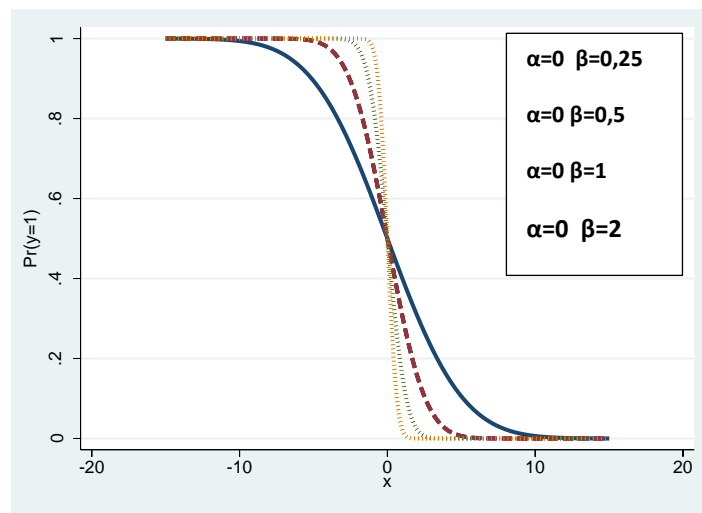
όταν αλλάζει η κλίση της καμπύλης β . Για το $\alpha=0$, οι καμπύλες περνάνε από το σημείο $(0, 0.50)$. Όσο πιο μικρό γίνεται το β , τόσο πιο πολύ τεντώνεται προς τα άκρα η καμπύλη. Γενικά, όσο το β αυξάνεται, η καμπύλη αυξάνεται πιο απότομα καθώς το X τείνει στο 0.

Εικόνα 4: Οι επιδράσεις της αλλαγής της κλίσης και του σταθερού όρου σε ένα απλό διχοτομικό υπόδειγμα παλινδρόμησης

4.A: Επιδράσεις της αλλαγής του σταθερού όρου



4.B: Επιδράσεις της αλλαγής της κλίσης



Πηγή: (Long, 1997)

Εφόσον εκτιμηθεί το υπόδειγμα logit ή probit, με τη μέθοδο της μέγιστης πιθανοφάνειας, ο έλεγχος υποθέσεων και η οικονομετρική ανάλυση μπορεί να γίνει όπως την γνωρίζουμε. Τα πρόσημα των συντελεστών έχουν την ίδια ερμηνεία όπως και στο γραμμικό υπόδειγμα πιθανότητας, και ο έλεγχος με τη στατιστική t μπορεί να χρησιμοποιηθεί για να ελέγξουμε

υποθέσεις σχετικά με τους συντελεστές του logit ή probit υποδείγματος. Ωστόσο, η οικονομική ερμηνεία των εκτιμημένων συντελεστών σε αυτά τα υποδείγματα διαφέρει από την ερμηνεία που δώσαμε στο γραμμικό υπόδειγμα πιθανότητας. Συγκεκριμένα, τα απόλυτα μεγέθη, των εκτιμώμενων συντελεστών μη-γραμμικών υποδειγμάτων είναι πολύ διαφορετικά από τα μεγέθη των εκτιμημένων συντελεστών του γραμμικού υποδείγματος πιθανότητας για την ίδια εξειδίκευση και τα ίδια δεδομένα. Όπως μπορούμε να παρατηρήσουμε και στην εικόνα 4.B, οι οριακές επιδράσεις δεν είναι σταθερές εφόσον η κλίση (η μεταβολή δηλαδή της πιθανότητας) της γραφικής παράστασης του υποδείγματος probit μεταβάλλεται, καθώς το y κινείται από το 0 προς το 1. Οι εκτιμώμενοι συντελεστές στα υποδείγματα logit και probit έχουν μόνο ποιοτική ερμηνεία και όχι ποσοτική. Γι' αυτό πολλοί ερευνητές χρησιμοποιούν για την ερμηνεία των αποτελεσμάτων την έννοια των οριακών επιδράσεων και των διακριτών αλλαγών που συζητείται παρακάτω.

2.5 Το τεστ του λόγου της πιθανοφάνειας (Likelihood Ratio Test- LR test)

Το τεστ του λόγου της πιθανοφάνειας (LR test) χρησιμοποιείται για να ελέγξουμε πιθανούς περιορισμούς στο υπόδειγμα. Αυτοί οι περιορισμοί συνήθως είναι μη-γραμμικοί και περίπλοκοι, αλλά για λόγους απλότητας θα αναφερθούμε μόνο στην περίπτωση που ο περιορισμός αφορά στην απαλοιφή μιας ερμηνευτικής μεταβλητής από το υπόδειγμα μας. Για παράδειγμα, θα μπορούσαμε να έχουμε τα παρακάτω λογιστικά υποδείγματα:

$$\begin{aligned} M_1: \Pr(y = 1|x) &= \Lambda(\beta_0 + \beta_1 X_1 + \beta_2 X_2) \\ M_2: \Pr(y = 1|x) &= \Lambda(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3) \\ M_3: \Pr(y = 1|x) &= \Lambda(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4) \end{aligned} \quad (27)$$

Το υπόδειγμα M_1 προέκυψε από το υπόδειγμα M_2 εφαρμόζοντας τον περιορισμό $\beta_3 = 0$, το υπόδειγμα M_1 προέκυψε από το υπόδειγμα M_3 εφαρμόζοντας τον περιορισμό $\beta_4 = \beta_3 = 0$ και το υπόδειγμα M_2 προέκυψε από το υπόδειγμα M_3 εφαρμόζοντας τον περιορισμό $\beta_4 = 0$. Όταν ένα υπόδειγμα προκύπτει από ένα άλλο υπόδειγμα εφαρμόζοντας στο ένα από αυτά περιορισμούς, το περιορισμένο υπόδειγμα λέγεται «ένθετο» στο μη περιορισμένο υπόδειγμα. Έτσι, το υπόδειγμα M_1 είναι ένθετο στο M_2 και στο M_3 . Επίσης, το M_2 είναι ένθετο στο M_3 . Το τεστ του λόγου πιθανοφάνειας έχει ως εξής. Το περιορισμένο μοντέλο M_c με παραμέτρους β_c είναι ένθετο στο υπόδειγμα χωρίς περιορισμούς M_u με παραμέτρους β_u . Η μηδενική υπόθεση είναι ότι οι περιορισμοί που εφαρμόζονται στο περιορισμένο υπόδειγμα, είναι αληθείς. Έστω ότι $L(M_u)$ είναι η τιμή της συνάρτησης πιθανοφάνειας που εκτιμάται για το υπόδειγμα χωρίς περιορισμούς, και ότι $L(M_c)$ είναι η τιμή της συνάρτησης πιθανοφάνειας στο υπόδειγμα με περιορισμούς. Η στατιστική του λόγου πιθανοφάνειας δίνεται από:

$$G^2(M_c | M_u) = 2 \ln L(M_u) - 2 \ln L(M_c) = \ln \left(\frac{L(M_u)}{L(M_c)} \right)^2 \quad (28)$$

H_0 : Οι περιορισμοί που επιβάλλονται στο υπόδειγμα χωρίς περιορισμούς (M_u) για να προκύψει το υπόδειγμα με περιορισμούς, είναι αληθείς (M_c).

Για να μπορεί να πραγματοποιηθεί ο έλεγχος λόγου της πιθανοφάνειας όμως πρέπει να ισχύουν οι παρακάτω προϋποθέσεις:

- A. Θα πρέπει τα δύο υποδείγματα να είναι το ένα ένθετο στο άλλο. Δηλαδή, ο έλεγχος ισχύει μόνο όταν το ένα υπόδειγμα μπορεί να προέλθει από το άλλο με την επιβολή περιορισμών στους συντελεστές του υποδείγματος χωρίς περιορισμό.
- B. Θα πρέπει τα δύο υποδείγματα να αφορούν το ίδιο δείγμα. Δηλαδή, δεν είναι έγκυρος ο έλεγχος εάν το ένα υπόδειγμα έχει περισσότερες παρατηρήσεις από το άλλο.
- Γ. Στην περίπτωση που έχουμε ένα υπόδειγμα που περιλαμβάνει κατηγορικές μεταβλητές θα πρέπει να υπάρχουν παρατηρήσεις για κάθε κατηγορία των κατηγορικών μεταβλητών.

3. Οριακές Επιδράσεις (Marginal Effects) και Όροι Αλληλεπίδρασης (Interaction Effects)

3.1 Οριακές Επιδράσεις (Marginal Effects)

Στα μη-γραμμικά υποδείγματα πολλές φορές θέλουμε να υπολογίσουμε πως μεταβάλλεται η προβλεπόμενη πιθανότητα της y καθώς μεταβάλλεται μια ανεξάρτητη μεταβλητή ενώ οι υπόλοιπες μεταβλητές παραμένουν σταθερές. Στο γραμμικό υπόδειγμα πιθανότητας η μεταβολή αυτή δίνεται κατευθείαν από τους συντελεστές της παλινδρόμησης. Στην περίπτωση όμως των μη-γραμμικών παλινδρομήσεων κάτι τέτοιο δεν ισχύει. Σε αυτά τα υποδείγματα τα β_i μπορούν να χρησιμοποιηθούν για να υπολογιστεί η μερική αλλαγή στην πιθανότητα εμφάνισης ενός χαρακτηριστικού. Έστω η παρακάτω πιθανότητα ενός μη γραμμικού υποδείγματος:

$$Pr(y = 1|X) = F(\mathbf{X}\boldsymbol{\beta}) \quad (29)$$

Όπου F είναι η συνάρτηση αθροιστικής κατανομής είτε για την κανονική κατανομή (Φ) είτε για την λογιστική κατανομή (Λ).

Η μερική αλλαγή στην πιθανότητα ονομάζεται οριακή επίδραση και υπολογίζεται παίρνοντας την μερική παράγωγο της παραπάνω εξίσωσης ως προς μια μεταβλητή X_k .

$$\frac{\partial Pr(y = 1 | X)}{\partial X_k} = \frac{\partial F(\mathbf{X}\boldsymbol{\beta})}{\partial X_k} = \frac{\partial F(\mathbf{X}\boldsymbol{\beta})}{\partial \mathbf{X}\boldsymbol{\beta}} * \frac{\partial \mathbf{X}\boldsymbol{\beta}}{\partial X_k} = f(\mathbf{X}\boldsymbol{\beta})\beta_k \quad (30)$$

Όπου η f είναι η συνάρτηση πυκνότητας πιθανότητας

Η οριακή επίδραση είναι η κλίση της συνάρτησης αθροιστικής κατανομής ως προς την X , διατηρώντας τις υπόλοιπες μεταβλητές σταθερές. Το πρόσημο της οριακής επίδρασης καθορίζεται από το πρόσημο του συντελεστή β_k , αφού ο όρος $f(\mathbf{X}\boldsymbol{\beta})$ είναι πάντα θετικός. Το μέγεθος της αλλαγής εξαρτάται τόσο από το μέγεθος β_k όσο και από την ποσότητα $X\boldsymbol{\beta}$. Αυτό σημαίνει πως το μέγεθος της οριακής επίδρασης εξαρτάται από τις τιμές όλων των υπόλοιπων ανεξάρτητων μεταβλητών αφού **η f υπολογίζεται βάσει των $\mathbf{X}\boldsymbol{\beta}$** . Εφόσον η τιμή της οριακής επίδρασης εξαρτάται από τα επίπεδα τιμών των άλλων ανεξάρτητων μεταβλητών πρέπει να αποφασίσουμε σε ποιες τιμές των μεταβλητών θα υπολογίσουμε την οριακή επίδραση.

Μια μέθοδος είναι να υπολογίσουμε την μέση οριακή επίδραση αθροίζοντας τις οριακές επιδράσεις όλων των παρατηρήσεων και διαιρώντας με το πλήθος αυτών:

$$\frac{\partial \Pr(y = 1 | X)}{\partial X_{\kappa}} = \frac{\sum_{i=1}^N f(X\beta)\beta_{\kappa}}{N} \quad (31)$$

Μια άλλη μέθοδος είναι να υπολογίσουμε την οριακή επίδραση στον μέσο όλων των ανεξάρτητων μεταβλητών:

$$\frac{\partial \Pr(y = 1 | \bar{X})}{\partial X_{\kappa}} = f(\bar{X}\beta)\beta_{\kappa} \quad (32)$$

Οι οριακές επιδράσεις χρησιμοποιούνται κυρίως για συνεχείς μεταβλητές ώστε να υπολογίσουμε πόσο αλλάζει η πιθανότητα εμφάνισης ενός χαρακτηριστικού για μια οριακή αλλαγή στην μεταβλητή X_{κ} (Long, 1997)

3.2 Διακριτή Αλλαγή (Discrete Change)

Η αλλαγή της προβλεπόμενης πιθανότητας για μια διακριτή αλλαγή της ανεξάρτητης μεταβλητής είναι μια εναλλακτική της οριακής επίδρασης που χρησιμοποιείται κυρίως στην περίπτωση των ψευδομεταβλητών. Έστω $\Pr(y = 1 | \mathbf{X}, X_{\kappa})$ η πιθανότητα εμφάνισης ενός χαρακτηριστικού δεδομένου των \mathbf{X} για την μεταβλητή X_{κ} . Έτσι, η $\Pr(y = 1 | \mathbf{X}, X_{\kappa} + \delta)$ είναι η πιθανότητα του X_{κ} καθώς αυτό αυξάνεται κατά δ , ενώ οι υπόλοιπες μεταβλητές παραμένουν σταθερές (Driouchotis, 2011).

Η διακριτή αλλαγή στην πιθανότητα για μια αλλαγή της X_{κ} κατά δ δίνεται:

$$\frac{\Delta \Pr(y = 1 | \mathbf{X})}{\Delta X_{\kappa}} = \Pr(y = 1 | \mathbf{X}, X_{\kappa} + \delta) - \Pr(y = 1 | \mathbf{X}, X_{\kappa}) \quad (33)$$

Έτσι στην περίπτωση μια ψευδομεταβλητής θα παίρναμε:

$$\frac{\Delta \Pr(y = 1 | \mathbf{X})}{\Delta X_{\kappa}} = \Pr(y = 1 | \mathbf{X}, X_{\kappa} = 1) - \Pr(y = 1 | \mathbf{X}, X_{\kappa} = 0) \quad (34)$$

Είναι σημαντικό να υπογραμμίσουμε ότι μια οριακή επίδραση δεν είναι ισοδύναμη της διακριτής αλλαγής, δηλαδή:

$$\frac{\partial \Pr(y = 1 | \mathbf{X})}{\partial X_k} \neq \frac{\Delta \Pr(y = 1 | \mathbf{X})}{\Delta X_k} \quad (35)$$

3.3 Όροι αλληλεπίδρασης (Interaction Effects)

Πολύ συχνά στα εφαρμοσμένα οικονομικά χρησιμοποιούνται οι όροι αλληλεπίδρασης προκειμένου να εκτιμηθεί πως η επίδραση μιας ανεξάρτητης μεταβλητής επί της εξαρτημένης εξαρτάται από το μέγεθος μιας άλλης ανεξάρτητης μεταβλητής.

Στα γραμμικά υποδείγματα η ερμηνεία του συντελεστή αλληλεπίδρασης μεταξύ δύο μεταβλητών είναι άμεση. Έστω ότι έχουμε το παρακάτω υπόδειγμα:

$$y_i = b_0 + b_1 X_{1i} + b_2 X_{2i} + b_{12} X_{1i} X_{2i} + u_i \quad (36)$$

Αν οι X_1 και X_2 είναι συνεχείς, ο συντελεστής του όρου αλληλεπίδρασης των δύο ανεξάρτητων μεταβλητών δίνεται από την σταυροειδή μερική παράγωγο της προσδοκώμενης τιμής της Y :

$$\frac{\partial E[y | X_{1i}, X_{2i}, \mathbf{X}]}{\partial X_1} = b_1 + b_{12} X_2 \quad (37)$$

ή

$$\frac{\partial^2 E[y | X_{1i}, X_{2i}, \mathbf{X}]}{\partial X_1 \partial X_2} = b_{12} \quad (38)$$

Ενώ αν οι X_1 και X_2 είναι ψευδομεταβλητές τότε η επίδραση αλληλεπίδρασης μιας αλλαγής και στην X_1 και στην X_2 από τιμή =0 σε τιμή =1, λαμβάνεται από την διακριτή διαφορά:

$$\frac{\Delta^2 E[y | X_{1i}, X_{2i}, \mathbf{X}]}{\Delta X_1 \Delta X_2} = b_{12} \quad (39)$$

Η στατιστική σημαντικότητα του όρου αλληλεπίδρασης μπορεί σε αυτή την περίπτωση να ελεγχθεί απευθείας με τον έλεγχο με την στατιστική t του συντελεστή b_{12} .

Έστω ότι έχουμε το ίδιο υπόδειγμα με πριν μόνο που η y είναι μια διχοτομική εξαρτημένη μεταβλητή:

$$E[y | X_{1i}, X_{2i}, \mathbf{X}] = F(b_1 X_{1i} + b_2 X_{2i} + b_{12} X_{1i} X_{2i} + Xb) = F(\cdot) \quad (40)$$

Έστω ότι οι X_1 και X_2 είναι συνεχείς μεταβλητές. Η επίδραση της X_1 επί της πιθανότητας είναι:

$$\frac{\partial F(\cdot)}{\partial X_1} = (b_1 + b_{12}X_{2i})F'(\cdot) \quad (41)$$

ή

$$\frac{\partial^2 F(\cdot)}{\partial X_1 \partial X_2} = b_{12}F'(\cdot) + (b_1 + b_{12}X_{2i})(b_2 + b_{12}X_{1i})F''(\cdot) \quad (42)$$

Θα μπορούσαμε αντί αυτού να υπολογίσουμε την οριακή επίδραση του όρου αλληλεπίδρασης, κάτι τέτοιο όμως θα αναλύσουμε παρακάτω (Ai & Norton, 2003).

3.4 Οριακές επιδράσεις των όρων αλληλεπίδρασης

Στο σημείο αυτό θα συγκρίνουμε τις οριακές επιδράσεις των όρων αλληλεπίδρασης σε γραμμικά και μη-γραμμικά υποδείγματα σύμφωνα με τον Driscoutis (2011).

3.4.1 OLS υποδείγματα παλινδρόμησης

Έστω ότι έχουμε το παρακάτω υπόδειγμα παλινδρόμησης:

$$y_i = b_1X_{1i} + b_2X_{2i} + b_{12}X_{1i}X_{2i} + \mathbf{b}'\mathbf{X} + u_i \quad (43)$$

Ο σταθερός όρος b_0 περιλαμβάνεται στο διάνυσμα \mathbf{b}' . Η οριακή επίδραση όποιας μεταβλητής X_j αλληλεπιδρά μπορεί να γραφτεί ως:

$$\frac{\partial y_i}{\partial X_{ji}} = b_j + b_{12}X_{ki} \quad \text{για } k = 1,2 \text{ όταν } j = 2,1 \quad (44)$$

Η Εξίσωση (44) μπορεί να εκτιμηθεί για διάφορες τιμές της μεταβλητής X_k . Αν τώρα η X_k είναι ψευδομεταβλητή η οριακή επίδραση μπορεί να υπολογιστεί για τις τιμές 0 και 1 αντίστοιχα (και στο μέσο όλων των υπόλοιπων μεταβλητών):

$$\frac{\partial y_i}{\partial X_{ji}} \Big|_{X_k=0} = b_j \quad \text{και} \quad \frac{\partial y_i}{\partial X_{ji}} \Big|_{X_k=1} = b_j + b_{12} \quad (45)$$

3.4.2 Υποδείγματα παλινδρόμησης probit και logit

Έστω η παρακάτω παλινδρόμηση λανθάνουσας μεταβλητής:

$$Y_i^* = b_1X_{1i} + b_2X_{2i} + b_{12}X_{1i}X_{2i} + \mathbf{b}'\mathbf{X} + u_i \quad (46)$$

Με παρατηρούμενη μεταβλητή την $y=1$ αν και μόνο αν $y^* > 0$. Η προσδοκώμενη μέση τιμή για την παρατηρούμενη y είναι:

$$E(y_i|\mathbf{X}_i) = 0 * \text{Prob}[Y = 0|\mathbf{X}] + 1 * \text{Prob}[Y = 1|\mathbf{X}] \quad (47)$$

Οι οριακές επιδράσεις ενός διχοτομικού μοντέλου παλινδρόμησης για μια συνεχή μεταβλητή δίνονται από:

$$ME = \frac{\partial F(\mathbf{b}'\mathbf{X})}{\partial \mathbf{X}} = bf(\mathbf{b}'\mathbf{X}) \quad (48)$$

και αν πρόκειται για ψευδομεταβλητή τότε δίνεται από:

$$\begin{aligned} DC &= \text{Prob}[Y = 1|X_j = 1] - \text{Prob}[Y = 0|X_j = 0] \\ &= F(\mathbf{b}'\mathbf{X}|X_j = 1) - F(\mathbf{b}'\mathbf{X}|X_j = 0) \end{aligned} \quad (49)$$

Σε σύγκριση με το γραμμικό υπόδειγμα παλινδρόμησης, στα μη-γραμμικά υποδείγματα έχει διαφορά αν η εκτιμώμενη αλλαγή αφορά μια συνεχή ή μια διακριτή μεταβλητή. Αν έχουμε συνεχή μεταβλητή τότε η οριακή επίδραση του όρου αλληλεπίδρασης υπολογίζεται ως:

$$ME_j = (b_j + b_{12}X_k) * f(\mathbf{b}'\mathbf{X}) \text{ για } k = 1,2 \text{ όταν } j = 2,1 \quad (50)$$

Ενώ αν πρόκειται για ψευδομεταβλητή τότε η κατάλληλη οριακή αλλαγή είναι η διακριτή αλλαγή που υπολογίζεται ως:

$$\begin{aligned} DC &= \text{Prob}[Y = 1|X_j = 1] - \text{Prob}[Y = 0|X_j = 0] \\ &= F(\mathbf{b}'\mathbf{X}|X_j = 1) - F(\mathbf{b}'\mathbf{X}|X_j = 0) \text{ όταν } j = 1,2 \end{aligned} \quad (51)$$

Υπάρχουν 3 σημαντικές επιπτώσεις των οριακών επιδράσεων ενός όρου αλληλεπίδρασης για τα μη-γραμμικά υποδείγματα με όρους αλληλεπίδρασης:

- 1) Το φαινόμενο αλληλεπίδρασης μπορεί να είναι μη-μηδενικό ακόμα και αν ο συντελεστής του όρου αλληλεπίδρασης, b_{12} είναι μηδενικός. Δηλαδή η αλληλεπίδραση σε μη-γραμμικά υποδείγματα παρατηρείται ακόμα και αν δεν υπάρχει όρος αλληλεπίδρασης μέσα στο υπόδειγμα.
- 2) Σε αντίθεση με τα γραμμικά μοντέλα η επίδραση αλληλεπίδρασης εξαρτάται απ' τις υπόλοιπες ανεξάρτητες μεταβλητές.
- 3) Το πρόσημο της επίδρασης αλληλεπίδρασης μπορεί να διαφέρει για διαφορετικές τιμές των ανεξάρτητων μεταβλητών. Οπότε το πρόσημο του b_{12} δεν υποδηλώνει αναγκαστικά το πρόσημο της επίδρασης αλληλεπίδρασης (Ai & Norton, 2003).

