



ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ

ΕΡΕΥΝΑ ΤΗΣ ΕΠΙΔΡΑΣΗΣ ΠΕΡΙΒΑΛΛΟΝΤΙΚΩΝ ΜΕΤΑΒΛΗΤΩΝ ΣΤΗΝ ΚΑΤΑΝΟΜΗ ΕΙΔΩΝ ΤΗΣ ΧΛΩΡΙΔΑΣ ΤΟΥ ΟΡΟΥΣ ΕΡΥΜΑΝΘΟΥ ΜΕ ΤΗΝ ΧΡΗΣΗ ΤΟΥ ΜΟΝΤΕΛΟΥ PROPORTIONAL ODDS

Σωτήριος Γ. Βαλμάς

ΕΡΓΑΣΙΑ

Που υποβλήθηκε στο Τμήμα Στατιστικής
του Οικονομικού Πανεπιστημίου Αθηνών
ως μέρος των απαιτήσεων για την απόκτηση

Μεταπτυχιακού Διπλώματος

Συμπληρωματικής Ειδίκευσης στη Στατιστική
Μερικής Παρακολούθησης (Part-time)

Αθήνα
Δεκέμβριος 2005

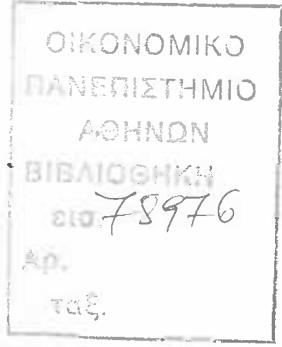


ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΟ ΑΘΗΝΩΝ
ΚΑΤΑΛΟΓΟΣ



0 000000 559584





ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ

**ΕΡΕΥΝΑ ΤΗΣ ΕΠΙΔΡΑΣΗΣ ΠΕΡΙΒΑΛΛΟΝΤΙΚΩΝ
ΜΕΤΑΒΛΗΤΩΝ ΣΤΗΝ ΚΑΤΑΝΟΜΗ ΕΙΔΩΝ ΤΗΣ
ΧΛΩΡΙΔΑΣ ΤΟΥ ΟΡΟΥΣ ΕΡΥΜΑΝΘΟΥ ΜΕ ΤΗΝ
ΧΡΗΣΗ ΤΟΥ ΜΟΝΤΕΛΟΥ PROPORTIONAL ODDS.**

Σωτήριος Γ. Βαλμάς

ΕΡΓΑΣΙΑ

Που υποβλήθηκε στο Τμήμα Στατιστικής
του Οικονομικού Πανεπιστημίου Αθηνών
ως μέρος των απαιτήσεων για την απόκτηση



Μεταπτυχιακού Διπλώματος

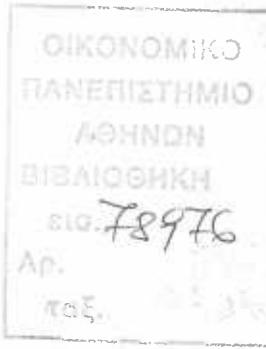
Συμπληρωματικής Ειδίκευσης στη Στατιστική

Μερικής Παρακολούθησης (Part-time)

Αθήνα

Νοέμβριος 2005





ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ

Εργασία που υποβλήθηκε ως μέρος των απαιτήσεων για την απόκτηση
Μεταπτυχιακού Διπλώματος Συμπληρωματικής Ειδίκευσης στη Στατιστική
Μερικής Παρακολούθησης (Part-time)

ΕΡΕΥΝΑ ΤΗΣ ΕΠΙΛΑΡΑΣΗΣ ΠΕΡΙΒΑΛΛΟΝΤΙΚΩΝ ΜΕΤΑΒΛΗΤΩΝ ΣΤΗΝ ΚΑΤΑΝΟΜΗ ΕΙΔΩΝ ΤΗΣ ΧΛΩΡΙΔΑΣ ΤΟΥ ΟΡΟΥΣ ΕΡΥΜΑΝΘΟΥ ΜΕ ΤΗΝ ΧΡΗΣΗ ΤΟΥ ΜΟΝΤΕΛΟΥ PROPORTIONAL ODDS

Σωτήριος Γ. Βαλμάς

Υπεύθυνο μέλος ΔΕΠ:

I. Παπαγεωργίου

Λέκτορας



Ο Διευθυντής Μεταπτυχιακών Σπουδών

Μιχαήλ Λαζάνης
Καθηγητής



ΑΦΙΕΡΩΣΗ

Στη Γωγώ.



ΕΥΧΑΡΙΣΤΙΕΣ

Με την ολοκλήρωση της εργασίας μου θα ήθελα να ευχαριστήσω όλους όσους συνέβαλαν σ' αυτό το αποτέλεσμα.

Την Λέκτορα του Οικονομικού Πανεπιστήμιου Αθηνών, κ. Παπαγεωργίου Ιουλία. Που ως επιβλέποντας της εργασίας, με το ενεργό ενδιαφέρον, της υποδείξεις και της χρήσιμες παρατηρήσεις, ήταν εκείνη που βοήθησε στο να αναδεικνύεται στην διατριβή μου μια επιστημονική επάρκεια και την ευχαριστώ θερμά.