4. Η έννοια της διατροφικής ετικέτας

Η διατροφική επισήμανση μπορεί να οριστεί ως η ετικέτα που δίνει όλες τις απαραίτητες πληροφορίες ενός προϊόντος, στην οποία συμπεριλαμβάνονται από απλές πληροφορίες διατροφικών στοιχείων μέχρι τις ημερήσιες τιμές αναφοράς, τη σύσταση του προϊόντος, τις επιδράσεις στην υγεία και την αποποίηση ευθυνών από τυχόν παρενέργειες από την κατανάλωση του προϊόντος (π.χ. μια συσκευασία μπορεί να αναφέρει ότι περιέχει «ίχνη ξηρών καρπών» ώστε κάποιος με αλλεργία στους ξηρούς καρπούς να μην διεκδικήσει από την εταιρεία αποζημίωση) (Ghani et al., 1999).

Οι διατροφικές ετικέτες περιγράφουν την περιεκτικότητα σε θρεπτικά συστατικά των τροφίμων και προορίζονται για να καθοδηγήσουν τον καταναλωτή μέσω της παροχής πληροφοριών σε μια επιλογή η οποία θα είναι προϊόν καλής πληροφόρησης του. Οι διατροφικές πληροφορίες θα πρέπει να είναι ειλικρινείς και να μην παραπλανούν τους καταναλωτές. Ταυτόχρονα, οι κανονισμοί επισήμανσης θα πρέπει να παρέχουν κίνητρα στους κατασκευαστές να αναπτύξουν προϊόντα που προωθούν τη δημόσια υγεία και να βοηθούν τους καταναλωτές να ακολουθούν διατροφικές συστάσεις (Wijngaart., 2002).

Η διατροφική επισήμανση έχει λάβει μεγάλη προσοχή στη βιβλιογραφία λόγω του αυξημένου ενδιαφέροντος των καταναλωτών για θέματα υγείας και διατροφής. Ωστόσο, οι δυνατότητές της, δεν αξιοποιούνται πάντα με τον καλύτερο τρόπο. Υπάρχουν ενδείξεις ότι οι ετικέτες διατροφής ενδέχεται να μην χρησιμοποιούνται ακόμη και αν οι καταναλωτές λένε ότι το κάνουν, και ότι μπορεί να μην γίνονται πάντα κατανοητές από τους καταναλωτές. Οικονομικές μελέτες σχετικά με τις διατροφικές πληροφορίες έχουν διερευνήσει τους προσδιοριστικούς παράγοντες της χρήσης αυτού του είδους πληροφοριών από τους καταναλωτές και τη σχέση μεταξύ διατροφής και υγείας, αναλύοντας συγκεκριμένα τη χρήση ετικετών διατροφής για τον προσανατολισμό της συμπεριφοράς των καταναλωτών προς την υγιεινή διατροφή (Nayga, 1996). Τα διαθέσιμα στοιχεία δείχνουν ότι οι καταναλωτές που βλέπουν τις ετικέτες διατροφής μπορούν να κατανοήσουν μερικούς από τους όρους που χρησιμοποιούνται, αλλά μπερδεύονται από άλλους τύπους πληροφοριών. Οι περισσότεροι άνθρωποι φαίνονται ικανοί για ανάκτηση απλών πληροφοριών και για να κάνουν απλούς υπολογισμούς και συγκρίσεις μεταξύ των προϊόντων, αλλά η ικανότητά τους να ερμηνεύουν τη διατροφική ετικέτα μειώνεται καθώς αυξάνεται η πολυπλοκότητα της ετικέτας (Drichoutis et al., 2005).

4.1 Παράγοντες που επιδρούν στην χρήση της διατροφικής ετικέτας

Σύμφωνα με τους Drichoutis , Lazaridis , και Nayga (2005) , οι παράγοντες που επηρεάζουν τη χρήση πληροφοριών στις ετικέτες τροφίμων μπορούν να κατηγοριοποιηθούν στις ακόλουθες κατηγορίες:

- i) Ατομικά χαρακτηριστικά
- ii) Καταναλωτική συμπεριφορά και στάση
- iii) Παράγοντες εμπλοκής με το προϊόν
- iv) Παράγοντες γνώσης
- v) Παράγοντες παρακίνησης
- vi) Άλλοι παράγοντες

i. Ατομικά χαρακτηριστικά:

Ένα από τα ατομικά χαρακτηριστικά που φαίνεται να επηρεάζει την ανάγνωση των διατροφικών ετικετών από τους καταναλωτές είναι η ηλικία. Η επίδραση της ηλικίας σύμφωνα με την βιβλιογραφία έχει βρεθεί ότι επηρεάζει τη χρήση των διατροφικών ετικετών με διαφορετικούς τρόπους. Για παράδειγμα, οι Burton και Andrews (1996) διαπίστωσαν ότι οι ηλικιωμένοι αντιλαμβάνονται τις ετικέτες ως λιγότερο κατανοητές ενώ άλλοι μελετητές έχουν βρει το αντίθετο. Δηλαδή, ότι καθώς αυξάνεται η ηλικία αυξάνεται και η πιθανότητα χρήσης των πληροφοριών σχετικά με την περιεκτικότητα σε λιπαρά, περιεχόμενο χοληστερόλης και οφέλη για την υγεία (Nayga, 1996).

Ένα ακόμη ατομικό χαρακτηριστικό που επιδρά στην χρήση διατροφικών ετικετών αποτελεί το φύλο για το οποίο οι μελέτες δείχνουν αντιφατικά αποτελέσματα. Από τη μία πλευρά, ορισμένες μελέτες αναφέρουν ότι οι γυναίκες χρησιμοποιούν και εμπιστεύονται τις ετικέτες των τροφίμων περισσότερο από τους άνδρες (Drichoutis et al., 2007). Από την άλλη πλευρά, μια σειρά από μελέτες αναφέρουν ότι δεν υπάρχουν στατιστικά σημαντικές διαφορές μεταξύ των φύλων (Nayga, 2000).

Η εκπαίδευση και το εισόδημα έχουν εξεταστεί ως καθοριστικοί παράγοντες για τη χρήση των ετικετών των τροφίμων. Από τη μία πλευρά, φαίνεται ότι όσοι χρησιμοποιούν τις ετικέτες των τροφίμων έχουν υψηλότερο επίπεδο εκπαίδευσης από τους μη χρήστες. Αυτό δεν σημαίνει απαραίτητα ότι οι λιγότερο μορφωμένοι δεν εμπλέκονται καθόλου σε συμπεριφορές αναζήτησης πληροφοριών, αλλά εστιάζουν μόνο στις ετικέτες διατροφής, ενώ τα καλύτερα μορφωμένα άτομα εξετάζουν τόσο τις διατροφικές ετικέτες όσο και τις λίστες συστατικών (Bender & Derby, 1992). Όσον αφορά τον παράγοντα του εισοδήματος κάποιες μελέτες καταδεικνύουν ότι οι καταναλωτές με υψηλό εισόδημα είναι πιο πιθανό να χρησιμοποιήσουν τις ετικέτες των τροφίμων ενώ άλλες μελέτες βρίσκουν είτε το αντίθετο αποτέλεσμα ή δεν βρίσκουν καμία επίδραση (Drichoutis et al., 2007).

ii. Καταναλωτική συμπεριφορά και στάση

Σε αυτούς τους παράγοντες μπορούν να συμπεριληφθούν ο χρόνος που αφιερώνεται για ψώνια, το εάν το άτομο ακολουθεί κάποια συγκεκριμένη δίαιτα, η επίγνωση των καταναλωτών σχετικά με την υγιεινή διατροφή, το εάν αγοράζουν ή όχι οργανικά τρόφιμα, ο τύπος του νοικοκυριού, το εάν ένα άτομο είναι ο υπεύθυνος για την προετοιμασία των γευμάτων κ.α. Μελέτες δείχνουν μια θετική επίδραση της τρέχουσας κατάστασης διατροφής στην αναζήτηση πληροφοριών διατροφής. Φαίνεται ότι οι καταναλωτές με ιδιαίτερο ενδιαφέρον για τη διατροφή ή οι καταναλωτές που θεωρούν τη διατροφή ως σημαντική για τον τρόπο ζωής τους, είναι πιο πιθανό να χρησιμοποιούν τη διατροφική επισήμανση περισσότερο (Nayga, 2000). Ο τύπος του νοικοκυριού επηρεάζει επίσης τη χρήση της ετικέτας. Για παράδειγμα, τα νοικοκυριά με παιδιά προσχολικής ηλικίας και παντρεμένους καταναλωτές είναι πιο πιθανό να αναζητήσουν πληροφορίες διατροφής (McLean & Meynsse, 2001). Άτομα που αποτελούν τους βασικούς αγοραστές από παντοπωλεία βρέθηκαν πιο πιθανό να χρησιμοποιούν τις διατροφικές ετικέτες αλλά οι υπεύθυνοι σχεδιασμού γεύματος βρέθηκαν λιγότερο πιθανό να τις χρησιμοποιούν ίσως επειδή αποδίδουν μεγαλύτερη αξία στη γεύση. Επιπλέον, φαίνεται ότι ο χρόνος που

διατίθεται για ψώνια επιδρά θετικά στη χρήση των ετικετών από τους καταναλωτές (Drichoutis et al., 2007).

iii. Παράγοντες εμπλοκής με το προϊόν

Η σημασία που αποδίδουν οι καταναλωτές σε ορισμένες ιδιότητες των τροφίμων έχει βρεθεί ότι επηρεάζει τη χρήση της διατροφικής ετικέτας. Οι καταναλωτές που δίνουν σημασία στην τιμή είναι λιγότερο πιθανό να χρησιμοποιούν γενικά διατροφικές ετικέτες. Αυτό μπορεί να εξηγηθεί από το γεγονός ότι οι καταναλωτές που δίνουν μεγάλη σημασία στην τιμή του προϊόντος αναζητούν πληροφορίες σχετικά με τις τιμές, οι οποίες στη συνέχεια μπορούν να τους εμποδίσουν να εξετάσουν τις διατροφικές ετικέτες είτε για να αποφύγουν την υπερφόρτωση πληροφοριών είτε απλώς λόγω έλλειψης διαθέσιμου χρόνου για να το κάνουν. Αντίθετα, η επίδραση της σημασίας της γεύσης στη θρεπτική χρήση της ετικέτας δεν είναι ξεκάθαρη, δεδομένου ότι διάφορες μελέτες έχουν βρει αντιφατικά αποτελέσματα (Drichoutis et al., 2007).

iv. Γνωστικοί παράγοντες

Ένας παράγοντας που βρέθηκε να σχετίζεται επίσης με την χρήση των ετικετών είναι η γνώση της θρέψης και των σημάνσεων, η γνώση της υγιεινής διατροφής και ο έλεγχος του βάρους. Η διατροφική γνώση μπορεί να διευκολύνει τη χρήση της ετικέτας αυξάνοντας τα αντιληπτά οφέλη και αυξάνοντας την αποτελεσματικότητα της χρήσης της ετικέτας, μειώνοντας έτσι το κόστος χρήσης των ετικετών.

v. Παράγοντες παρακίνησης

Στους παράγοντες αυτούς συγκαταλέγονται το ενδιαφέρον για την ανάγνωση των διατροφικών πληροφοριών κατά την αγορά τροφίμων, η συσχέτιση της διαφήμισης με τις αναγραφόμενες στο τρόφιμο διατροφικές πληροφορίες, η απόφαση για αλλαγές στην αγοραστική συμπεριφορά και στη διατροφή κ.α.

vi. Άλλοι προσδιοριστικοί παράγοντες

Στους άλλους παράγοντες περιλαμβάνεται η συχνότητα ανάγνωσης και ο τυχόν σκεπτικισμός έναντι των διατροφικών ισχυρισμών. Οι ερευνητές έχουν δείξει ότι ο σκεπτικισμός έναντι των διατροφικών ισχυρισμών έχει αρνητική επίδραση στη χρήση ισχυρισμών υγείας και θετική επίδραση στη χρήση διατροφικής ετικέτας. Ακόμη, η επίδραση της ανάγνωσης στην τελική αξιολόγηση του προϊόντος, καθώς και η στάση έναντι των διατροφικών συνηθειών επηρεάζουν την χρήση της ετικέτας (Drichoutis et al., 2005).

5. Σκοπός της μελέτης

Σκοπός της παρούσας πτυχιακής εργασίας είναι να μελετήσει τις μη-γραμμικές επιδράσεις σε γραμμικά και μη-γραμμικά υποδείγματα και να εξάγει κάποια μέτρα μη-γραμμικότητας του συγκεκριμένου υποδείγματος. Ακόμη, η πτυχιακή έχει στόχο να διερευνήσει τις επιδράσεις αλληλεπίδρασης των μεταβλητών και να διερευνήσει κατά πόσο αυτές οδηγούν σε ορθά συμπεράσματα για την επίδραση των ανεξάρτητων μεταβλητών επάνω στην εξαρτημένη. Προκειμένου να μελετηθεί αυτό, χρησιμοποιήθηκαν δευτερογενή δεδομένα από την Εθνική Έρευνα για την Υγεία και τη Διατροφή (National Health and Nutrition Examination Survey-NHANES) των ΗΠΑ, που αφορούν δεδομένα διατροφής και υγείας.

Περιγραφή της Έρευνας

Η Εθνική Έρευνα για την Υγεία και τη Διατροφή (NHANES) είναι ένα πρόγραμμα μελετών με σκοπό την αξιολόγηση της υγειονομικής και διατροφικής κατάστασης των ενηλίκων και των παιδιών στις Ηνωμένες Πολιτείες. Η έρευνα είναι μοναδική στο ότι συνδυάζει συνεντεύξεις και σωματικές εξετάσεις. Η NHANES είναι ένα σημαντικό πρόγραμμα του Εθνικού Κέντρου Στατιστικών Υγείας (NCHS). Το NCHS αποτελεί μέρος των κέντρων ελέγχου και πρόληψης νόσων (Center for Disease Control Prevention-CDC) και έχει την ευθύνη για την παραγωγή ζωτικών και υγειονομικών στατιστικών για τις ΗΠΑ.

Το πρόγραμμα NHANES ξεκίνησε στις αρχές της δεκαετίας του 1960 και έχει διεξαχθεί ως σειρά ερευνών που επικεντρώνονται σε διαφορετικές πληθυσμιακές ομάδες ή θέματα υγείας. Η έρευνα εξετάζει ένα εθνικό αντιπροσωπευτικό δείγμα 5.000 περίπου ατόμων κάθε χρόνο. Τα άτομα αυτά βρίσκονται σε κομητείες σε όλη τη χώρα, 15 από τις οποίες οι ερευνητές τις επισκέπτονται κάθε χρόνο.

Η συνέντευξη, στα πλαίσια της έρευνας NHANES περιλαμβάνει δημογραφικές, κοινωνικοοικονομικές, διατροφικές και σχετικές με την υγεία ερωτήσεις. Η έρευνα αποτελείται από ιατρικές, οδοντιατρικές και μετρήσεις φυσιολογίας, καθώς και εργαστηριακές δοκιμές που χορηγούνται από εκπαιδευμένο ιατρικό προσωπικό. Τα συμπεράσματα της έρευνας χρησιμοποιούνται για τον προσδιορισμό των σοβαρών ασθενειών και των παραγόντων κινδύνου για τις ασθένειες. Οι πληροφορίες χρησιμοποιούνται για την αξιολόγηση της διατροφικής κατάστασης και της σύνδεσής της με την προαγωγή της υγείας και την πρόληψη των ασθενειών. Επιπλέον, τα στοιχεία της έρευνας αυτής χρησιμοποιούνται σε επιδημιολογικές μελέτες και έρευνα στον τομέα των επιστημών υγείας, οι οποίες συμβάλλουν στην ανάπτυξη προγραμμάτων υγείας καθώς και στην ανάπτυξη των υγειονομικών γνώσεων για το Έθνος (NHANES).

5.1 Περιγραφή των δεδομένων

Στην παρούσα πτυχιακή εργασία χρησιμοποιήθηκαν τα δεδομένα από την έρευνα NHANES της χρονολογίας 2009-2010 με κύριο ερευνητικό ερώτημα το ποιοι είναι οι προσδιοριστικοί παράγοντες που παρακινούν ένα άτομο να διαβάσει τις διατροφικές ετικέτες των τροφίμων που αγοράζει. Τα δεδομένα αφορούν την χρονολογία 2009-2010 επειδή δεν υπήρχαν πιο

πρόσφατα δεδομένα που να περιλαμβάνουν ερωτήσεις που αφορούν τις προτιμήσεις των καταναλωτών ως προς τις διατροφικές ετικέτες.

Τα δεδομένα που χρησιμοποιήθηκαν προέρχονται από διαφορετικές βάσεις δεδομένων των ερευνών NHANES όπου το περιεχόμενο τους και ο αριθμός συμμετεχόντων περιγράφονται αναλυτικά παρακάτω:

A. Δημογραφικά δεδομένα που αφορούν σε ερωτήσεις δημογραφικών στοιχείων παρέχοντας πληροφορίες σε οικογενειακό και ατομικό επίπεδο. Οι ηλικιακές ομάδες-στόχοι για τις δημογραφικές ερωτήσεις διαφέρουν ανάλογα με το θέμα. Τα ερωτηματολόγια για τα δημογραφικά στοιχεία του δείγματος συλλέχθηκαν στο σπίτι, από εκπαιδευμένους ερευνητές που χρησιμοποιούν το σύστημα προσωπικών συνεντεύξεων με την βοήθεια Η/Υ (Computer-Assisted Personal Interviewing - CAPI). Άτομα ηλικίας 16 ετών και άνω συμμετείχαν στην συνέντευξη κατευθείαν, ενώ ένα πληρεξούσιο άτομο παρείχε πληροφορίες για συμμετέχοντες στην έρευνα που ήταν κάτω των 16 ετών και για άτομα που δεν μπορούσαν να απαντήσουν οι ίδιοι στις ερωτήσεις. Συνολικά ο αριθμός του δείγματος από τα δημογραφικά δεδομένα ανέρχεται στις 10537 παρατηρήσεις.

B. Δεδομένα συμπεριφοράς του καταναλωτή σχετικά με την διατροφή με τηλεφωνική συνέντευξη σε άτομα ηλικίας άνω των 16 ετών. Τα θέματα που περιλαμβάνονται στην ενότητα αυτή είναι:

- Παράγοντες που επηρεάζουν τις αποφάσεις σχετικά με το φαγητό.
- Χρήση των διατροφικών πληροφοριών των ετικετών των φαγητών εστιατορίου.
- Ευαισθητοποίηση σχετικά με το πρόγραμμα My Pyramid. Αυτό αφορά σε ένα διάγραμμα των διατροφικών αναγκών των ανθρώπων που έχει σχήμα πυραμίδας. Οι σπόροι και τα δημητριακά αντιπροσωπεύουν τη βάση της πυραμίδας. Πάνω από αυτά είναι φρούτα και λαχανικά, και στη συνέχεια κρέατα και γαλακτοκομικά προϊόντα. Τα λίπη και τα γλυκά βρίσκονται στην κορυφή.
- Στάσεις σχετικά με την αλλαγή της τρέχουσας διατροφής του ατόμου.
- Παράγοντες που επηρεάζουν τις επιλογές τροφίμων σε ένα μανάβικο ή σούπερ μάρκετ.
- Χρήση ετικετών τροφίμων και χρήση βιολογικών τροφίμων.

Τα δεδομένα αυτά προέρχονται από συμμετέχοντες ηλικίας 16 ετών και άνω. Τα δεδομένα που συλλέγονται για συμμετέχοντες ηλικίας 1-15 ετών δημοσιεύονται σε ξεχωριστό σύνολο δεδομένων. Για την έρευνα NHANES 2009-2010, υπήρχαν 6.399 άτομα ηλικίας 16 ετών και άνω που ήταν επιλέξιμα για τη συνέντευξη.

Γ. Δεδομένα καταναλωτικής συμπεριφοράς για διάφορα διατροφικά θέματα συμπεριφοράς καταναλωτών σε επίπεδο οικογένειας που προέρχονται από προσωπική συνέντευξη με τους ερωτώμενους. Τα θέματα που περιλαμβάνονται στην ενότητα αυτή είναι:

- Χρήση ειδικής διατροφής από κάποιο μέλος της οικογένειας.
- Διαθεσιμότητα ορισμένων τύπων τροφίμων για την οικογένεια.
- Δαπάνες οικογενειακής διατροφής.
- Χρόνος που αφιερώνεται στο μαγείρεμα του φαγητού και αριθμός γευμάτων που πραγματοποιούνται ως οικογένεια.

Τα δεδομένα αυτά προέρχονται από συμμετέχοντες όλων των ηλικιών και γι' αυτό ο αριθμός των παρατηρήσεων ανέρχεται στις 10.537.

Δ. Δεδομένα που αφορούν στην αρτηριακή πίεση και την χοληστερόλη που παρέχει δεδομένα προσωπικής συνέντευξης σχετικά με την ευαισθητοποίηση, τη θεραπεία και τον έλεγχο της υψηλής αρτηριακής πίεσης και της υψηλής χοληστερόλης. Το επιλέξιμο δείγμα για τις ερωτήσεις αυτές απευθύνεται σε άτομα ηλικίας 16 ετών και άνω όταν πρόκειται για ερωτήσεις που αφορούν την αρτηριακή πίεση και σε άτομα ηλικίας 20 ετών και άνω όταν αφορούν ερωτήσεις σχετικά με την υψηλή χοληστερόλη.

Ε. Δεδομένα ιστορικού βάρους που περιλαμβάνουν δεδομένα προσωπικής συνέντευξης για διάφορα θέματα που σχετίζονται με το σωματικό βάρος, συμπεριλαμβανομένης της αυτοαντίληψης του βάρους, του αυτοαναφερόμενου βάρους κατά τη διάρκεια της ζωής του συμμετέχοντα, της απόπειρας απώλειας βάρους κατά τους τελευταίους 12 μήνες και τις μεθόδους που χρησιμοποιήθηκαν για να προσπαθήσουν να χάσουν βάρος και να αποφύγουν την αύξηση βάρους. Το δείγμα στόχος για τις ερωτήσεις σε αυτήν την ενότητα είναι άτομα ηλικίας 16 ετών και άνω.

5.2 Περιγραφή των μεταβλητών του υποδείγματος και περιγραφική στατιστική

Με βάση την βιβλιογραφία καταλήγουμε σε ένα υπόδειγμα που ως εξαρτημένη μεταβλητή θέτουμε την readlab η οποία αφορά στο πόσο συχνά ένα άτομο διαβάζει την διατροφική ετικέτα επάνω στα τρόφιμα που αγοράζει. Η μεταβλητή αυτή είχε 5 κατηγορίες (1=πάντα, 2=τις περισσότερες φορές, 3=μερικές φορές, 4=σπάνια, 5=ποτέ) και μετατράπηκε σε δυαδική μεταβλητή ως εξής: Οι κατηγορίες 1 έως 3 κατηγοριοποιήθηκαν σε μία και οι κατηγορίες 4 και 5 σε μία. Έτσι προέκυψε η μεταβλητή readlab ως: 1= διαβάζει τις διατροφικές ετικέτες 2=δεν διαβάζει τις διατροφικές ετικέτες. Η μεταβλητή readlab περιλαμβάνει 5175 παρατηρήσεις και έχει κατηγοριοποιηθεί στις εξής δύο κατηγορίες $y_i=0$: Πολύ συχνά-συχνά $y_i=1$: σπάνια-καθόλου. Στην ερώτηση που τέθηκε στους συμμετέχοντες 4003 (77%) άτομα απάντησαν ότι διαβάζουν συχνά έως πολύ συχνά τις διατροφικές ετικέτες ενώ 1172 (23%) άτομα απάντησαν ότι τις διαβάζουν πολύ σπάνια έως καθόλου.

Ως προς τις ανεξάρτητες μεταβλητές συμπεριλήφθηκαν οι παρακάτω μεταβλητές στο υπόδειγμα:

- **Age:** Η ηλικία του συμμετέχοντα (από 1 έτους έως 80 ετών)
- **Gender:** Το φύλλο του συμμετέχοντα (0=άντρας, 1=γυναίκα)
- **Educ:** Το επίπεδο εκπαίδευσης του συμμετέχοντα (1=μέχρι 2^η Λυκείου, 2=Απόφοιτος Λυκείου, 3=Απόφοιτος Πανεπιστημίου ή Τ.Ε.Ι., 4= Απόφοιτος Πανεπιστημίου και άνω)
- **Income:** Το ετήσιο εισόδημα του νοικοκυριού (χωρίζεται σε 12 εισοδηματικές κατηγορίες αύξοντα αριθμού)
- **Marital:** Η οικογενειακή κατάσταση του συμμετέχοντα (1=Παντρεμένος ή σε σχέση συγκατοίκησης, 2=Σε διάσταση ή διαζευγμένος, 3= Ανύπαντρος)

- **Nutrition:** Αφορά στο πόσο σημαντική είναι η διατροφική αξία για τον συμμετέχοντα όταν αγοράζει από ένα παντοπωλείο τρόφιμα (1=πολύ σημαντική, 2=σημαντική, 3=λίγο ή καθόλου σημαντική)
- **Price:** Αφορά στο πόσο σημαντική είναι η τιμή για τον συμμετέχοντα όταν αγοράζει από ένα παντοπωλείο τρόφιμα (1=πολύ σημαντική, 2=σημαντική, 3=λίγο ή καθόλου σημαντική)
- **Purchase:** Αφορά στο πόσο συχνά ένα άτομο αγοράζει είδη τροφίμων (1=περισσότερο από 1 φορά την εβδομάδα, 2=μία φορά την εβδομάδα, 3=μία φορά στις 2 εβδομάδες, 4=μία φορά τον μήνα ή και λιγότερο)
- **Diet:** Αφορά στο εάν το κάποιο άτομο της οικογένειας ακολουθεί κάποια συγκεκριμένη δίαιτα (0=ναι, 1=όχι)
- **Money:** Αφορά στο πόσα χρήματα ξοδεύει ο συμμετέχοντας κατά μέσο όρο όταν ψωνίζει σε ένα σούπερ μάρκετ
- **Bloodpres:** Αφορά στο εάν το άτομο έχει ή όχι υψηλή αρτηριακή πίεση στο αίμα (0=ναι, 1=όχι)

Ως προς τις ανεξάρτητες μεταβλητές του υποδείγματος μας για να προκύψει η παραπάνω κατηγοριοποίηση έγινε επεξεργασία ώστε να αφαιρεθούν τυχόν ελλιπής τιμές που προκύπτουν είτε διότι ο ερωτώμενος δεν έχει απαντήσει στην συγκεκριμένη ερώτηση είτε διότι απάντησε ότι δεν γνωρίζει είτε διότι η απάντηση του αποτελεί ακραία τιμή που είναι μη-λογική. Επιπλέον, προκειμένου να ικανοποιούνται οι προϋποθέσεις του Likelihood Ratio Test και συγκεκριμένα η προϋπόθεση ότι για να πραγματοποιηθεί το τεστ θα πρέπει να υπάρχουν τιμές για κάθε κατηγορία των μεταβλητών που περιλαμβάνονται στα δύο ένθετα υποδείγματα, έγινε συγχώνευση ορισμένων κατηγοριών στις μεταβλητές price nutrition marital educ και purchase. Συγκεκριμένα, οι μεταβλητές price και nutrition μετατράπηκαν από μεταβλητές με 4 κατηγορίες σε μεταβλητές με 3 κατηγορίες αφού συγχωνεύτηκε η κατηγορία 3 με την 4 (3=λίγο σημαντική, 4=καθόλου σημαντική), η μεταβλητή marital κατηγοριοποιήθηκε σε 3 κατηγορίες από 6 που είχε αρχικά ,αφού η κατηγορία 1 συγχωνεύτηκε με την κατηγορία 6 και οι κατηγορίες 2, 3 και 4 συγχωνεύτηκαν σε μία (1=παντρεμένος, 2=χήρος, 3=διαζευγμένος, 4=σε διάσταση, 5=ανύπαντρος, 6=σε σχέση συγκατοίκησης), η μεταβλητή educ μετατράπηκε από μεταβλητή με 5 κατηγορίες σε 4 κατηγορίες αφού συγχωνεύτηκε η κατηγορία 1 με την 2(1= απόφοιτος 2^{ος} Λυκείου, 2=απόφοιτος 3^{ος} Λυκείου) και τέλος η μεταβλητή purchase μετατράπηκε σε 4 κατηγορίες από 6 που είχε αρχικά, αφού οι κατηγορίες 5 και 6 μετατράπηκαν σε ελλιπής τιμές διότι είχαν πολύ λίγες παρατηρήσεις .

Στον Πίνακα 1, παρατίθενται μέτρα περιγραφικής στατιστικής για τις μεταβλητές του υποδείγματος μας.

Πίνακας 1: Μέτρα περιγραφικής στατιστικής για τις μεταβλητές του υποδείγματος

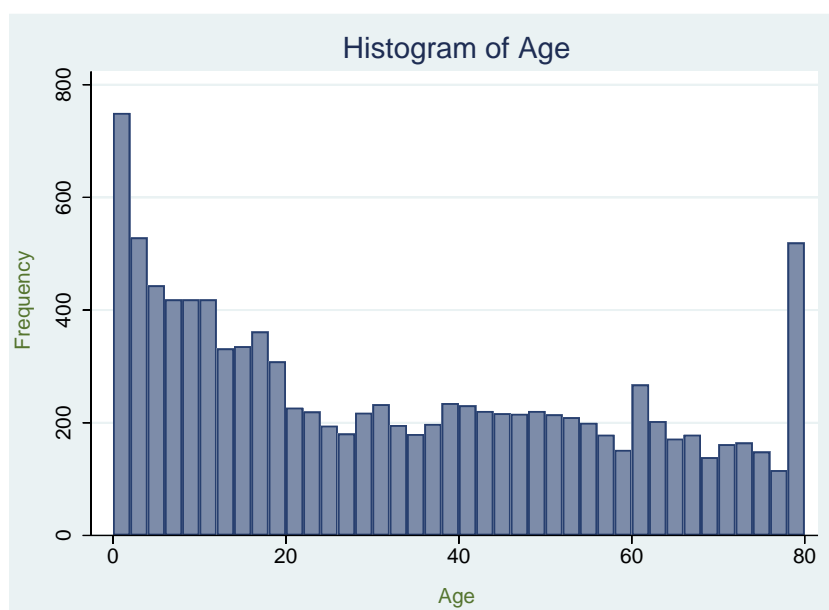
Μεταβλητή	Παρατηρήσεις	Μέσος Όρος	Τυπική Απόκλιση	Ελάχιστο	Μέγιστο
readlab	5,175	0,236	0,419	0	1
age	10,537	3,259	2,49	0	80
gender	10,537	0,504	0,5	0	1
educ	6,203	2,4	1,104	1	4

income	9,2	6,353	3,182	1	12
marital	6,214	1,587	0,772	1	3
price	5,185	1,647	0,672	1	3
purchase	10,207	2,436	0,883	1	4
nutrition	5,171	1,397	0,567	1	3
diet	6,119	0,687	0,464	0	1
bloodpres	6,876	0,677	0,468	0	1
money	10,227	4,418	312,97	0	6428

Παρακάτω παρατίθεται η περιγραφική στατιστική για τις ανεξάρτητες μεταβλητές του υποδείγματος:

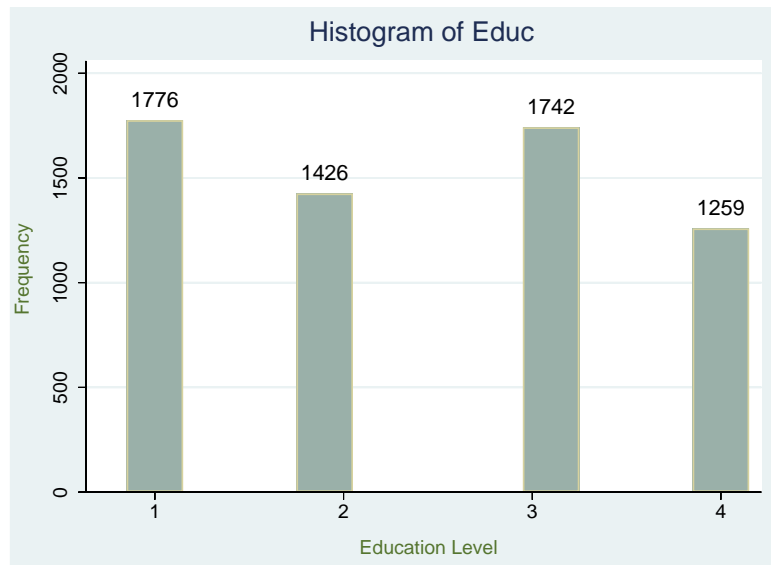
- Η μεταβλητή gender αποτελείται από 50% (5225 παρατηρήσεις) άντρες και 50% γυναίκες (5312) παρατηρήσεις.
- Η μεταβλητή age περιλαμβάνει περισσότερα άτομα ηλικίας από 1 έως 20 ετών σε σχέση με τα μεγαλύτερα ηλικιακά επίπεδα με εξαίρεση τα ηλικιακά επίπεδα 60 και 80 που εμφανίζουν μεγαλύτερη συχνότητα.

Εικόνα 5: Ιστόγραμμα της μεταβλητής Age



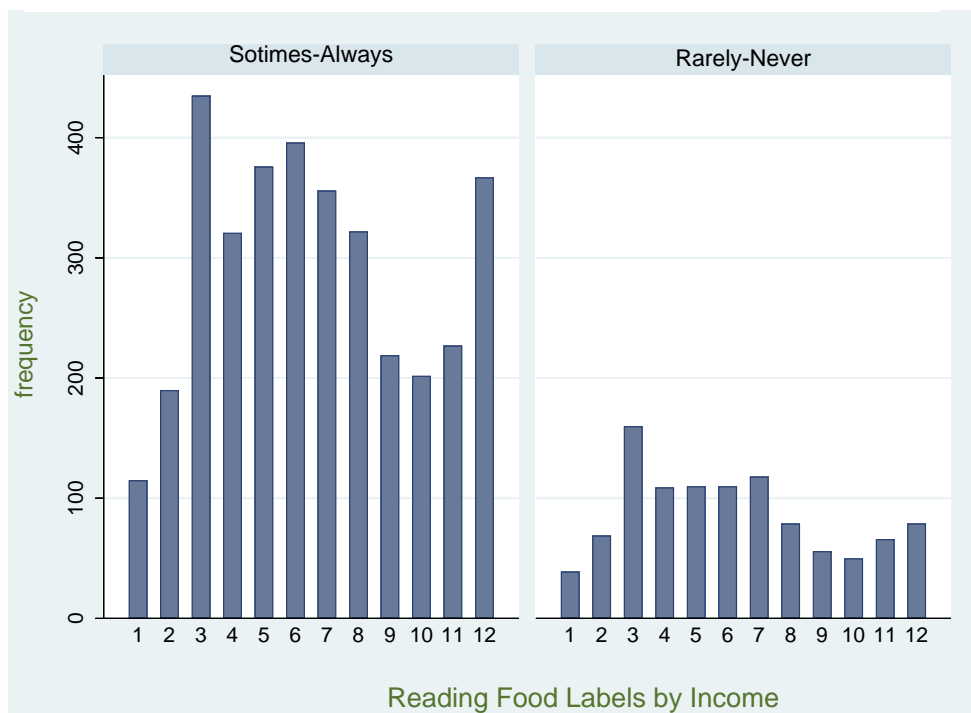
- Από το ιστόγραμμα της μεταβλητής educ παρατηρούμε ότι τα περισσότερα άτομα έχουν εκπαιδευτικό επίπεδο από Λύκειο έως Πανεπιστήμιο. Συγκεκριμένα, 23% του δείγματος είναι απόφοιτοι Λυκείου, 28% είναι απόφοιτοι Πανεπιστημίου και 20,4% είναι απόφοιτοι Πανεπιστημίου και άνω, ενώ μόλις το 28,6% του δείγματος είναι απόφοιτοι μέχρι και την 2^η Λυκείου.

Εικόνα 6: Ιστόγραμμα της μεταβλητής Educ



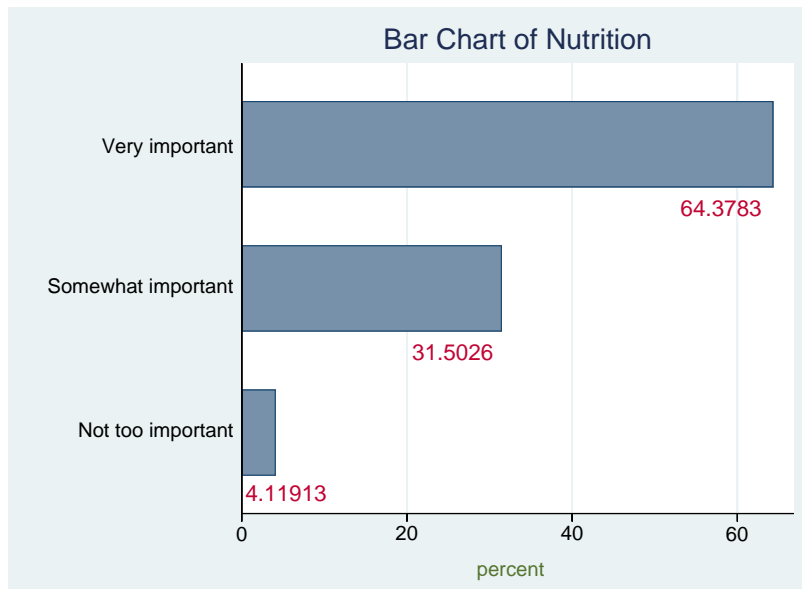
- Σχετικά με την μεταβλητή marital το μεγαλύτερο μέρος του δείγματος ανήκει στην κατηγορία «Παντρεμένος ή σε σχέση Συγκατοίκησης» με ποσοστό 59%, 18% του δείγματος δεν έχει παντρευτεί ποτέ ενώ το 23% του δείγματος ανήκει στην κατηγορία «Χωρισμένος ή Διαζευγμένος».
- Ως προς την μεταβλητή income βλέπουμε στο Ιστόγραμμα της μεταβλητής του εισοδήματος σε σχέση με την εξαρτημένη μεταβλητή readlab ότι τα περισσότερα άτομα που συνηθίζουν να διαβάζουν τις αναγραφόμενες ετικέτες στα τρόφιμα που αγοράζουν ανήκουν και σε μεγαλύτερη εισοδηματική κατηγορία.

Εικόνα 7: Ιστόγραμμα της μεταβλητής Income σε σχέση με την μεταβλητή Readlab



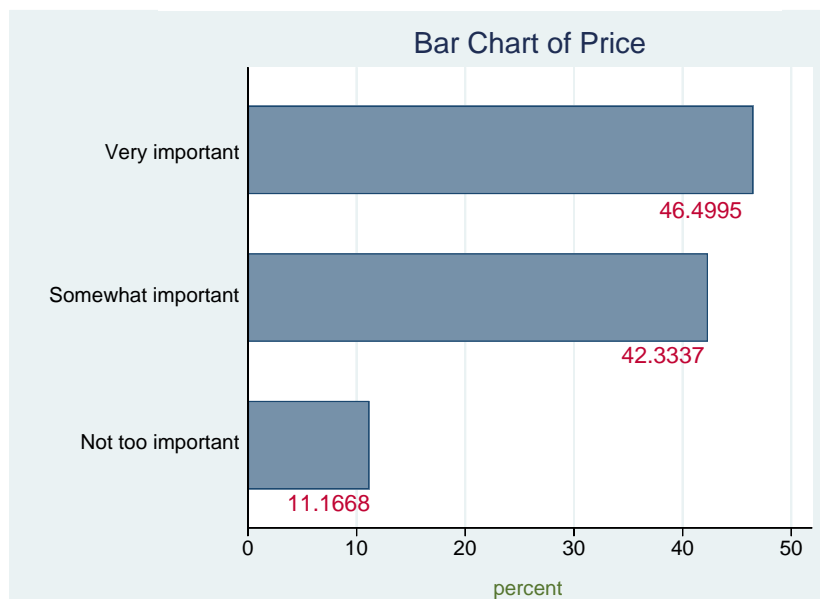
- Η μεταβλητή nutrition περιλαμβάνει κατά μεγαλύτερο ποσοστό άτομα που θεωρούν ως «Πολύ σημαντική-σημαντική» την διατροφική αξία όταν αγοράζουν τρόφιμα ενώ μόλις το 4% των συμμετεχόντων θεωρεί την διατροφική αξία ως «Λίγο έως καθόλου σημαντική».

Εικόνα 8: Ραβδόγραμμα της μεταβλητής Nutrition



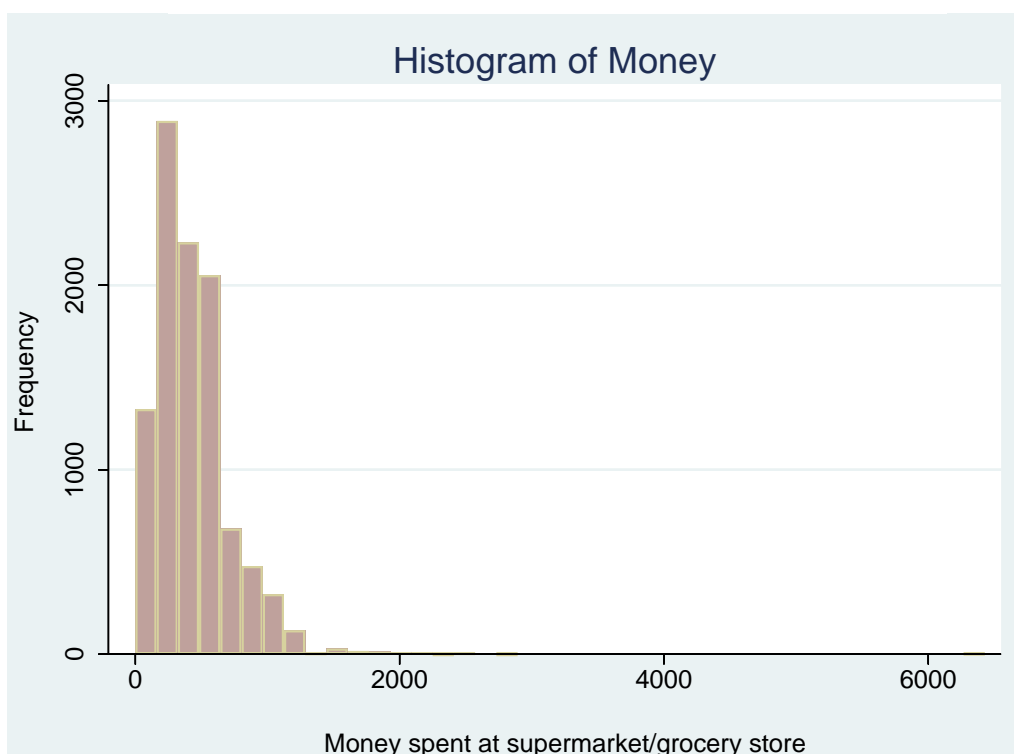
- Η μεταβλητή price περιλαμβάνει κατά μεγαλύτερο ποσοστό άτομα που θεωρούν ως «Πολύ σημαντική-σημαντική» την τιμή όταν αγοράζουν τρόφιμα ενώ μόλις το 11% των συμμετεχόντων θεωρούν την τιμή του προϊόντος ως «Λίγο έως Καθόλου σημαντική». Ως προς την μεταβλητή diet, το 69% των συμμετεχόντων δεν έχουν κάποιο μέλος της οικογένειάς τους να ακολουθεί κάποια συγκεκριμένη δίαιτα.

Εικόνα 9: Ραβδόγραμμα της μεταβλητής Price



- Για την μεταβλητή purchase παρατηρούμε ότι κατά 58% οι συμμετέχοντες αγοράζουν τουλάχιστον 1 φορά την εβδομάδα τρόφιμα από το super market ενώ το 28% αγοράζει τρόφιμα το πολύ κάθε 2 εβδομάδες και μόλις το 14% ψωνίζει το πολύ 1 φορά τον μήνα.
- Από το ιστόγραμμα της μεταβλητής money παρατηρούμε ότι εμφανίζει θετική ασυμμετρία εφόσον οι περισσότερες παρατηρήσεις είναι συγκεντρωμένες στις κατώτερες χρηματικές τιμές του δείγματος. Κάτι τέτοιο ήταν αναμενόμενο εφόσον κατά μέσο όρο ένα άτομο ξοδεύει 442€ στο σούπερ μάρκετ ενώ η μεταβλητή παίρνει τιμές μέχρι 6.500€.

Εικόνα 10: Ιστόγραμμα της μεταβλητής Money



- Τέλος, ως προς την μεταβλητή bloodpres παρατηρούμε ότι το 68% του δείγματος δεν έχει αρτηριακή πίεση ενώ 32% δήλωσε πως έχει.

6. Εξειδίκευση του Υποδείγματος- Το Διχοτομικό Υπόδειγμα Πιθανότητας

Προκειμένου να μελετηθεί το πώς οι ανεξάρτητες μεταβλητές που επιλέχθηκαν επιδρούν στην πιθανότητα το άτομο να διαβάζει τις διατροφικές ετικέτες όταν αγοράζει τρόφιμα και να προσδιοριστεί η σχέση αυτή χρησιμοποιήθηκε το υπόδειγμα λογιστικής παλινδρόμησης. Στην συνέχεια, με βάση αυτό το υπόδειγμα θα εξάγουμε μετρήσεις σχετικά με την γραμμικότητα και τη μη-γραμμικότητα των οριακών επιδράσεων των μεταβλητών μας. Για να τρέξουμε το υπόδειγμα μας και να εξάγουμε τις μετρήσεις μας, χρησιμοποιήθηκε το στατιστικό πρόγραμμα Stata και το υπόδειγμα logit.

Σύμφωνα με τις ανεξάρτητες μεταβλητές που έχουμε ήδη περιγράψει παραπάνω, το διχοτομικό υπόδειγμα παλινδρόμησης στην γενική του μορφή περιγράφεται από την εξίσωση:

$$Pr(\text{readlab} = 1) = F(b_0 + b_1\text{age} + b_2\text{gender} + b_3\text{educ} + b_4\text{income} + b_5\text{marital} + b_6\text{price} + b_7\text{purchase} + b_8\text{nutrition} + b_9\text{diet} + b_{10}\text{bloodpres} + b_{11}\text{money}) \quad (52)$$

όπου:

readlab: το εάν το άτομο διαβάζει ή όχι τις διατροφικές ετικέτες τροφίμων

b_0 : ο σταθερός όρος

$b_1 \dots b_{11}$: οι συντελεστές των ανεξάρτητων μεταβλητών

Ο Πίνακας 2 παρουσιάζει τα αποτελέσματα των εκτιμήσεων από το υπόδειγμα της παλινδρόμησης. Αναλυτικότερα, παρουσιάζονται το ποιοι παράγοντες είναι στατιστικά σημαντικοί σε επίπεδο σημαντικότητας 5%, το πρόσημο των επιδράσεων αυτών, δηλαδή το εάν επιδρούν θετικά ή αρνητικά στην πιθανότητα να διαβάζει κάποιος τις ετικέτες τροφίμων. Τέλος, εμπεριέχονται και μερικά μέτρα καλής προσαρμογής του υποδείγματος στα δεδομένα.

Πίνακας 2: Αποτελέσματα παλινδρόμησης της μεταβλητής Readlab

Εξαρτημένη Μεταβλητή: readlab	Συντελεστής	Τυπικό Σφάλμα	p-value
Ανεξάρτητες Μεταβλητές			
age	-.0191168**	.0033017	0.000
gender			
No	-.3807879**	.0904396	0.000
educ			
High School Grad/GED or Equivalent	-.2279602	.1232823	0.064

Some College or AA degree	-.3546587*	.1214401	0.003
College Graduate or above	-.72631**	.145477	0.000
income	-.0130845	.0162503	0.421
marital			
Widowed-Divorced-Separated	.0856909	.1220274	0.483
Never married	-.0930007	.1252905	0.458
price			
Somewhat important	-.1959642*	.0978355	0.045
Not too important	-.0164284	.1504456	0.913
nutrition			
Somewhat important	.9180126**	.0938215	0.000
Not too important	2.061.196**	.2031978	0.000
diet			
No	.4845188**	.0986578	0.000
purchase			
Once a week	.1406231	.150672	0.351
Once every two weeks	.3452258*	.1605791	0.032
Once a month or less	.5292765*	.1782466	0.003
bloodpres			
No	.1886232	.1104312	0.088
money	-.0002497	.0001903	0.189
_cons	-.7491914*	.3057103	0.014
Log likelihood	-1639.3638		
R² McFadden	0.106		
AIC	3.316.728		
R²_{count}	0.788		

- Από την στήλη των συντελεστών μπορούμε να ερμηνεύσουμε το πρόσημο των εκτιμήσεων και όχι το ίδιο το νούμερο ποσοτικά. Δηλαδή όταν το πρόσημο είναι θετικό τότε αυτό σημαίνει ότι η μεταβλητή επιδρά θετικά στην πιθανότητα ένα άτομο να διαβάσει τις διατροφικές ετικέτες και όταν το πρόσημο είναι αρνητικό αυτό σημαίνει ότι η μεταβλητή επιδρά αρνητικά στην πιθανότητα ένα άτομο να διαβάσει τις διατροφικές ετικέτες. Στην περίπτωση που αναφερόμαστε σε ψευδομεταβλητή τότε το θετικό πρόσημο σημαίνει ότι η συγκεκριμένη κατηγορία της ψευδομεταβλητής έχει μεγαλύτερη πιθανότητα το άτομο να διαβάσει τις διατροφικές ετικέτες σε σχέση με την κατηγορία αναφοράς, που στην περίπτωσή μας είναι η κατηγορία 1, και το αρνητικό πρόσημο σημαίνει ότι η συγκεκριμένη κατηγορία έχει μικρότερη πιθανότητα το άτομο να διαβάσει τις διατροφικές ετικέτες σε σχέση με την κατηγορία αναφοράς (Hosmer et al., 2000,2013).

- Από την στήλη p-value μπορούμε να εξάγουμε την στατιστική σημαντικότητα των μεταβλητών σε επίπεδο σημαντικότητας 5%. Η λογική του ελέγχου έχει ως εξής: Η μηδενική υπόθεση είναι ότι η τιμή του συντελεστή στον πληθυσμό μπορεί να είναι μηδέν, οπότε αν δεν απορρίψουμε την μηδενική υπόθεση τότε αυτό συνεπάγεται ότι η μεταβλητή δεν είναι στατιστικά σημαντική Η τιμή p-value ερμηνεύεται ως την πιθανότητα να κάνουμε λάθος αν απορρίψουμε την μηδενική υπόθεση. Οπότε όταν η τιμή p-value είναι μικρότερη του 5% (=0.05) τότε απορρίπτουμε την μηδενική υπόθεση (Greenland et al., 2016). Από τον Πίνακα 7 βλέπουμε πως στατιστικά σημαντικές είναι οι μεταβλητές:
 - ✓ Age
 - ✓ Gender
 - ✓ Educ (στατιστικά σημαντικές είναι οι κατηγορίες 3 και 4)
 - ✓ Price (στατιστικά σημαντική είναι η κατηγορία 2)
 - ✓ Nutrition
 - ✓ Diet
 - ✓ Purchase (στατιστικά σημαντικές είναι οι κατηγορίες 3 και 4)
- Στο τέλος του πίνακα 7 παίρνουμε κάποια μέτρα καλής προσαρμογής του υποδείγματος στα δεδομένα. Στην βιβλιογραφία έχουν προταθεί διάφορα μέτρα που βασίζονται στην πιθανοφάνεια, που κάποια από αυτά χρησιμοποιούν την λογική του R^2 . Αυτά τα μέτρα ονομάζονται *pseudo-R²* και παίρνουν τιμές από 0 έως 1, με το 0 να υποδεικνύει ένα μοντέλο με καθόλου ερμηνευτική ικανότητα και με το 1 να υποδεικνύει ένα υπόδειγμα με πλήρη ερμηνευτική ικανότητα. Η τιμή log-likelihood (LL) παίρνει πάντα αρνητικές τιμές και όσο μικρότερη είναι η τιμή της κατά απόλυτο τόσο καλύτερα προσαρμόζεται το υπόδειγμα στα δεδομένα. Παρόλο που η τιμή $-2LL$ μπορεί να χρησιμοποιηθεί ως δείγμα καλής προσαρμογής, επειδή η τιμή της επηρεάζεται από τον αριθμό των παρατηρήσεων, συνήθως χρησιμοποιείται ένα μέτρο που συγκρίνει την πιθανοφάνεια ενός μηδενικού υποδείγματος (δηλαδή ένα υπόδειγμα μόνο με σταθερό όρο) με το πλήρες υπόδειγμα (Best, 2013). Αυτή την λογική χρησιμοποιεί και το McFadden pseudo- R^2 με τον τύπο:

$$R^2_{McF} = 1 - \frac{\ln \hat{L}(M_{Full}) - k}{\ln \hat{L}(M_{Intercept})} \quad (53)$$

Όπου:

M_{Full} : Το πλήρες υπόδειγμα με όλες τις ανεξάρτητες μεταβλητές

$M_{Intercept}$: Το μηδενικό υπόδειγμα που περιλαμβάνει μόνο τον σταθερό όρο

k : Ο αριθμός των ανεξάρτητων μεταβλητών

Στο παράδειγμα μας έχουμε $R^2_{McF} = 0,106$ δηλαδή, μια τιμή πιο κοντά στο 0 απ' ότι στο 1 κάτι που υποδεικνύει μικρή ερμηνευτική ικανότητα του υποδείγματος. Παρόλο αυτά, τόσο ο λογάριθμος της πιθανοφάνειας όσο και το R^2_{McF} χρησιμοποιούνται κυρίως στην σύγκριση υποδειγμάτων και δεν ερμηνεύονται σαν νούμερα ποσοτικά. Όλα τα μέτρα καλής προσαρμογής όμως, υπόκεινται στον

περιορισμό ότι παίρνουν τόσο μεγαλύτερη τιμή όσο περισσότερες είναι οι ανεξάρτητες μεταβλητές του υποδείγματος γι' αυτό και πολλοί ερευνητές προτείνουν την χρήση του κριτηρίου του Akaike που δίνεται από τον τύπο:

$$AIC = -2\ln LL(M_{Full}) + 2(k + 1) \quad (54)$$

- Τέλος παίρνουμε ένα μέτρο που ονομάζεται R^2_{count} και μας δίνει το ποσοστό ορθών προβλέψεων. Το ποσοστό ορθό προβλέψεων προκύπτει από έναν πίνακα ταξινόμησης που αφορά σε ένα μέσο εξαγωγής της προβλεπτικής ικανότητας του υποδείγματος μας. Ουσιαστικά, πρόκειται για έναν πίνακα που ταξινομεί τις προβλεπόμενες και παρατηρούμενες τιμές έτσι ώστε να εκτιμάται ο αριθμός των περιπτώσεων όπου το μοντέλο προέβλεψε ορθά. Από τον πίνακα ταξινόμησης υπολογίζεται στην συνέχεια το μέτρο R^2_{count} βάσει του τύπου $R^2_{count} = \sum njj/n$, όπου n_{jj} : ορθές προβλέψεις και n : ο αριθμός των παρατηρήσεων. Στο παράδειγμα μας, το $R^2_{count} = 0,788$ που σημαίνει ότι το υπόδειγμα μας προβλέπει σωστά στο 78,8% των περιπτώσεων (Agresti, 1996).

7. Ερευνητικά Ερωτήματα, Μεθοδολογία Έρευνας και Αποτελέσματα

Οι εφαρμοσμένοι οικονομολόγοι χρησιμοποιούν συχνά όρους αλληλεπίδρασης στην εξειδίκευση των υποδειγμάτων τους για να διερευνήσουν πώς η επίδραση μιας ανεξάρτητης μεταβλητής στην εξαρτημένη μεταβλητή εξαρτάται από το μέγεθος μιας άλλης ανεξάρτητης μεταβλητής. Έτσι, είναι πολύ συνηθισμένο για τους ερευνητές να εισάγουν όρους αλληλεπίδρασης σε ένα υπόδειγμα όταν θέλουν να συμπεριλάβουν σε αυτό μη γραμμικές επιδράσεις. Η έννοια της μη-γραμμικότητας σε γραμμικά υποδείγματα συνεπάγεται τις περισσότερες φορές την έννοια της αλληλεπίδρασης. Παρόλο αυτά, όπως δείξαμε στο θεωρητικό κομμάτι (Εξίσωση 44), στα μη-γραμμικά υποδείγματα ακόμα και όταν ο συντελεστής του όρου αλληλεπίδρασης είναι μη-στατιστικά σημαντικός, δηλαδή η επίδραση του είναι μηδενική, παρατηρείται αλληλεπίδραση μεταξύ των ανεξάρτητων μεταβλητών.

Αυτό που θέλουμε να ερευνήσουμε στην συγκεκριμένη Πτυχιακή Εργασία είναι αφενός το κατά πόσο είναι ορθό να συμπεριλάβουμε σε ένα μη-γραμμικό υπόδειγμα όρους αλληλεπίδρασης όταν θέλουμε να λάβουμε υπόψη την μη-γραμμικότητα των παραμέτρων και αφετέρου εάν οδηγούμαστε σε λάθος συμπεράσματα για τον τρόπο που επιδρούν οι μεταβλητές στην εξαρτημένη μεταβλητή όταν σε ένα υπόδειγμα περιέχεται όρος αλληλεπίδρασης σε σχέση με ένα υπόδειγμα χωρίς όρο αλληλεπίδρασης.

Η μεθοδολογία που ακολουθήσαμε προκειμένου να ερευνήσουμε αυτά τα ερωτήματα χωρίζεται σε τρία μέρη. Στο πρώτο μέρος ασχοληθήκαμε με την διερεύνηση του κατά πόσο ένα υπόδειγμα με αλληλεπίδραση προσαρμόζεται καλύτερα στα δεδομένα σε σύγκριση με ένα υπόδειγμα χωρίς αλληλεπίδραση. Στο δεύτερο μέρος διερευνήσαμε το κατά πόσο οδηγούμαστε σε διαφορετικά συμπεράσματα για τον τρόπο που επιδρούν οι ανεξάρτητες μεταβλητές στην εξαρτημένη όταν συμπεριλαμβάνουμε στο υπόδειγμα μας όρο αλληλεπίδρασης σε σχέση με ένα υπόδειγμα που δεν περιλαμβάνει αλληλεπίδραση. Στο τρίτο και τελευταίο μέρος, καταλήξαμε στις περιπτώσεις όπου οδηγούμαστε σε εσφαλμένο συμπέρασμα εάν χρησιμοποιήσουμε σε ένα υπόδειγμα όρο αλληλεπίδρασης.

7.1 Όροι Αλληλεπίδρασης και Σύγκριση Υποδειγμάτων

Προκειμένου να μελετήσουμε το πρώτο ερευνητικό ερώτημα που αφορά την καλή προσαρμογή στα δεδομένα για υποδείγματα με και χωρίς αλληλεπίδραση εργαστήκαμε ως εξής:

- Αρχικά, δημιουργήσαμε υποδείγματα με όλες τις πιθανές αλληλεπιδράσεις των ανεξάρτητων μεταβλητών που συμπεριλάβαμε στο υπόδειγμα μας (Εξίσωση 52). Έτσι, εφόσον έχουμε 11 ανεξάρτητες μεταβλητές προέκυψαν 110 υποδείγματα για όλους τους συνδυασμούς αλληλεπιδράσεων των μεταβλητών. Στους συνδυασμούς αυτούς δεν περιλαμβάνονται οι συνδυασμοί που αφορούν στην αλληλεπίδραση μιας μεταβλητής με τον εαυτό της διότι στις περισσότερες των περιπτώσεων το

υπόδειγμα μας περιέχει κατηγορικές μεταβλητές που όταν πολλαπλασιάζονται με τον εαυτό τους προκύπτει η ίδια η μεταβλητή.

- Στην συνέχεια, προχωρήσαμε με την σύγκριση αυτών των υποδειγμάτων με το αρχικό υπόδειγμα που δεν περιέχει κανέναν όρο αλληλεπίδρασης. Όπως αναφέρθηκε και παραπάνω, για την σύγκριση 2 υποδειγμάτων μπορούμε να χρησιμοποιήσουμε ως μέτρα καλής προσαρμογής τα ψεύδο- R^2 και τα κριτήρια πληροφόρησης AIC και BIC. Όμως, επειδή τα μέτρα αυτά επηρεάζονται από τον αριθμό των ανεξάρτητων μεταβλητών μπορεί σε ορισμένες περιπτώσεις να οδηγήσουν σε λάθος συμπέρασμα. Γι' αυτό και τα μέτρα αυτά χρησιμοποιούνται για την επιβεβαίωση των αποτελεσμάτων. Έτσι, για να συγκρίνουμε τα υποδείγματα μεταξύ τους χρησιμοποιήσαμε ένα likelihood ratio test. Το συγκεκριμένο test έχει ως εξής:

Θεωρήσαμε ως το υπόδειγμα χωρίς περιορισμό αυτό που περιλαμβάνει όρο αλληλεπίδρασης και το υπόδειγμα με περιορισμό το υπόδειγμα χωρίς τον όρο αλληλεπίδρασης. Έτσι, το αρχικό υπόδειγμα είναι ένθετο στο υπόδειγμα με αλληλεπίδραση με τον περιορισμό ο συντελεστής του όρου αλληλεπίδρασης να ισούται με το μηδέν. Έτσι έχουμε τις υποθέσεις:

H₀: Ο περιορισμός που επιβάλλεται στο υπόδειγμα με αλληλεπίδραση για να προκύψει το υπόδειγμα χωρίς αλληλεπίδραση είναι αληθής. Δηλαδή, το υπόδειγμα με αλληλεπίδραση δεν προσαρμόζεται καλύτερα στα δεδομένα.

H₁: Ο περιορισμός που επιβάλλεται στο υπόδειγμα με αλληλεπίδραση για να προκύψει το υπόδειγμα χωρίς αλληλεπίδραση δεν είναι αληθής. Δηλαδή, το υπόδειγμα με αλληλεπίδραση προσαρμόζεται καλύτερα στα δεδομένα.

Όταν απορρίπτουμε την μηδενική υπόθεση συνεπάγεται ότι το υπόδειγμα με όρο αλληλεπίδρασης δείχνει καλύτερη προσαρμογή στα δεδομένα. Στην συνέχεια, επιβεβαιώνουμε την σύγκριση με το κριτήριο πληροφόρησης AIC.

Για να γίνει κατανοητή η διαδικασία που ακολουθήσαμε θα παραθέσουμε 2 παραδείγματα, ένα για την περίπτωση που δεν απορρίπτουμε την μηδενική υπόθεση και ένα για την αντίθετη περίπτωση.

❖ Παράδειγμα 1: Αλληλεπίδραση της μεταβλητής **gender** με την μεταβλητή **income**

Το υπόδειγμα με αλληλεπίδραση της μεταβλητής gender με την μεταβλητή income, που είναι και το υπόδειγμα χωρίς περιορισμό περιγράφεται από την εξίσωση:

$$\begin{aligned} Pr(readlab = 1) = F(b_0 + b_1age + b_2gender + b_{24}gender * \\ income + b_3educ + b_4income + b_5marital + b_6price + \\ b_7purchase + b_8nutrition + b_9diet + b_{10}bloodpres + b_{11}money) \end{aligned} \quad (55)$$

(REG4)

Ενώ το υπόδειγμα με περιορισμό περιγράφεται από την εξίσωση:

$$Pr(readlab = 1) = F(b_0 + b_1age + b_2gender + b_3educ + b_4income + b_5marital + b_6price + b_7purchase + b_8nutrition + b_9diet + b_{10}bloodpres + b_{11}money) \text{ (REG1)} \quad (56)$$

Ο περιορισμός που θέλουμε να εξετάσουμε εάν είναι αληθής ή όχι είναι ότι $b_{24} = 0$. Τα αποτελέσματα του ελέγχου ήταν τα εξής:

Πίνακας 3: Έλεγχος λόγου πιθανοφάνειας για την αλληλεπίδραση της μεταβλητής Gender και Income

Likelihood-ratio test	LR chi2(1)	= 2.04
(Assumption: REG1 nested in REG4)	p-value	= 0.1531

Βλέπουμε λοιπόν ότι $p\text{-value}=0,1531 > 0,05$ που σημαίνει ότι δεν απορρίπτεται η μηδενική υπόθεση. Άρα σε αυτή την περίπτωση, το υπόδειγμα με όρο αλληλεπίδρασης δεν προσαρμόζεται καλύτερα στα δεδομένα σε σχέση με το υπόδειγμα χωρίς αλληλεπίδραση. Αυτό το αποτέλεσμα επιβεβαιώνεται και από τα μέτρα πληροφόρησης AIC των δύο υποδειγμάτων όπου το μέτρο AIC για το υπόδειγμα χωρίς αλληλεπίδραση (REG1) έχει μικρότερη τιμή απ' ότι στο υπόδειγμα με αλληλεπίδραση (REG4) :

Πίνακας 4: Κριτήρια πληροφόρησης AIC για την αλληλεπίδραση της μεταβλητής Gender και Income

	REG1	REG4
AIC	3433.6139	3439.7243

❖ Παράδειγμα 2: Αλληλεπίδραση της μεταβλητής age με την μεταβλητή educ

Το υπόδειγμα με αλληλεπίδραση της μεταβλητής age με την μεταβλητή educ, που είναι και το υπόδειγμα χωρίς περιορισμό περιγράφεται από την εξίσωση:

$$Pr(readlab = 1) = F(b_0 + b_1age + b_{13}age * educ + b_2gender + b_3educ + b_4income + b_5marital + b_6price + b_7purchase + b_8nutrition + b_9diet + b_{10}bloodpres + b_{11}money) \text{ (REG13)} \quad (57)$$

Ενώ το υπόδειγμα με περιορισμό περιγράφεται από την εξίσωση:

$$Pr(readlab = 1) = F(b_0 + b_1age + b_2gender + b_3educ + b_4income + b_5marital + b_6price + b_7purchase + b_8nutrition + b_9diet + b_{10}bloodpres + b_{11}money) \text{ (REG1)} \quad (58)$$

Ο περιορισμός που θέλουμε να εξετάσουμε εάν είναι αληθής ή όχι είναι ότι $b_{13} = 0$. Τα αποτελέσματα του ελέγχου ήταν τα εξής:

Πίνακας 5: Έλεγχος λόγου πιθανοφάνειας για την αλληλεπίδραση της μεταβλητής Age και Educ

Likelihood-ratio test	LR chi2(1)	= 16.78
(Assumption: REG1 nested in REG4)	p-value	= 0.0008

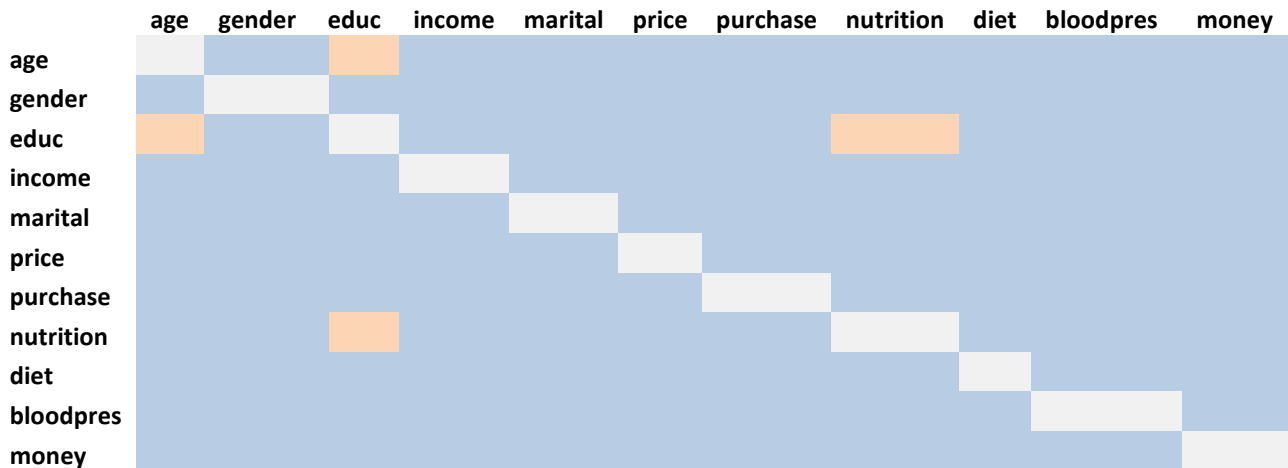
Βλέπουμε λοιπόν ότι $p\text{-value}=0.0008 < 0.05$ που σημαίνει ότι απορρίπτεται η μηδενική υπόθεση. Άρα σε αυτή την περίπτωση αλληλεπίδρασης, το υπόδειγμα με αλληλεπίδραση προσαρμόζεται καλύτερα στα δεδομένα σε σχέση με το υπόδειγμα χωρίς αλληλεπίδραση. Αυτό το αποτέλεσμα επιβεβαιώνεται και από τα μέτρα πληροφόρησης AIC των δύο υποδειγμάτων όπου το μέτρο AIC για το υπόδειγμα χωρίς αλληλεπίδραση (REG1) έχει μεγαλύτερη τιμή απ' ότι στο υπόδειγμα με αλληλεπίδραση (REG13) :



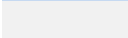
Πίνακας 6: Κριτήρια πληροφόρησης AIC για την αλληλεπίδραση της μεταβλητής Age και Educ

	REG1	REG13
AIC	3316.7276	3305.9514

- Στην συνέχεια, για να συνοψίσουμε τα αποτελέσματα όλων των συγκρίσεων των υποδειγμάτων δημιουργήσαμε τον παρακάτω πίνακα:

Πίνακας 7: Σύγκριση υποδειγμάτων με και χωρίς αλληλεπίδραση



-  Το υπόδειγμα με αλληλεπίδραση προσαρμόζεται καλύτερα στα δεδομένα
-  Το υπόδειγμα με αλληλεπίδραση δεν προσαρμόζεται καλύτερα στα δεδομένα
-  Δεν υπάρχουν μετρήσεις για αυτά τα υποδείγματα

Το γκρι χρώμα στον Πίνακα 7 αφορά στις περιπτώσεις που δεν υπήρξε σύγκριση κάποιων υποδειγμάτων εφόσον έχουμε αλληλεπίδραση μιας μεταβλητής με τον εαυτό της. Από τον παραπάνω πίνακα μπορούμε να εξάγουμε το ποσοστό των περιπτώσεων όπου το υπόδειγμα με αλληλεπίδραση δεν προσαρμόζεται καλύτερα στα δεδομένα το οποίο προκύπτει να είναι 96%. Αυτό σημαίνει ότι στο 96% των περιπτώσεων το υπόδειγμα χωρίς αλληλεπίδραση δείχνει καλύτερη προσαρμογή στα δεδομένα οπότε και η αλληλεπίδραση δεν προσέθεσε κάτι στην ερμηνευτική ικανότητα του υποδείγματος.

7.2 Οριακές Επιδράσεις σε Γραμμικά και Μη-γραμμικά Υποδείγματα

Το δεύτερο ερευνητικό ερώτημα μας, αφορά στο κατά πόσον οδηγούμαστε σε διαφορετικά συμπεράσματα για τον τρόπο που επιδρούν οι ανεξάρτητες μεταβλητές στην εξαρτημένη όταν συμπεριλαμβάνουμε στο υπόδειγμα μας όρο αλληλεπίδρασης σε σχέση με ένα υπόδειγμα χωρίς αλληλεπίδραση. Προκειμένου να διερευνήσουμε αυτές τις επιδράσεις εργαστήκαμε ως εξής:

- Αρχικά, αφού έχουμε εκτιμήσει τα 110 υποδείγματα με τις αλληλεπιδράσεις των ανεξάρτητων μεταβλητών, υπολογίσαμε τις οριακές επιδράσεις όλων των συνδυασμών των μεταβλητών και των επιπέδων τους τόσο στο υπόδειγμα χωρίς αλληλεπίδραση όσο και σε εκείνα τα υποδείγματα με αλληλεπίδραση. Για να υπολογίσουμε τις οριακές επιδράσεις χρησιμοποιήσαμε την εντολή του Stata (“margins”).
- Στην συνέχεια για κάθε συνδυασμό ανεξάρτητων μεταβλητών δημιουργήσαμε το διάγραμμα των οριακών επιδράσεων με την εντολή του Stata (“marginsplot”) ώστε να απεικονίσουμε γραφικά αυτές τις επιδράσεις. Από το διάγραμμα διασποράς των οριακών επιδράσεων βλέπουμε σε πρώτο στάδιο αν φαίνεται να υπάρχει κάποια αλληλεπίδραση στον τρόπο που επιδρούν οι μεταβλητές αλλά και αν υποδεικνύεται κάποια διαφορά στον τρόπο επίδρασης των ανεξάρτητων μεταβλητών σε υποδείγματα με αλληλεπίδραση συγκρινόμενα με αυτά χωρίς αλληλεπίδραση.
- Από μόνο του το διάγραμμα διασποράς των οριακών επιδράσεων δεν εγγυάται ότι η επίδραση ενός συνδυασμού μεταβλητών διαφέρει μεταξύ του υποδείγματος με αλληλεπίδραση σε σχέση με εκείνου χωρίς αλληλεπίδραση. Για να διαπιστώσουμε λοιπόν, αν η επίδραση μιας μεταβλητής, καθώς μεταβάλλεται μια άλλη μεταβλητή, διαφέρει μεταξύ των υποδειγμάτων με και χωρίς αλληλεπίδραση πραγματοποιήσαμε την εξής διαδικασία:
- Έστω ότι θέλουμε να εξετάσουμε πώς επιδρά η μεταβλητή του φύλου (gender) στην πιθανότητα ένα άτομο να διαβάζει τις διατροφικές ετικέτες, καθώς μεταβάλλεται η ηλικία (age). Στο υπόδειγμα χωρίς αλληλεπίδραση πραγματοποιήσαμε ένα t-test για να ελέγξουμε αν η οριακή επίδραση του φύλου για το πρώτο ηλικιακό επίπεδο μπορεί να ισούται με την οριακή επίδραση για το τελευταίο ηλικιακό επίπεδο. Στην περίπτωση που δεν απορρίψουμε την μηδενική υπόθεση αυτό σημαίνει ότι δεν υπάρχει στατιστικά σημαντική διαφορά στην επίδραση του φύλου στην εξαρτημένη μεταβλητή καθώς μεταβάλλεται η ηλικία.

Ακολούθως, εφαρμόζουμε τον ίδιο έλεγχο για το υπόδειγμα που περιλαμβάνει τον όρο αλληλεπίδρασης του φύλου με την ηλικία, και έτσι διακρίνουμε δύο περιπτώσεις:

- ✓ Αν δεν απορρίψουμε την μηδενική υπόθεση, σημαίνει ότι και στο υπόδειγμα με αλληλεπίδραση δεν υπάρχει στατιστικά σημαντική διαφορά στην επίδραση του φύλου καθώς μεταβάλλεται η ηλικία.
- ✓ Αν απορρίψουμε την μηδενική υπόθεση, σημαίνει ότι στο υπόδειγμα με αλληλεπίδραση η επίδραση του φύλου στην πιθανότητα το άτομο να διαβάσει τις διατροφικές ετικέτες, μεταβάλλεται καθώς αυξάνεται η ηλικία. Το συμπέρασμα αυτό είναι διαφορετικό από εκείνο που προέκυψε από το t-test στο υπόδειγμα χωρίς αλληλεπίδραση κάτι που μας δείχνει ότι θα οδηγούμασταν σε διαφορετικό συμπέρασμα για την επίδραση του φύλου καθώς μεταβάλλεται η ηλικία εάν λαμβάναμε υπόψη την αλληλεπίδραση μεταξύ των μεταβλητών.

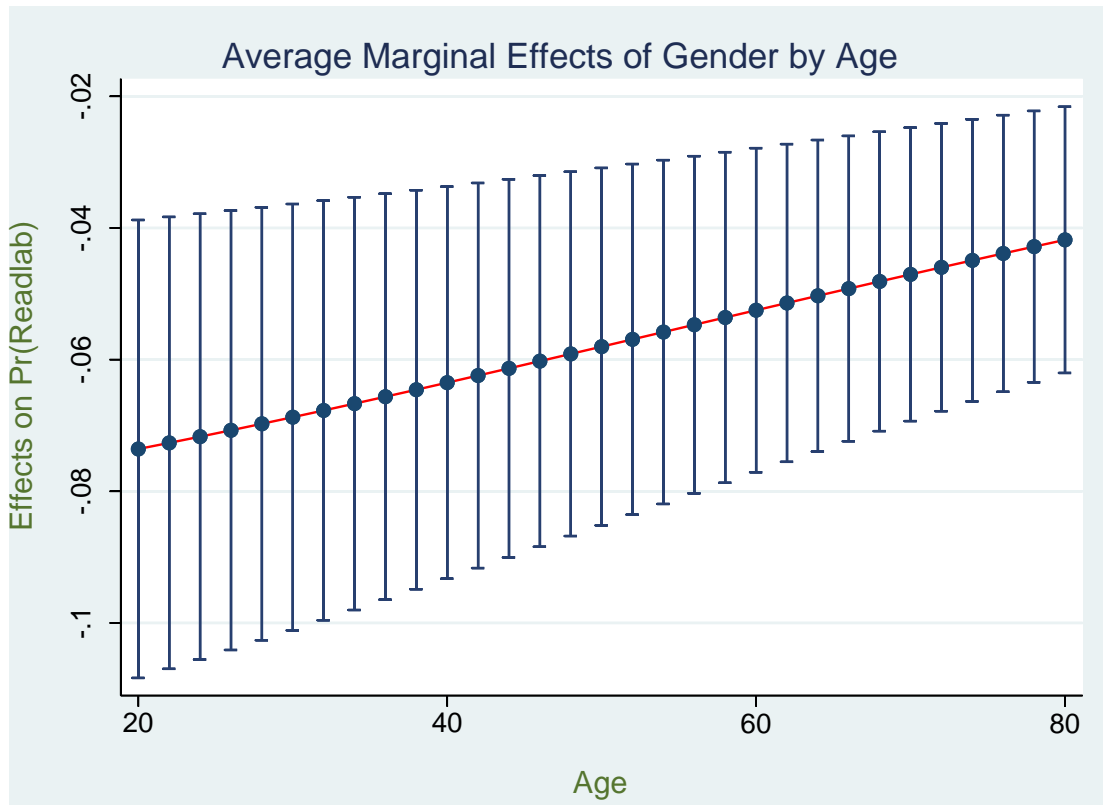
Παρακάτω θα περιγράψουμε την διαδικασία για μερικά αντιπροσωπευτικά παραδείγματα αλληλεπιδράσεων και στο τέλος θα συνοψίσουμε τα συμπεράσματα σε ένα πίνακα.

❖ **Παράδειγμα 1: Οριακή επίδραση του φύλου (gender) καθώς μεταβάλλεται η ηλικία (age)**

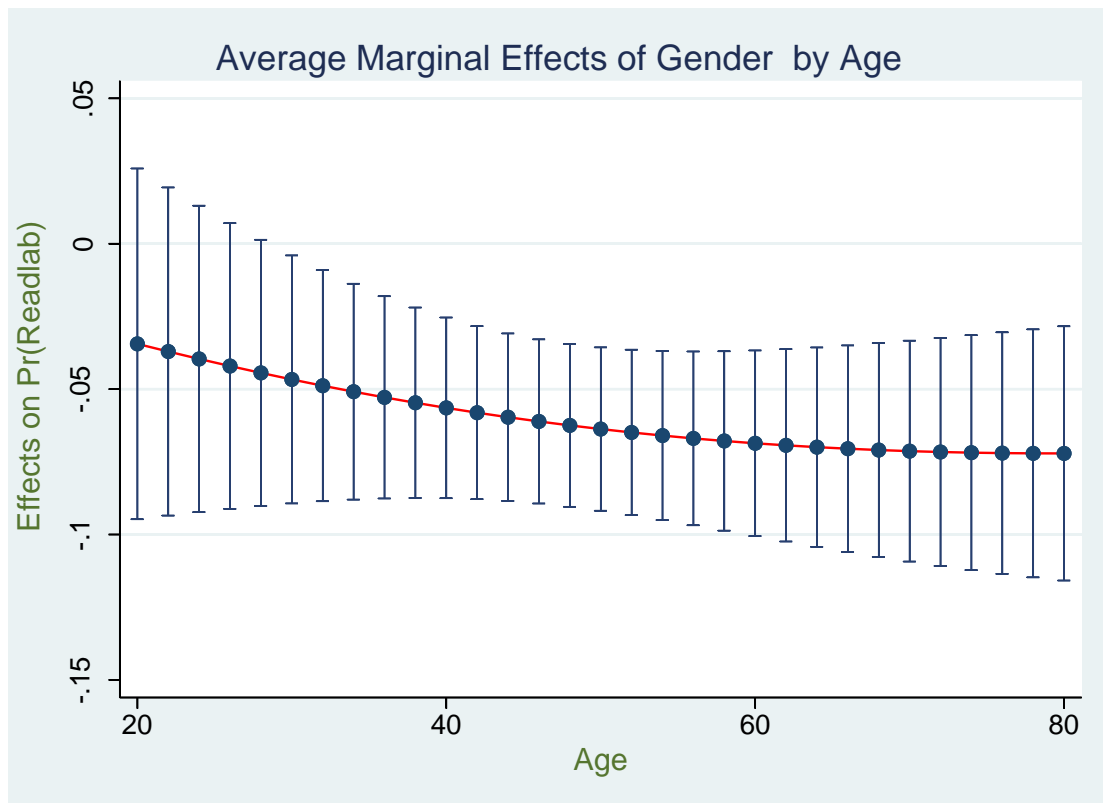
Αρχικά, υπολογίζουμε τις οριακές επιδράσεις του φύλου καθώς μεταβάλλεται η ηλικία. Επειδή η ηλικία είναι συνεχής μεταβλητή και παίρνει τιμές από 20 έως 80 θα υπολογίσουμε την οριακή επίδραση του φύλου για το διάστημα ηλικιών από [20,80]. Για το υπόδειγμα χωρίς αλληλεπίδραση δημιουργούμε αρχικά το διάγραμμα διασποράς των οριακών επιδράσεων και στην συνέχεια κάνουμε το ίδιο για το υπόδειγμα με αλληλεπίδραση ώστε τα συγκρίνουμε σε πρώτη φάση μόνο γραφικά.

Από το διάγραμμα της εικόνας 11, των οριακών επιδράσεων του φύλου καθώς μεταβάλλεται η ηλικία για το υπόδειγμα χωρίς αλληλεπίδραση παρατηρούμε ότι ένα άντρας έχει μικρότερη πιθανότητα να διαβάσει τις διατροφικές ετικέτες απ' ό τι μια γυναίκα για κάθε ηλικιακό επίπεδο. Ακόμη, παρατηρούμε ότι η επίδραση του φύλου στην πιθανότητα να διαβάσει τις διατροφικές ετικέτες αυξάνεται καθώς αυξάνεται η ηλικία του ατόμου. Στο σημείο αυτό είναι σημαντικό να παρατηρήσουμε ότι παρόλο που σε αυτό το υπόδειγμα δεν υπάρχει όρος αλληλεπίδρασης του φύλου με την ηλικία από το διάγραμμα διασποράς φαίνεται ξεκάθαρα ότι υπάρχει αλληλεπίδραση των μεταβλητών. Δηλαδή, οι μεταβλητές αυτές επιδρούν με μη-γραμμικό τρόπο επάνω στην εξαρτημένη ακόμα και χωρίς όρο αλληλεπίδρασης. Τέλος, στο διάγραμμα φαίνονται και διαστήματα εμπιστοσύνης των οριακών επιδράσεων από τα οποία βλέπουμε ότι όλες οι οριακές επιδράσεις είναι στατιστικά σημαντικές.

Εικόνα 11: Οριακές επιδράσεις του φύλου καθώς μεταβάλλεται η ηλικία για το υπόδειγμα χωρίς αλληλεπίδραση



Εικόνα 12: Οριακές επιδράσεις του φύλου καθώς μεταβάλλεται η ηλικία για το υπόδειγμα με αλληλεπίδραση



Από το διάγραμμα της εικόνας 12, των οριακών επιδράσεων του φύλλου καθώς μεταβάλλεται η ηλικία για το υπόδειγμα με αλληλεπίδραση παρατηρούμε ότι ένα άντρας έχει μικρότερη πιθανότητα να διαβάσει τις διατροφικές ετικέτες απ' ότι μια γυναίκα για κάθε ηλικιακό επίπεδο. Ακόμη, παρατηρούμε ότι η επίδραση του φύλου στην πιθανότητα να διαβάσει τις διατροφικές ετικέτες μειώνεται καθώς αυξάνεται η ηλικία του ατόμου. Το συμπέρασμα αυτό διαφέρει από το συμπέρασμα στο οποίο καταλήξαμε στο υπόδειγμα χωρίς αλληλεπίδραση. Τέλος, βλέπουμε από τα διαστήματα εμπιστοσύνης ότι για τα πρώτα ηλικιακά επίπεδα οι οριακές επιδράσεις δεν είναι στατιστικά σημαντικές.

Για να ελέγξουμε εάν όντως διαφέρει η επίδραση του φύλου καθώς μεταβάλλεται η ηλικία μεταξύ των δύο υποδειγμάτων πραγματοποιούμε τους παρακάτω ελέγχους:

Ελέγχουμε με ένα t-test το αν η οριακή επίδραση του φύλου για το πρώτο ηλικιακό επίπεδο μπορεί να είναι η ίδια με την οριακή επίδραση του φύλου για το μεσαίο και το τελευταίο ηλικιακό επίπεδο. Δηλαδή η μηδενική υπόθεση που ελέγχουμε είναι:

H₀: Η οριακή επίδραση του φύλου δεν διαφέρει μεταξύ διαφορετικών ηλικιακών επιπέδων

H₁: Η οριακή επίδραση του φύλου διαφέρει μεταξύ διαφορετικών ηλικιακών επιπέδων

Για το υπόδειγμα χωρίς αλληλεπίδραση:

Τα αποτελέσματα του t-test φαίνονται στον παρακάτω:

Πίνακας 8: Αποτελέσματα t-test για την μεταβλητή Age και Gender χωρίς αλληλεπίδραση

chi2 = 17.68
p-value = 0.0005

Από τον Πίνακα 8 βλέπουμε ότι **p-value=0.0005 < 0.05** οπότε απορρίπτουμε την μηδενική υπόθεση ότι η επίδραση του φύλου μπορεί να είναι ίδια μεταξύ των διαφορετικών ηλικιακών επιπέδων.

Για το υπόδειγμα με αλληλεπίδραση:

Τα αποτελέσματα του t-test φαίνονται στον παρακάτω:

Πίνακας 9: Αποτελέσματα t-test για την μεταβλητή Age και Gender με αλληλεπίδραση

chi2 = 4.32
p-value = 0.2290

Από τον Πίνακα 9 που αφορά στο υπόδειγμα με αλληλεπίδραση της μεταβλητής του φύλου με την ηλικία βλέπουμε ότι **p-value=0.2290 > 0.05**, οπότε δεν απορρίπτουμε την μηδενική

υπόθεση η επίδραση του φύλου μπορεί να είναι ίδια μεταξύ των διαφορετικών ηλικιακών επιπέδων. Δηλαδή, καταλήξαμε σε διαφορετικό συμπέρασμα απ' ό τι καταλήξαμε στο υπόδειγμα χωρίς αλληλεπίδραση, κάτι που είχαμε υποθέσει και από τα διαγράμματα διασποράς.

❖ **Παράδειγμα 2: Οριακή επίδραση της ηλικίας (age) καθώς μεταβάλλεται το εκπαιδευτικό ή επίπεδο (educ)**

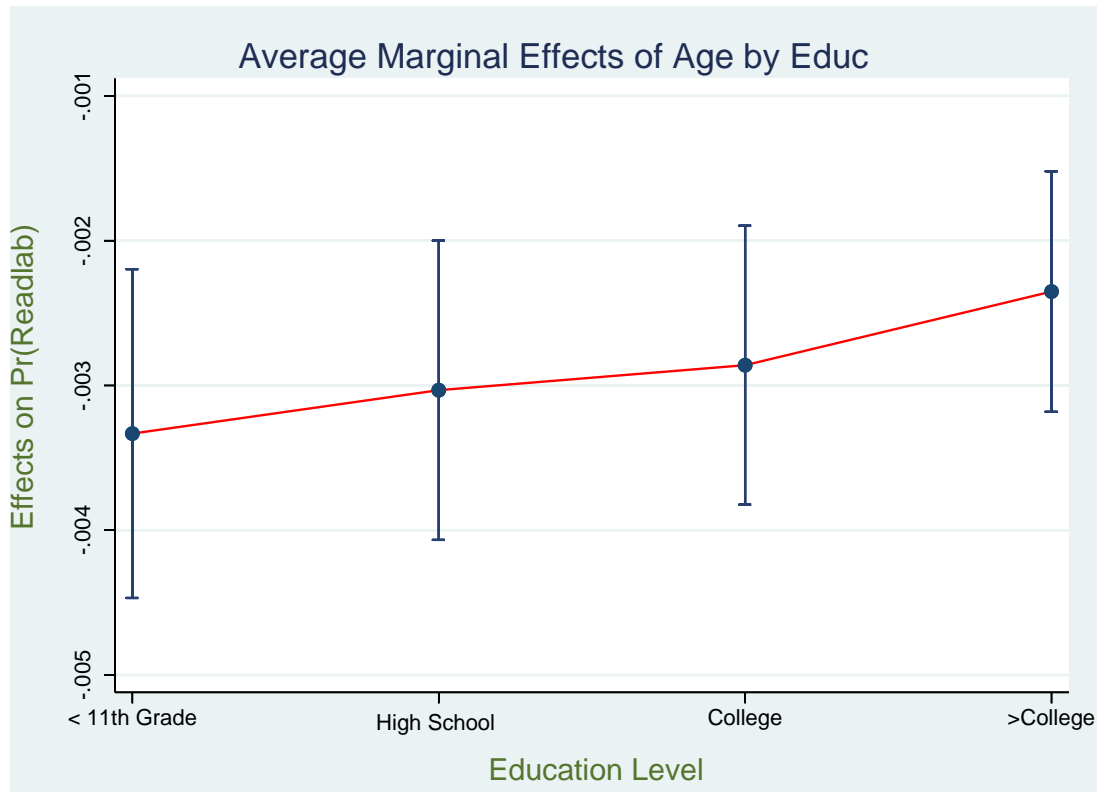
Αρχικά, υπολογίζουμε τις οριακές επιδράσεις της ηλικίας καθώς μεταβάλλεται το εκπαιδευτικό επίπεδο. Επειδή η μεταβλητή εκπαίδευση είναι ψευδομεταβλητή θα υπολογίσουμε την οριακή επίδραση της ηλικίας για κάθε εκπαιδευτικό επίπεδο .

Για το υπόδειγμα χωρίς αλληλεπίδραση δημιουργούμε αρχικά το διάγραμμα των οριακών επιδράσεων και στην συνέχεια κάνουμε το ίδιο για το υπόδειγμα με αλληλεπίδραση ώστε τα συγκρίνουμε σε πρώτη φάση μόνο γραφικά.

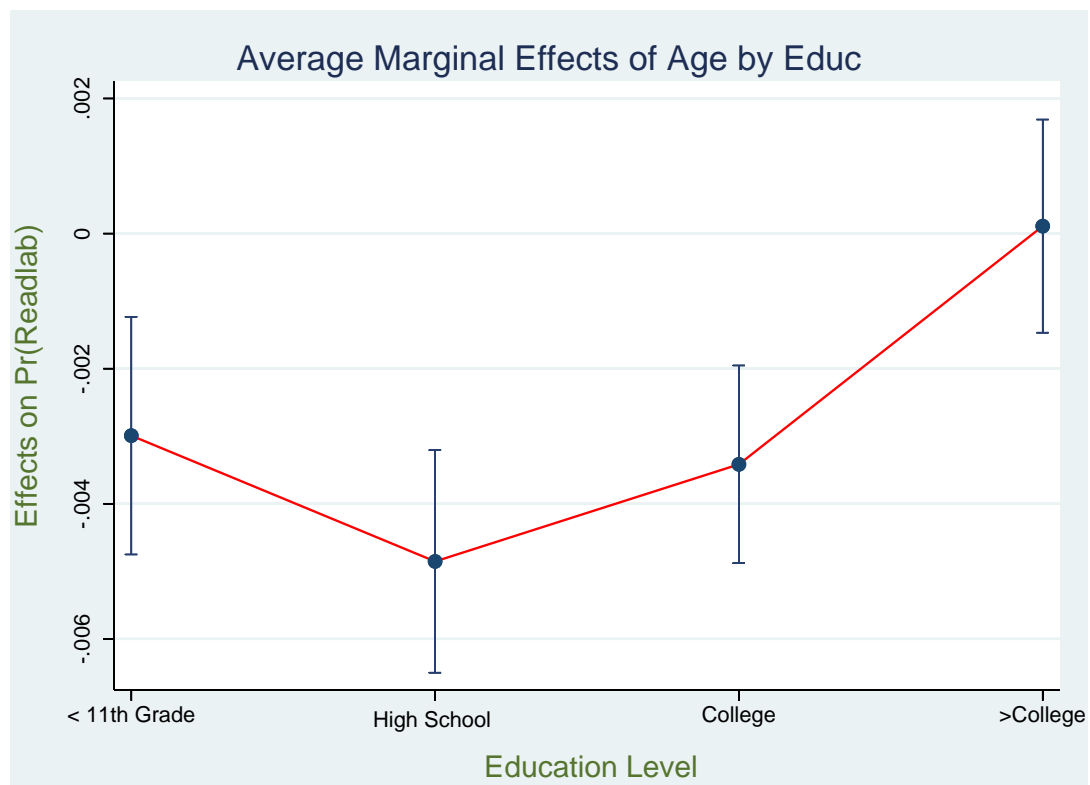
Από το γράφημα της Εικόνας 13, για το υπόδειγμα χωρίς αλληλεπίδραση, παρατηρούμε ότι για κάθε εκπαιδευτικό επίπεδο οι οριακές επιδράσεις έχουν σε κάθε περίπτωση αρνητικό πρόσημο. Το συμπέρασμα που εξάγεται είναι ότι καθώς αυξάνεται η ηλικία μειώνεται η πιθανότητα ένα άτομο να διαβάζει τις διατροφικές ετικέτες. Επιπλέον, παρατηρούμε ότι καθώς μεταβαίνουμε από το μικρότερο εκπαιδευτικό επίπεδο προς το μεγαλύτερο, η πιθανότητα να διαβάζει ένα άτομο τις διατροφικές ετικέτες αυξάνεται. Τέλος, βλέπουμε από τα διαστήματα εμπιστοσύνης ότι όλες οι οριακές επιδράσεις της ηλικίας καθώς αλλάζει το εκπαιδευτικό επίπεδο είναι στατιστικά σημαντικές. Υπογραμμίζεται πάλι το γεγονός ότι η ηλικία επιδρά με μη-γραμμικό τρόπο στην εξαρτημένη μεταβλητή εφόσον μεταβάλλεται καθώς μεταβάλλεται η μεταβλητή της εκπαίδευσης.

Από το γράφημα της Εικόνας 14, στο υπόδειγμα με αλληλεπίδραση, παρατηρούμε ότι καθώς αυξάνεται η ηλικία μειώνεται η πιθανότητα ένα άτομο να διαβάζει τις διατροφικές ετικέτες για όλα τα εκπαιδευτικά επίπεδα εκτός του τελευταίου επιπέδου που δείχνει μια θετική επίδραση της ηλικίας εφόσον έχει θετικό πρόσημο. Επιπλέον, βλέπουμε τα άτομα που έχουν τελειώσει το Λύκειο έχουν μικρότερη πιθανότητα να διαβάζουν τις διατροφικές ετικέτες σε σχέση με ένα άτομο που δεν το έχει τελειώσει. Ενώ, μετά το επίπεδο του Λυκείου και καθώς μεταβαίνουμε στο μεγαλύτερο εκπαιδευτικό επίπεδο αυξάνεται η πιθανότητα ένα άτομο να διαβάζει τις διατροφικές ετικέτες. Τέλος, παρατηρούμε ότι η οριακή επίδραση της ηλικίας για το τελευταίο εκπαιδευτικό επίπεδο δεν είναι στατιστικά σημαντική.

Εικόνα 13: Οριακές επιδράσεις της ηλικίας καθώς μεταβάλλεται το εκπαιδευτικό επίπεδο για το υπόδειγμα χωρίς αλληλεπίδραση



Εικόνα 14: Οριακές επιδράσεις της ηλικίας καθώς μεταβάλλεται το εκπαιδευτικό επίπεδο για το υπόδειγμα με αλληλεπίδραση



Τα αποτελέσματα μεταξύ των δύο υποδειγμάτων διαφέρουν. Για να ελέγξουμε εάν όντως διαφέρει η επίδραση της ηλικίας καθώς μεταβάλλεται το εκπαιδευτικό επίπεδο μεταξύ των δύο υποδειγμάτων πραγματοποιούμε τους παρακάτω ελέγχους:

Ελέγχουμε με ένα t-test το αν η οριακή επίδραση της ηλικίας για το πρώτο εκπαιδευτικό επίπεδο μπορεί να είναι η ίδια με την οριακή επίδραση της ηλικίας για το δεύτερο, το τρίτο και το τελευταίο εκπαιδευτικό επίπεδο. Δηλαδή η μηδενική υπόθεση που ελέγχουμε είναι:

H₀: Η οριακή επίδραση της ηλικίας δεν διαφέρει μεταξύ διαφορετικών εκπαιδευτικών επιπέδων

H₁: Η οριακή επίδραση της ηλικίας διαφέρει μεταξύ διαφορετικών εκπαιδευτικών επιπέδων

Για το υπόδειγμα χωρίς αλληλεπίδραση:

Τα αποτελέσματα του t-test φαίνονται στον παρακάτω πίνακα:

Πίνακας 10: Αποτελέσματα t-test για την μεταβλητή Age και Educ χωρίς αλληλεπίδραση

chi2 = 14.80
p-value = 0.0020

Από τον Πίνακα 10 βλέπουμε ότι **p-value=0.002 < 0.05** οπότε απορρίπτουμε την μηδενική υπόθεση ότι η επίδραση της ηλικίας μπορεί να είναι ίδια μεταξύ των διαφορετικών εκπαιδευτικών επιπέδων.

Για το υπόδειγμα με αλληλεπίδραση:

Τα αποτελέσματα του t-test φαίνονται στον παρακάτω:

Πίνακας 11: Αποτελέσματα t-test για την μεταβλητή Age και Educ με αλληλεπίδραση

chi2 = 22.07
p-value = 0.0001

Από τον πίνακα 11 βλέπουμε ότι **p-value=0.0001 < 0.05**, οπότε απορρίπτουμε την μηδενική υπόθεση ότι η επίδραση της ηλικίας μπορεί να είναι ίδια μεταξύ των διαφορετικών εκπαιδευτικών επιπέδων. Δηλαδή, και στα δύο υποδείγματα καταλήξαμε στο ίδιο συμπέρασμα ότι η οριακή επίδραση της ηλικίας διαφέρει για τα διαφορετικά εκπαιδευτικά επίπεδα.

❖ **Παράδειγμα 3: Οριακή επίδραση της εκπαίδευσης (educ) καθώς μεταβάλλεται η συχνότητα που αγοράζει κάποιος τρόφιμα (purchase)**

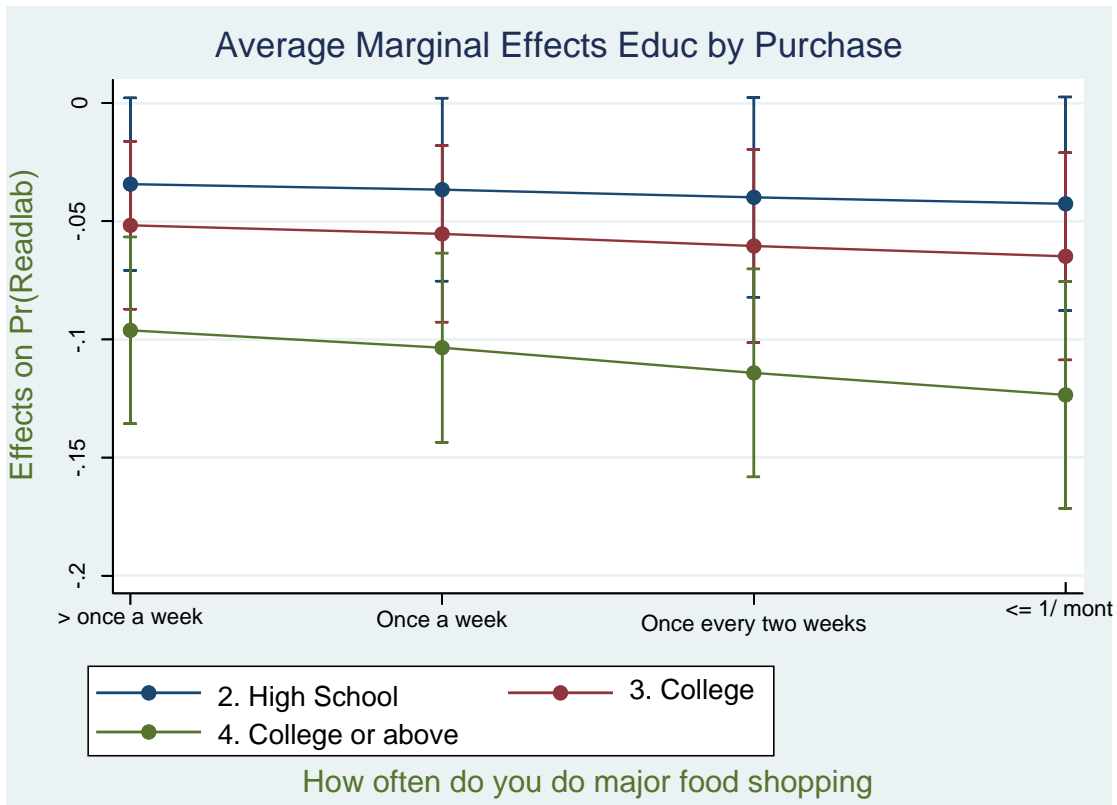
Αρχικά, υπολογίζουμε τις οριακές επιδράσεις της εκπαίδευσης καθώς μεταβάλλεται η συχνότητα όπου αγοράζει κάποιος τρόφιμα. Επειδή η μεταβλητή purchase είναι ψευδομεταβλητή θα υπολογίσουμε την οριακή επίδραση της εκπαίδευσης για κάθε επίπεδο της μεταβλητής purchase.

Για το υπόδειγμα χωρίς αλληλεπίδραση δημιουργούμε αρχικά το διάγραμμα διασποράς των οριακών επιδράσεων και στην συνέχεια κάνουμε το ίδιο για το υπόδειγμα με αλληλεπίδραση ώστε τα συγκρίνουμε σε πρώτη φάση μόνο γραφικά.

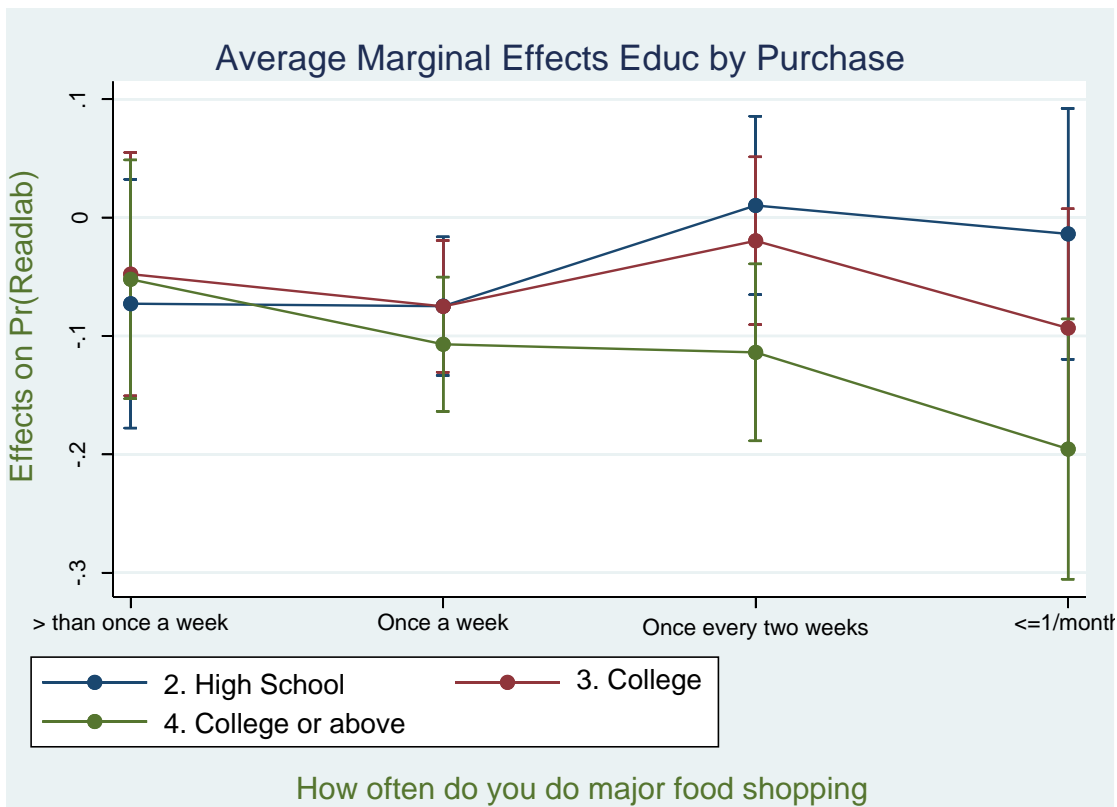
Από το γράφημα της εικόνας 15, για το υπόδειγμα χωρίς αλληλεπίδραση, παρατηρούμε ότι για κάθε εκπαιδευτικό επίπεδο η πιθανότητα κάποιος να διαβάζει τις ετικέτες τροφίμων δείχνει να έχει μειωτική τάση. Ακόμη, για κάθε επίπεδο συχνότητας αγοράς τροφίμων ένα άτομο με επίπεδο εκπαίδευσης Λυκείου έχει μικρότερη πιθανότητα να διαβάζει τις ετικέτες τροφίμων σε σχέση με ένα άλλο άτομο με τα ίδια χαρακτηριστικά αλλά με επίπεδο εκπαίδευσης μέχρι 2^η Λυκείου. Το ίδιο ισχύει και για ένα άτομο με επίπεδο εκπαίδευσης Πανεπιστημίου σε σχέση με ένα άτομο που έχει τελειώσει το Λύκειο και για ένα άτομο που έχει επίπεδο εκπαίδευσης μεγαλύτερο του πτυχίο Πανεπιστημίου σε σχέση με ένα άτομο που έχει πτυχίο Πανεπιστημίου. Επιπρόσθετα, παρατηρούμε ότι όλες οι οριακές επιδράσεις είναι στατιστικά σημαντικές.

Από το γράφημα της εικόνας 16, για το υπόδειγμα με αλληλεπίδραση, παρατηρούμε ότι για κάθε εκπαιδευτικό επίπεδο η πιθανότητα ένα άτομο να διαβάζει τις διατροφικές ετικέτες μεταβάλλεται καθώς μεταβαίνουμε από τον πρώτο επίπεδο της μεταβλητής purchase προς το τελευταίο. Αρχικά παρατηρούμε μια μικρή μείωση στην πιθανότητα μετά μια αύξηση και μετά πάλι μια μείωση. Για παράδειγμα, ένα άτομο που έχει πτυχίο Πανεπιστημίου και αγοράζει τρόφιμα μία φορά την εβδομάδα έχει μικρότερη πιθανότητα να διαβάζει τις ετικέτες τροφίμων σε σχέση με ένα άλλο άτομο με τα ίδια χαρακτηριστικά αλλά που αγοράζει τρόφιμα περισσότερο από μία φορά την εβδομάδα. Ακόμη, βλέπουμε πως για κάθε επίπεδο συχνότητας αγοράς τροφίμων η πιθανότητα να διαβάζει ένα άτομο τις ετικέτες τροφίμων μειώνεται καθώς μεταβαίνουμε από το μεγαλύτερο εκπαιδευτικό επίπεδο στο ακριβώς μικρότερο του, εφόσον το πρόσθετο της οριακής επίδρασης είναι αρνητικό. Τέλος, παρατηρούμε ότι υπάρχουν ορισμένες οριακές επιδράσεις που δεν είναι στατιστικά σημαντικές .

Εικόνα 15: Οριακές επιδράσεις της εκπαίδευσης καθώς μεταβάλλεται η συχνότητα αγοράς τροφίμων για το υπόδειγμα χωρίς αλληλεπίδραση



Εικόνα 16: Οριακές επιδράσεις της εκπαίδευσης καθώς μεταβάλλεται η συχνότητα αγοράς τροφίμων για το υπόδειγμα με αλληλεπίδραση



Τα συμπεράσματα για το γράφημα του υποδείγματος με αλληλεπίδραση φαίνεται ότι γραφικά είναι αρκετά διαφορετικά από αυτά που προέκυψαν στο υπόδειγμα χωρίς αλληλεπίδραση. Για να ελέγξουμε λοιπόν εάν όντως διαφέρει η επίδραση της εκπαίδευσης καθώς μεταβάλλεται η συχνότητα αγοράς τροφίμων μεταξύ των δύο υποδειγμάτων πραγματοποιούμε τους παρακάτω ελέγχους:

Στην συνέχεια, ελέγχουμε με ένα t-test το αν η οριακή επίδραση της εκπαίδευσης για το πρώτο επίπεδο συχνότητας αγοράς τροφίμων μπορεί να είναι η ίδια με την οριακή επίδραση της εκπαίδευσης για το δεύτερο, το τρίτο και το τελευταίο επίπεδο συχνότητας αγοράς τροφίμων. Στην προκειμένη περίπτωση επειδή η μεταβλητή educ περιλαμβάνει 4 κατηγορίες θα πραγματοποιήσουμε αυτόν τον έλεγχο για κάθε εκπαιδευτικό επίπεδο σε σχέση με το προηγούμενο εκπαιδευτικό επίπεδο. Δηλαδή, θα ελέγξουμε την μηδενική υπόθεση για το δεύτερο εκπαιδευτικό επίπεδο σε σύγκριση με το πρώτο, το τρίτο με το δεύτερο και το τέταρτο με το τρίτο. Η μηδενική υπόθεση του ελέγχου είναι:

H_0 : Η οριακή επίδραση της εκπαίδευσης δεν διαφέρει μεταξύ διαφορετικών επιπέδων συχνότητας αγοράς τροφίμων

H_1 : Η οριακή επίδραση της εκπαίδευσης διαφέρει μεταξύ διαφορετικών επιπέδων συχνότητας αγοράς τροφίμων

Για το υπόδειγμα χωρίς αλληλεπίδραση:

Τα αποτελέσματα των t-test φαίνονται παρακάτω:

Πίνακας 12: Αποτελέσματα t-test για την μεταβλητή Educ και Purchase με αλληλεπίδραση

educ=2
chi2 = 2.66
p-value = 0.4470
educ=3
Chi2 = 5.16
p-value = 0.1605
educ=4
chi2 = 9.04
p-value = 0.0287

Από τον Πίνακα 12 βλέπουμε ότι για το επίπεδο εκπαίδευσης 2 σε σχέση με το επίπεδο 1 παίρνουμε $p\text{-value}=0.4470 > 0.05$ οπότε δεν απορρίπτουμε την μηδενική υπόθεση ότι η επίδραση της εκπαίδευσης δεν διαφέρει μεταξύ διαφορετικών επιπέδων συχνότητας αγοράς τροφίμων .

Για το επίπεδο εκπαίδευσης 3 σε σχέση με το επίπεδο 2 παίρνουμε $p\text{-value}=0.1605 > 0.05$ οπότε δεν απορρίπτουμε την μηδενική υπόθεση ότι η επίδραση της εκπαίδευσης δεν διαφέρει μεταξύ διαφορετικών επιπέδων συχνότητας αγοράς τροφίμων .

Για το επίπεδο εκπαίδευσης 4 σε σχέση με το επίπεδο 3 παίρνουμε $p\text{-value}=0.0287 < 0.05$ οπότε απορρίπτουμε την μηδενική υπόθεση ότι η επίδραση της εκπαίδευσης δεν διαφέρει μεταξύ διαφορετικών επιπέδων συχνότητας αγοράς τροφίμων

Για το υπόδειγμα με αλληλεπίδραση:

Τα αποτελέσματα του t-test φαίνονται παρακάτω:

Πίνακας 13: Αποτελέσματα t-test για την μεταβλητή Educ και Purchase χωρίς αλληλεπίδραση

educ=2
chi2 = 3.71
p-value = 0.2940
educ=3
chi2 = 2.06
p-value = 0.5594
educ=4
chi2 = 3.71
p-value = 0.2951

Από τον Πίνακα 13 βλέπουμε ότι για το επίπεδο εκπαίδευσης 2 σε σχέση με το επίπεδο 1 παίρνουμε $p\text{-value}=0.294 > 0.05$ οπότε δεν απορρίπτουμε την μηδενική υπόθεση ότι η επίδραση της εκπαίδευσης δεν διαφέρει μεταξύ διαφορετικών επιπέδων συχνότητας αγοράς τροφίμων.

Για το επίπεδο εκπαίδευσης 3 σε σχέση με το επίπεδο 2 παίρνουμε $p\text{-value}=0.5594 > 0.05$ οπότε δεν απορρίπτουμε την μηδενική υπόθεση ότι η επίδραση της εκπαίδευσης δεν διαφέρει μεταξύ διαφορετικών επιπέδων συχνότητας αγοράς τροφίμων.

Για το επίπεδο εκπαίδευσης 4 σε σχέση με το επίπεδο 3 παίρνουμε $p\text{-value}=0.2952 > 0.05$ οπότε δεν απορρίπτουμε την μηδενική υπόθεση ότι η επίδραση της εκπαίδευσης δεν διαφέρει μεταξύ διαφορετικών επιπέδων συχνότητας αγοράς τροφίμων.

Συγκρίνοντας τα αποτελέσματα των t-test για τα 2 υποδείγματα με και χωρίς αλληλεπίδραση παρατηρούμε ότι τόσο για το 2^ο όσο και για το 3^ο εκπαιδευτικό επίπεδο η επίδραση της εκπαίδευσης στην πιθανότητα ένα άτομο να διαβάζει τις ετικέτες τροφίμων δεν μεταβάλλεται καθώς μεταβάλλεται η συχνότητα αγοράς τροφίμων. Οπότε τα δύο υποδείγματα έδωσαν ίδιο αποτέλεσμα όσον αφορά τις οριακές επιδράσεις για το 2^ο και το 3^ο εκπαιδευτικό επίπεδο. Ωστόσο, για το 4^ο εκπαιδευτικό επίπεδο τα αποτελέσματα μεταξύ του υποδείγματος χωρίς αλληλεπίδραση και με αλληλεπίδραση διαφέρουν.

- ❖ Παράδειγμα 4: Οριακή επίδραση του εάν κάποιος ακολουθεί μια συγκεκριμένη διαίτα (diet) καθώς μεταβάλλεται η σημασία της διατροφικής αξίας (nutrition)

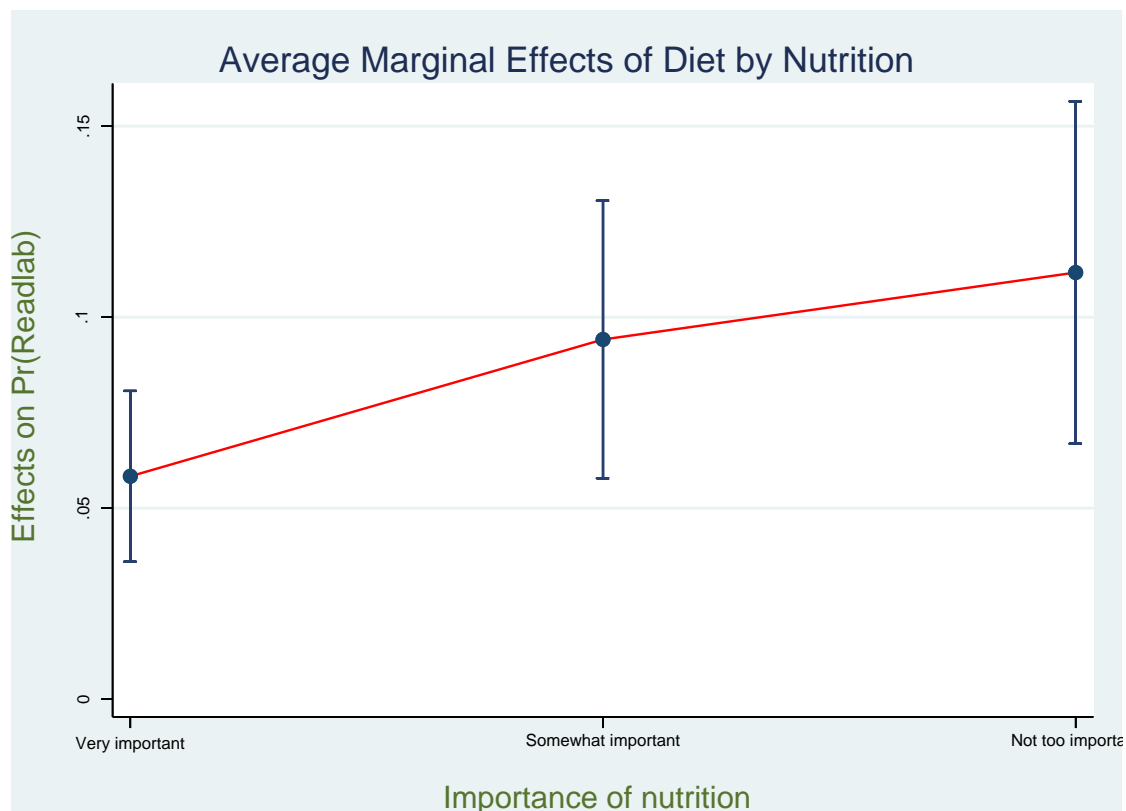
Αρχικά, υπολογίζουμε τις οριακές επιδράσεις της μεταβλητής diet καθώς μεταβάλλεται η σημασία της διατροφικής αξίας. Επειδή η μεταβλητή nutrition είναι κατηγορική θα υπολογίσουμε την οριακή επίδραση της μεταβλητής diet για κάθε επίπεδο της μεταβλητής nutrition.

Για το υπόδειγμα χωρίς αλληλεπίδραση δημιουργούμε αρχικά το διάγραμμα διασποράς των οριακών επιδράσεων και στην συνέχεια κάνουμε το ίδιο για το υπόδειγμα με αλληλεπίδραση και τα συγκρίνουμε γραφικά.

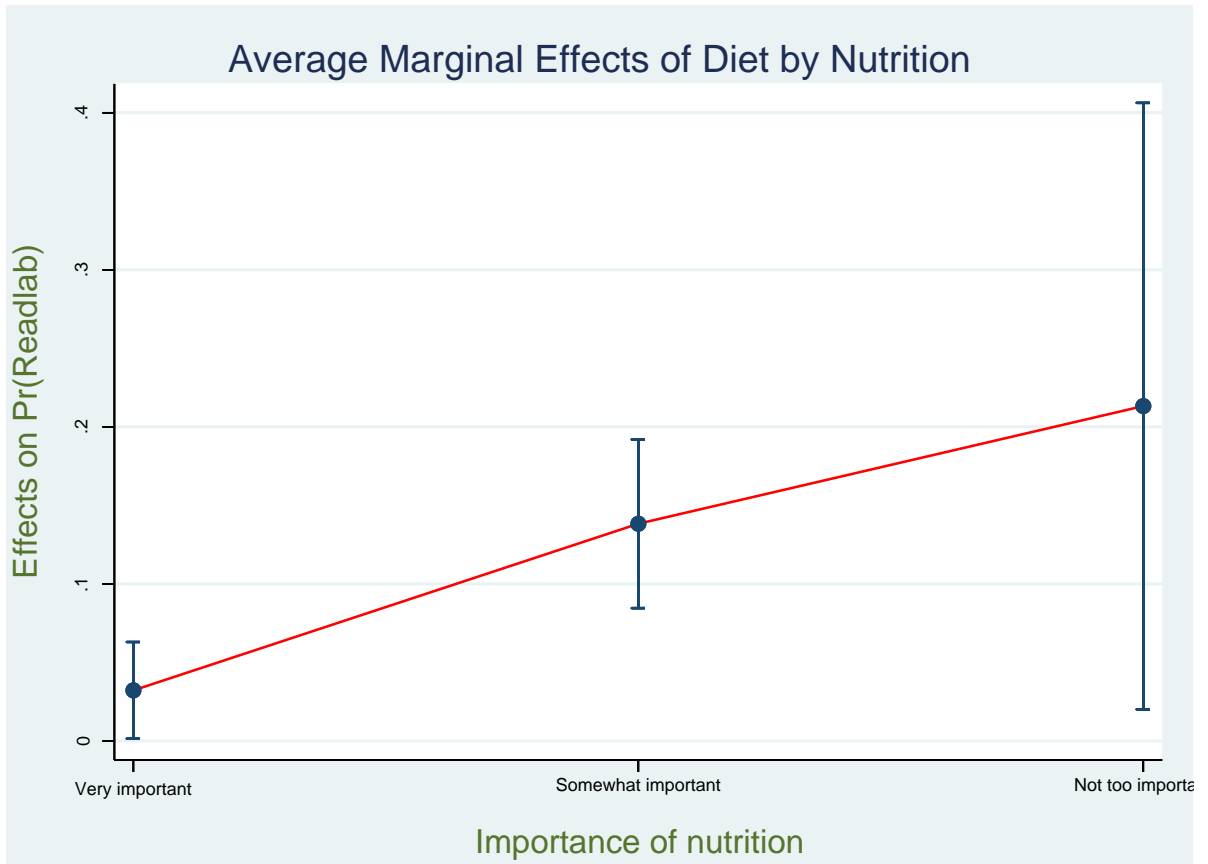
Από το γράφημα της Εικόνας 17 για το υπόδειγμα χωρίς αλληλεπίδραση, για κάθε επίπεδο της μεταβλητής nutrition ένα άτομο που ακολουθεί μια συγκεκριμένη διαίτα έχει μεγαλύτερη πιθανότητα να διαβάσει τις ετικέτες τροφίμων σε σχέση με ένα άτομο με τα ίδια χαρακτηριστικά αλλά που δεν ακολουθεί κάποια συγκεκριμένη διαίτα. Επιπλέον, βλέπουμε πως ένα άτομο που ακολουθεί μια συγκεκριμένη διαίτα σε σχέση με ένα που δεν ακολουθεί εμφανίζει μεγαλύτερη πιθανότητα να διαβάσει τις ετικέτες τροφίμων όσο μειώνεται η σημασία της διατροφικής αξίας. Επιπλέον, παρατηρούμε από τα διαστήματα εμπιστοσύνης ότι όλες οι οριακές επιδράσεις είναι στατιστικά σημαντικές .

Από το γράφημα της Εικόνας 18, οριακών επιδράσεων της μεταβλητής diet καθώς μεταβάλλεται η σημασία της διατροφικής αξίας, για το υπόδειγμα με περιορισμό, παρατηρούμε σχεδόν παρόμοια αποτελέσματα γραφικά σε σύγκριση με εκείνα του υποδείγματος χωρίς αλληλεπίδραση.

Εικόνα 17: Οριακές επιδράσεις του εάν κάποιος ακολουθεί μια συγκεκριμένη διαίτα (diet) καθώς μεταβάλλεται η σημασία της διατροφικής αξίας (nutrition) για το υπόδειγμα χωρίς αλληλεπίδραση



Εικόνα 18: Οριακές επιδράσεις του εάν κάποιος ακολουθεί μια συγκεκριμένη διαίτα (diet) καθώς μεταβάλλεται η σημασία της διατροφικής αξίας (nutrition) για το υπόδειγμα με αλληλεπίδραση



Στην συνέχεια, ελέγχουμε με ένα t-test το αν η οριακή επίδραση της μεταβλητής diet για το πρώτο επίπεδο της μεταβλητής nutrition μπορεί να είναι η ίδια με την οριακή επίδραση της μεταβλητής diet για το δεύτερο και το τρίτο επίπεδο της μεταβλητής nutrition

H_0 : Η οριακή της μεταβλητής diet δεν διαφέρει μεταξύ των διαφορετικών επιπέδων της μεταβλητής nutrition

H_1 : Η οριακή της μεταβλητής diet διαφέρει μεταξύ των διαφορετικών επιπέδων της μεταβλητής nutrition

Για το υπόδειγμα χωρίς αλληλεπίδραση:

Τα αποτελέσματα του t-test φαίνονται παρακάτω:

Πίνακας 14: Αποτελέσματα t-test για την μεταβλητή Diet και Nutrition χωρίς αλληλεπίδραση

chi2 = 20.62
p-value = 0.0000

Από τον Πίνακα 14 βλέπουμε ότι **p-value=0.0000 < 0.05** οπότε απορρίπτουμε την μηδενική υπόθεση ότι η επίδραση της μεταβλητής diet δεν διαφέρει μεταξύ των διαφορετικών επιπέδων της μεταβλητής nutrition.

Για το υποδείγμα με αλληλεπίδραση:

Τα αποτελέσματα του t-test φαίνονται παρακάτω:

Πίνακας 15: Αποτελέσματα t-test για την μεταβλητή Diet και Nutrition χωρίς αλληλεπίδραση

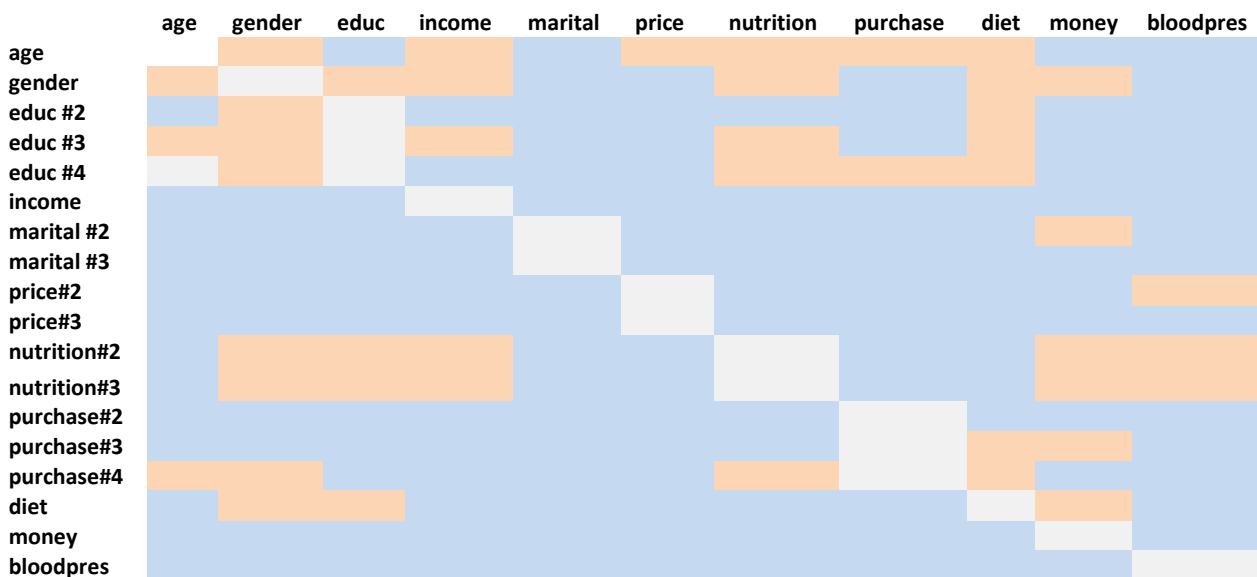
chi2 = 13.92
p-value = 0.0010

Από τον Πίνακα 15 βλέπουμε ότι **p-value=0.001 < 0.05** οπότε απορρίπτουμε την μηδενική υπόθεση ότι η επίδραση της μεταβλητής diet δεν διαφέρει μεταξύ των διαφορετικών επιπέδων της μεταβλητής nutrition.

Οπότε τελικά, παρατηρούμε ότι και στα 2 υποδείγματα, με και χωρίς περιορισμό, η επίδραση της μεταβλητής diet διαφέρει καθώς μεταβάλλεται η μεταβλητή nutrition, κάτι που συμφωνεί και με αυτό που υποψιαστήκαμε και από τα διαγράμματα διασποράς.

Στην συνέχεια, προκειμένου να συνοψίσουμε τα αποτελέσματα όλων των συγκρίσεων των υποδειγμάτων δημιουργήσαμε τον παρακάτω πίνακα:

Πίνακας 16: Σύγκριση οριακών επιδράσεων μεταξύ των υποδειγμάτων με και χωρίς αλληλεπίδραση



- Τα δύο υποδείγματα κατέληξαν σε διαφορετική επίδραση στην εξαρτημένη μεταβλητή
- Τα δύο υποδείγματα κατέληξαν σε ίδια επίδραση στην εξαρτημένη μεταβλητή
- Δεν υπάρχουν μετρήσεις για αυτά τα υποδείγματα

% περιπτώσεων όπου τα 2 συγκρινόμενα υποδείγματα έδειξαν διαφορετικές επιδράσεις	0,24
--	-------------

Ο Πίνακας 16 απεικονίζει με μπλε χρώμα τις περιπτώσεις όπου από τα t-test που περιγράφηκαν προηγουμένως προέκυψε ότι ένα υπόδειγμα χωρίς αλληλεπίδραση και ένα υπόδειγμα με αλληλεπίδραση οδηγεί σε ίδιο συμπέρασμα σχετικά με τις οριακές επιδράσεις των ανεξάρτητων μεταβλητών. Το πορτοκαλί χρώμα του πίνακα αναπαριστά τις περιπτώσεις όπου τα δύο υποδείγματα, χωρίς και με αλληλεπίδραση, οδήγησαν σε διαφορετικές οριακές επιδράσεις των ανεξάρτητων μεταβλητών. Με γκρι χρώμα απεικονίζονται οι περιπτώσεις όπου δεν υπήρχαν μετρήσεις διότι αφορά σε αλληλεπίδραση μιας μεταβλητής με τον εαυτό της. Από τον πίνακα αυτόν, μπορούμε να εξάγουμε το ποσοστό των περιπτώσεων όπου τα δύο συγκρινόμενα υποδείγματα κατέληξαν σε διαφορετικές οριακές επιδράσεις. Οι μετρήσεις που είχαμε ήταν στο σύνολο 198 από τις οποίες οι 44 οδηγούν σε διαφορετικό συμπέρασμα. Έτσι το ποσοστό των περιπτώσεων όπου το υπόδειγμα χωρίς αλληλεπίδραση οδηγεί σε διαφορετικό πόρισμα για τις οριακές επιδράσεις των μεταβλητών από το υπόδειγμα με αλληλεπίδραση είναι **24%**.

7.3 Όροι αλληλεπίδρασης και εσφαλμένα συμπεράσματα

Οι ερευνητές πολύ συχνά όταν θέλουν να λάβουν υπόψη την μη-γραμμικότητα σε ένα υπόδειγμα παλινδρόμησης εισάγουν σε αυτό όρους αλληλεπίδρασης. Στο 3^ο μέρος της ανάλυσης μας θέλουμε να διαπιστώσουμε σε ποιες περιπτώσεις η εισαγωγή ενός όρου αλληλεπίδρασης στο υπόδειγμα μας οδηγεί σε εσφαλμένα συμπεράσματα σχετικά με τον τρόπο που επιδρούν οι ανεξάρτητες μεταβλητές επάνω στην πιθανότητα ένα άτομο να διαβάσει τις ετικέτες τροφίμων. Προκειμένου να εξάγουμε το ποσοστό των περιπτώσεων αυτών ακολουθήσαμε την παρακάτω διαδικασία:

Από τους Πίνακες 7 και 16 προκύπτουν 4 διαφορετικοί τύποι συμπερασμάτων για την ορθότητα των συμπερασμάτων που εξάγονται από τις οριακές επιδράσεις των ανεξάρτητων μεταβλητών. Οι τέσσερις αυτοί τύποι συμπερασμάτων είναι οι ακόλουθοι:

- Όταν στον Πίνακα 7 έχουμε μπλε χρώμα και στον Πίνακα 16 έχουμε επίσης μπλε χρώμα. Αυτό σημαίνει ότι από τον έλεγχο του λόγου πιθανοφάνειας προέκυψε ότι το υπόδειγμα χωρίς αλληλεπίδραση προσαρμόζεται καλύτερα στα δεδομένα και από τους ελέγχους με την στατιστική t φαίνεται ότι με και χωρίς όρο αλληλεπίδρασης σε ένα υπόδειγμα οδηγούμαστε στο ίδιο συμπέρασμα για τις οριακές επιδράσεις των μεταβλητών. Οπότε θα μπορούσαμε να πούμε ότι σε αυτήν την περίπτωση ένας ερευνητής θα ήταν αδιάφορος για το αν χρησιμοποιεί ένα υπόδειγμα με ή χωρίς όρο αλληλεπίδρασης. Δηλαδή, είτε χρησιμοποιήσει το υπόδειγμα με αλληλεπίδραση είτε αυτό χωρίς, οδηγείται στο ίδιο συμπέρασμα οπότε και **δεν οδηγούμαστε σε εσφαλμένο συμπέρασμα**.
- Όταν στον Πίνακα 7 έχουμε μπλε χρώμα και στον Πίνακα 16 έχουμε ροζ χρώμα. Αυτό σημαίνει ότι από τον έλεγχο του λόγου πιθανοφάνειας προέκυψε ότι το

υπόδειγμα χωρίς αλληλεπίδραση προσαρμόζεται καλύτερα στα δεδομένα και από τους ελέγχους με την στατιστική t φαίνεται ότι εάν χρησιμοποιήσουμε όρο αλληλεπίδρασης οδηγούμαστε σε διαφορετικό συμπέρασμα για τον τρόπο που επιδρούν οι μεταβλητές επάνω στην εξαρτημένη. Δηλαδή, σε αυτή την περίπτωση εάν ένας ερευνητής χρησιμοποιούσε το υπόδειγμα με αλληλεπίδραση θα οδηγούνταν σε **εσφαλμένο συμπέρασμα** και θα έπρεπε να είχε χρησιμοποιήσει ένα υπόδειγμα χωρίς αλληλεπίδραση.

- Όταν στον Πίνακα 7 έχουμε ροζ χρώμα και στον Πίνακα 16 έχουμε επίσης ροζ χρώμα. Αυτό σημαίνει ότι από τον έλεγχο του λόγου πιθανοφάνειας προέκυψε ότι το υπόδειγμα με αλληλεπίδραση προσαρμόζεται καλύτερα στα δεδομένα και από τους ελέγχους με την στατιστική t φαίνεται ότι τα 2 υποδείγματα, με και χωρίς αλληλεπίδραση, οδηγούν σε διαφορετικό συμπέρασμα για τις οριακές επιδράσεις των μεταβλητών. Οπότε θα μπορούσαμε να πούμε ότι σε αυτήν την περίπτωση ένας ερευνητής που χρησιμοποιήσει ένα υπόδειγμα με αλληλεπίδραση έπραξε ορθά και **δεν οδηγείται σε εσφαλμένο συμπέρασμα**, εφόσον αυτό το υπόδειγμα προσαρμόζεται καλύτερα στα δεδομένα και αν δεν είχε συμπεριλάβει αλληλεπίδραση θα οδηγούνταν σε διαφορετικό συμπέρασμα. Μάλιστα, σε αυτή την περίπτωση θα οδηγούμασταν σε εσφαλμένο συμπέρασμα εάν είχαμε χρησιμοποιήσει το υπόδειγμα χωρίς αλληλεπίδραση.
- Όταν στον Πίνακα 7 έχουμε ροζ χρώμα και στον Πίνακα 16 έχουμε μπλε χρώμα. Αυτό σημαίνει ότι από τον έλεγχο του λόγου πιθανοφάνειας προέκυψε ότι το υπόδειγμα με αλληλεπίδραση προσαρμόζεται καλύτερα στα δεδομένα και από τους ελέγχους με την στατιστική t φαίνεται ότι με και χωρίς όρο αλληλεπίδρασης σε ένα υπόδειγμα οδηγούμαστε στο ίδιο συμπέρασμα για τις οριακές επιδράσεις των μεταβλητών. Αυτό σημαίνει ότι ένα ερευνητής έπρεπε να χρησιμοποιήσει το υπόδειγμα με αλληλεπίδραση αφού δείχνει καλύτερη προσαρμογή αλλά και να μην το είχε κάνει δεν **θα οδηγούνταν σε εσφαλμένο συμπέρασμα** εφόσον τα 2 υποδείγματα καταλήγουν σε ίδιες επιδράσεις. Ακόμη και, να είχε χρησιμοποιήσει το υπόδειγμα χωρίς αλληλεπίδραση δεν θα οδηγούνταν σε λάθος συμπέρασμα οπότε θα ήταν αδιάφορος μεταξύ των δύο υποδειγμάτων.

Από τους 4 παραπάνω τύπους συμπερασμάτων για τις οριακές επιδράσεις, άρα και για την ερμηνεία των επιδράσεων των ανεξάρτητων μεταβλητών, παρατηρούμε μόνο σε μία περίπτωση οδηγούμαστε σε εσφαλμένο συμπέρασμα όταν χρησιμοποιούμε το υπόδειγμα με όρο αλληλεπίδρασης. Αυτή είναι η περίπτωση στον Πίνακα 7 να έχουμε μπλε χρώμα και στον Πίνακα 16 να έχουμε ροζ χρώμα. Σε όλες τις υπόλοιπες περιπτώσεις δεν παρατηρούμε σφάλμα. Έτσι, θα δημιουργήσουμε ένα πίνακα που να συνοψίζει αυτά τα αποτελέσματα απεικονίζοντας με κόκκινο χρώμα την περίπτωση σφάλματος και με πράσινο χρώμα την περίπτωση μη-σφάλματος.

8. Συμπεράσματα

Κύριος σκοπός της παρούσας πτυχιακής μελέτης ήταν να μελετηθούν οι μη-γραμμικές επιδράσεις σε γραμμικά και μη-γραμμικά υποδείγματα παλινδρόμησης. Ένα από τα κύρια ερωτήματα αφορά στο κατά πόσο ένας ερευνητής καταλήγει σε σωστά συμπεράσματα για τις οριακές επιδράσεις των ανεξάρτητων μεταβλητών όταν, περιλαμβάνει σε ένα υπόδειγμα διχοτομικής απόκρισης όρους αλληλεπίδρασης. Προκειμένου να μελετήσουμε την έννοια της μη-γραμμικότητας χρησιμοποιήσαμε δευτερογενή δεδομένα που προέρχονται από τον τομέα τον τροφίμων και αφορούν στο εάν κάποιος διαβάζει ή όχι τις ετικέτες των τροφίμων που αγοράζει. Στο κεφάλαιο αυτό θα παρατεθούν οι παρατηρήσεις και τα συμπεράσματα από την ανάλυση μας.

Αρχικά, όσον αφορά την σύγκριση των υποδειγμάτων με και χωρίς όρους αλληλεπίδρασης παρατηρήσαμε ότι με βάση τα τεστ του λόγου της πιθανοφάνειας στο 96% των περιπτώσεων, το υπόδειγμα χωρίς αλληλεπίδραση προσαρμοζόταν καλύτερα στα δεδομένα κάτι που υποδηλώνει ότι η προσθήκη όρων αλληλεπίδρασης δεν προσέθεσε τίποτα στην ερμηνευτική ικανότητα του υποδείγματος.

Επιπροσθέτως, σχετικά με την έννοια της μη-γραμμικότητας, όπως μας υπέδειξε και η θεωρία, διαπιστώσαμε μέσα από τον υπολογισμό των οριακών επιδράσεων μιας μεταβλητής όταν μια άλλη μεταβλητή μεταβάλλεται, ότι ακόμη και σε γραμμικά υποδείγματα ως προς τις ανεξάρτητες μεταβλητές υπάρχει αλληλεπίδραση μεταξύ των μεταβλητών. Δηλαδή, παρόλο που δεν υπήρχε κάποιος συντελεστής αλληλεπίδρασης οι μεταβλητές έδειξαν μια μη-γραμμική επίδραση επάνω στην εξαρτημένη μας μεταβλητή.

Επιπλέον, αναφορικά με τις οριακές επιδράσεις των ανεξάρτητων μεταβλητών από τις μετρήσεις που πήραμε μέσω των t-test παρατηρήσαμε ότι σε πολλές περιπτώσεις η επίδραση διέφερε μεταξύ των δύο υποδειγμάτων, με και χωρίς αλληλεπίδραση. Αυτό μας δείχνει ότι, ανάλογα με το υπόδειγμα που θα επιλέξουμε θα οδηγηθούμε σε διαφορετικό πόρισμα για τον τρόπο επίδρασης των ανεξάρτητων μεταβλητών. Το ποσοστό των περιπτώσεων αυτών υπολογίστηκε να είναι 24%.

Ακόμη, καταλήξαμε στον υπολογισμό του ποσοστού ορθών προβλέψεων του υποδείγματος σχετικά με τις οριακές επιδράσεις διαχωρίζοντας τις περιπτώσεις όπου θα οδηγούμασταν σε σωστή εκτίμηση των οριακών επιδράσεων, όταν χρησιμοποιούμε ένα υπόδειγμα με αλληλεπίδραση, από αυτές όπου θα καταλήγαμε σε σφάλμα. Το ποσοστό των περιπτώσεων όπου καταλήγουμε σε εσφαλμένο συμπέρασμα υπολογίστηκε σε 20,2%. Ένα ποσοστό που είναι αρκετά μεγάλο όταν μιλάμε για σφάλμα στις μετρήσεις μας. Κάτι τέτοιο σημαίνει στο 20,2% των περιπτώσεων θα καταλήξουμε σε λανθασμένη ερμηνεία των επιδράσεων των μεταβλητών εάν εισάγουμε όρο αλληλεπίδρασης στο υπόδειγμα μας.

Κλείνοντας, από τα αποτελέσματα και της έρευνας μας, γίνεται εμφανές, ότι είναι σημαντικό για έναν ερευνητή να αναλογιστεί καλύτερα κατά πόσο πρέπει να εισάγει όρους αλληλεπίδρασης σε ένα υπόδειγμα διακριτής απόκρισης προκειμένου να συμπεριλάβει μη-γραμμικές επιδράσεις των μεταβλητών. Εκτός από το, ότι μη-γραμμικές επιδράσεις υφίστανται ακόμη και σε γραμμικά υποδείγματα ως προς τις ανεξάρτητες μεταβλητές, είναι

πολύ πιθανό να καταλήξει, ο ερευνητής, σε λανθασμένες εκτιμήσεις και ερμηνεία της επίδρασης των ανεξάρτητων μεταβλητών. Φυσικά, είναι επιτακτική ανάγκη να γίνουν περαιτέρω μελέτες, με περισσότερα δεδομένα και μεταβλητές, έτσι ώστε να διαπιστωθούν αυτά τα συμπεράσματα και για πιο γενικές περιπτώσεις. Κάτι τέτοιο θα μπορούσε να υλοποιηθεί με μια μελέτη προσομοίωσης τέτοιων υποδειγμάτων με σκοπό την επαλήθευση των πορισμάτων και την εξαγωγή πιο έγκυρων αποτελεσμάτων.

9. Βιβλιογραφία

- Agresti, A. (1996). *An Introduction to Categorical Data Analysis*. John Wiley and Sons, New York, 372 pages.
- Ai, C., and E.C. Norton (2003). Interaction terms in logit and probit models. *Economics Letters* 80(1), 123-129
- Bender, M.M., and Derby, B.M. (1992). Prevalence of reading nutrition and ingredient information on food labels among adult Americans: 1982-1988. *Journal of Nutrition Education*, 24, 292-297.
- Best, H. & C., Wolf, (2015). *The SAGE Handbook of Regression Analysis and Causal Inference* First edition 424 pages.
- Burton, S., and Andrews, J.C. (1996). Age, product nutrition, and label format effects on consumer perceptions and product evaluations. *The Journal of Consumer Affairs*, 30(1), 68-89.
- Collett, D. (2003). *Modelling Binary Data*, 2nd ed. Chapman & Hall, London, 344 p.
- Drichoutis, A.ndreas, (2011). "Interpreting interaction terms in linear and non-linear models: A cautionary tale," MPRA Paper 33251, University Library of Munich, Germany.
- Drichoutis, A.C., Lazaridis, P., and Nayga, R.M., Jr. (2005). Nutrition knowledge and consumer use of nutritional food labels. *European Review of Agricultural Economics*, 32(1), 93-118.
- Drichoutis, A. C., Lazaridis, P., and Nayga, R.M., Jr. (2007). An assessment of product class involvement in food purchasing behaviour. *European Journal of Marketing*, 40(forthcoming).
- Ghani, W.I., and N.M. Childs. (1999). Wealth Effects of the Passage of the Nutrition Labeling and Education Act of 1990 for Large U.S. Multinational Food Corporations. *Journal of Public Policy & Marketing*, 18: 147–158.
- Greene, W. (2012) *Econometric Analysis*. 7th Edition, Prentice Hall, Upper Saddle River.
- Greenland, S., Senn, S.J., Rothman, K.J., Carlin, J.B., Poole, C., Goodman, S.N. & D.G. Altman (2016). Statistical tests, *P* values, confidence intervals, and power: a guide to misinterpretations. *Eur J Epidemiol* **31**, 337–350.
- Hosmer, D.W. & S. Lemeshow. (2000). *Applied Logistic Regression* (Wiley Series in Probability and Statistics) 2nd Edition. Wiley Series in Probability & Statistics. ISBN-13: 978-0471356325, 397 Pages.
- Hosmer, D.W., Lemeshow, S., & R.X. Sturdivant (2013). *Applied Logistic Regression*, 3rd Edition. Wiley Series in Probability & Statistics, ISBN: 978-0-470-58247, 528 Pages.

Long, J.S. (1997). *Regression Models for Categorical and Limited Dependent Variables*, Volume: 7, Series: *Advanced Quantitative Techniques in the Social Sciences*, 1997 | 328 pages | SAGE Publications, Inc.

McLean-Meyinsse, P. E. (2001). An analysis of nutritional label use in the Southern United States. *Journal of Food Distribution Research*, 32(1), 110-114.

Nayga, R. M., Jr. (1996). Determinants of consumers' use of nutritional information on food packages. *Journal of Agricultural and Applied Economics*, 28(2), 303-312.

Nayga, R. M., Jr. (2000). Nutrition knowledge, gender, and food label use. *The Journal of Consumer Affairs*, 34(1), 97-112.

Wijngaart, A. (2002). Nutrition labelling: purpose, scientific issues and challenges. *Asia Pacific Journal of Clinical Nutrition*; 11(2): S68–S71.

ΠΑΡΑΡΤΗΜΑ

Στο παράρτημα της πτυχιακής μας μελέτης παραθέτουμε ένα μέρος του κώδικα που χρησιμοποιήσαμε για την οικονομετρική ανάλυση με το Στατιστικό Πρόγραμμα Stata.

**Εισαγωγή και επεξεργασία δεδομένων*

```
use demo.dta, replace

merge 1:1 seqn using foodlabel.dta
rename _merge matching
merge 1:1 seqn using blood.dta
rename _merge matching1
merge 1:1 seqn using weight.dta
rename _merge matching2
merge 1:1 seqn using income.dta
rename _merge matching3
merge 1:1 seqn using behavior.dta
rename _merge matching4
merge 1:1 seqn using nutrition.dta

rename riagendr gender
rename ridageyr age
rename dmdeduc2 education
rename ind235 income
rename dmdmart1 marstatus
rename cbq660 prc
rename cbq665 nut
rename whq070 diet
rename cbq140 purchase
rename bpq020 bloodpres
rename cbd070 money
rename dbq750 label

foreach x of varlist gender age education income organic household
marstatus prc nut diet purchase bloodpres chol weight money allergies
planner label {
tab `x'
}
*ΑΝΑΚΩΔΙΚΟΠΟΙΗΣΗ ΤΩΝ ΜΕΤΑΒΛΗΤΩΝ
*Μετατροπή της εξαρτημένης μεταβλητής σε διχοτομική
recode label(1/3=0) (4/5=1) , generate (readlab)

recode marstatus (1=1)(2/4=2)(5=3)(6=1), generate (marital)

recode nut(1=1) (2=2) (3/4=3), generate (nutrition)

recode education(1/2=1) (3=2)(4=3)(5=4), generate (educ)

recode prc (1=1) (2=2) (3/4=3), generate (price)

replace marital=. if marital==99 | marital==77
replace purchase=. if purchase==99 | purchase==77
replace price=. if price==9 | price==5
replace nutrition=. if nutrition==9
replace money=. if money==777777 | money==999999 | money==171728
replace income=. if income==77 | income==99
replace diet=. if diet==7 | diet==9
replace bloodpres=. if bloodpres==9
```

```

replace purchase=. if purchase==77 | purchase==99 | purchase==5 |
purchase==6
replace educ=. if educ==7 | educ==9
replace readlab=. if readlab==7 | readlab==9 | readlab==6

```

```

replace gender=0 if gender==1
replace gender=1 if gender==2
replace diet=0 if diet==1
replace diet=1 if diet==2
replace bloodpres=0 if bloodpres==1
replace bloodpres=1 if bloodpres==2

```

*Προσθήκη Ετικετιών

```

label define var 0 "Yes" 1 "No"
label values bloodpres var
label values gender var
label values diet var

```

```

label define var1 1 "Married-Living with partner" 2 "Widowed-
Divorced-Separated" 3 "Never married"
label values marital var1

```

```

label define var2 1 "Less Than 11th Grade" 2 "High School Grad/GED or
Equivalent" 3 "Some College or AA degree" 4 "College Graduate or
above"
label values educ var2

```

```

label define var3 1 "Very important" 2 "Somewhat important" 3 "Not
too important"
label values price var3
label values nutrition var3

```

```

label define var4 1 "More than once a week" 2 "Once a week" 3 "Once
every two weeks" 4 "Once a month or less"
label values purchase var4

```

```

label define var5 0 "Sotimes-Always" 1 "Rarely-Never"
label values readlab var5

```

*Οικονομετρική Ανάλυση - Εκτίμηση Οριακών Επιδράσεων

*gender-age

```

logit readlab c.age i.gender i.educ c.income i.marital i.price
i.nutrition i.diet i.purchase i.bloodpres c.money
estimates store REG1
sum age if e(sample)==1
margins, dydx(gender) at(age=(20(2)80)) post
marginsplot
test [1.gender]_b[1._at] = [1.gender]_b[9._at]= [1.gender]_b[18._at]=
[1.gender]_b[31._at] // Απορρίπτεται

```

```

logit readlab i.gender##c.age i.educ c.income i.marital i.price
i.nutrition i.diet i.purchase i.bloodpres c.money
estimates store REG2
lrtest REG1 REG2 // Δεν απορρίπτεται
margins, dydx(gender) at(age=(20(2)80)) post
marginsplot
test [1.gender]_b[1._at] = [1.gender]_b[9._at]= [1.gender]_b[18._at]=
[1.gender]_b[31._at] // Δεν απορρίπτεται

```

*gender-educ

```
logit readlab c.age i.gender i.educ c.income i.marital i.price  
i.nutrition i.diet i.purchase i.bloodpres c.money  
estimates store REG1  
margins, dydx(gender) at(educ=(1 2 3 4)) post  
marginsplot  
test [1.gender]_b[1._at] = [1.gender]_b[2._at]= [1.gender]_b[3._at] =  
[1.gender]_b[4._at] // Απορρίπτεται
```

```
logit readlab i.gender##i.educ c.age c.income i.marital i.price  
i.nutrition i.diet i.bloodpres c.money i.purchase  
estimates store REG3  
lrtest REG1 REG3 // Δεν απορρίπτεται  
margins, dydx(gender) at(educ=(1 2 3 4)) post  
marginsplot  
test [1.gender]_b[1._at] = [1.gender]_b[2._at]= [1.gender]_b[3._at] =  
[1.gender]_b[4._at] // Δεν απορρίπτεται
```

*gender-income

```
logit readlab c.age i.gender i.educ c.income i.marital i.price  
i.nutrition i.diet i.purchase i.bloodpres c.money  
estimates store REG1  
margins, dydx(gender) at(income=(1(1)12)) post  
marginsplot  
test [1.gender]_b[1._at]= [1.gender]_b[6._at] = [1.gender]_b[12._at]  
// Απορρίπτεται
```

```
logit readlab i.gender##c.income i.educ c.age i.marital i.price  
i.nutrition i.diet i.bloodpres c.money i.purchase  
estimates store REG4  
lrtest REG1 REG4 // Δεν απορρίπτεται  
margins, dydx(gender) at(income=(1(1)12)) post  
marginsplot  
test [1.gender]_b[1._at]= [1.gender]_b[6._at] = [1.gender]_b[12._at]  
// Δεν απορρίπτεται
```

*gender-marital

```
logit readlab c.age i.gender i.educ c.income i.marital i.price  
i.nutrition i.diet i.purchase i.bloodpres c.money  
estimates store REG1  
margins, dydx(gender) at(marital=(1 2 3)) post  
marginsplot  
test [1.gender]_b[1._at] = [1.gender]_b[2._at] = [1.gender]_b[3._at]  
// Δεν απορρίπτεται
```

```
logit readlab i.gender##i.marital i.educ c.age c.income i.price  
i.nutrition i.diet i.bloodpres c.money i.purchase  
estimates store REG5  
lrtest REG1 REG5 // Δεν απορρίπτεται  
margins, dydx(gender) at(marital=(1 2 3)) post  
marginsplot  
test [1.gender]_b[1._at] = [1.gender]_b[2._at] = [1.gender]_b[3._at]  
Δεν απορρίπτεται
```

*marital-age

```
logit readlab c.age i.gender i.educ c.income i.marital i.price
i.nutrition i.diet i.purchase i.bloodpres c.money
estimates store REG1
margins, dydx(marital) at(age=(16(2)80)) post
marginsplot
matlist e(b)
foreach x of numlist 2/3 {
test [`x'.marital]_b[1._at] = [`x'.marital]_b[9._at]=
[`x'.marital]_b[18._at]= [`x'.marital]_b[31._at]
}
* #2 Δεν απορρίπτεται
* #3 Δεν απορρίπτεται
```

```
logit readlab c.age##i.marital i.educ i.gender c.income i.price
i.nutrition i.diet i.bloodpres c.money i.purchase
estimates store REG42
lrtest REG1 REG42 // den aporriptetai
margins, dydx(marital) at(age=(16(2)80)) post
marginsplot
matlist e(b)
foreach x of numlist 2/3 {
test [`x'.marital]_b[1._at] = [`x'.marital]_b[9._at]=
[`x'.marital]_b[18._at]= [`x'.marital]_b[31._at]
}
* #2 Δεν απορρίπτεται
* #3 Δεν απορρίπτεται
```

*marital-gender

```
logit readlab c.age i.gender i.educ c.income i.marital i.price
i.nutrition i.diet i.purchase i.bloodpres c.money
estimates store REG1
margins, dydx(marital) at(gender=(0 1)) post
marginsplot
matlist e(b)
foreach x of numlist 2/3 {
test [`x'.marital]_b[1._at] = [`x'.marital]_b[2._at]
}
* #2 Δεν απορρίπτεται
* #3 Δεν απορρίπτεται
```