Επίσης ευχαριστώ τον φίλο και συνάδελφο Ph.D. κ. Μαρούλη Γεώργιο για την διάθεση των στοιχείων και την συμμετοχή του στην επιλογή του θέματος. Η βοήθεια που μου προσέφερε όμως ήταν κάθετη και ολόπλευρη καθ' όλη την διάρκεια της έρευνας και συγγραφής της εργασίας μου. Οι υποδείξεις του οποιαδήποτε στιγμή και αν τις χρειαζόμουν ήταν που με έκαναν να αντιληφθώ το ενδιαφέρον και τις δυνατότητες που έκρυβαν τα φαινομενικά ψυχρά αριθμητικά δεδομένα ως προς την φυτοκοινονιολογία και τους φυσικούς και περιβαλλοντικούς παράγοντες που επικρατούν και διέπουν τον ορεινό όγκο του όρους Ερυμάνθου.

Θα ήθελα ακόμη να ευχαριστήσω τον φίλο και μεταπτυχιακό συμφοιτητή μου κ. Σκολαρίκο Μαρίνο όπου με τις χρήσιμες επισημάνσεις του κατάφερα να παρουσιάσω μια άρτια μορφοποιημένη εργασία.

Τέλος εκφράζω τις βαθύτατες ευχαριστίες στην οικογένεια μου για την υπομονή, και την ενθάρρυνση τους ώστε να ολοκληρωθεί και η συγκεκριμένη φάσης των σπουδών μου.

Οφείλω όμως πάρα πολλά στην συμπαράσταση της αρραβωνιαστικιάς μου Μπάκα Γεωργίας και την ευχαριστώ για την ανοχή και αντοχή της.

ΒΙΟΓΡΑΦΙΚΟ ΣΗΜΕΙΩΜΑ

Γεννήθηκα 1972 στο Νεοχώρι - Ηεσπιάν - Ηηβών. Σπούδασα στο Μαθηματικό τμήμα του Πανεπιστήμιου Πατρών την περίοδο 1989-1994.

Μετά την ολοκλήρωση των στρατιωτικών μου υποχρεώσεων εργάσθηκα ως ωρομίσθιος εκπαιδευτής στα δημόσια I.E.K Ηήβας διδάσκοντας το μάθημα της Στατιστικής την περίοδο 1997-1999.

Από το 1998 έως 2001 εργάσθηκα σε Φροντιστήρια Μ.Ε. διδάσκοντας Μαθηματικά, σε μαθητές Γυμνασίου Λυκείου και υποψηφίους ανωτέρων και ανωτάτων σχολών.

Από το 2001 έως το 2004 εργάσθηκα παραδίδοντας κατ'οίκον μαθήματα μαθηματικών σε μαθητές Γυμνασίου –Λυκείου και υποψηφίους ανωτέρων και ανωτάτων σχολών.

Από τον Σεπτέμβριο του 2004 εργάζομαι ως μόνιμος καθηγητής μέσης εκπαίδευσης στο Γυμνάσιο Παλαιοκατούνας Ευρυτανίας, λόγω της πρόσληψης μου μέσω γραπτού διαγωνισμού του Α.Σ.Ε.Π που διεξήχθη τον Δεκέμβριο του 2002.

ABSTRACT

SUMMARY

Sotiris Valmas

Research on the effect of environmental variables on the distribution of plant species of mt. Erimanthos (NW Peloponissos) flora, using Proportional Odds model.

November 2005

It is known that the human impact on nature can cause changes that can be negative not only to the environment but also to the quality of our life. Thus, it is clear that any interference in the environment has to be done with the proper care and respect.

This research is based on an older research on the flora and vegetation of mt Erimanthos. It includes a large number of vegetation samplings through which the examination of plant species distribution in the mountain area of Erimanthos is possible.

Erimanthos is a mountain neighboring the urban center of Patras. This is the reason why it has suffered an intense human impact. It is its forest vegetation that has suffered most changes and degradation. Therefore, today's plant vegetation is a jigsaw puzzle of long-term changes and successions.

This study is trying to present the outcomes concerning the dynamic of vegetation of this particular mountain, taking into consideration the flora and the various ecological as well as environmental factors (such as height, inclination, orientation and the geological substrate of samplings). The understanding of vegetation dynamics is crucial for the future decisions on the management of the area.

The structure of the work originally includes a long descriptive analysis of all the recorded environmental factors.

After that there has been attempted an investigation of the interrelation between these factors. The last and most important stage is an attempt to describe the dynamic of vegetation through the model of Proportional Odds according to the above environmental factors as well as an appraisal of this model.

ΠΕΡΙΛΗΨΗ

Σωτήρης Βαλμάς.

Έρευνα της επίδρασης περιβαλλοντικών μεταβλητών στην κατανομή ειδών χλωρίδας του όρους Ερυμάνθου με τη χρήση του μοντέλου Proportional Odds.

Μάιος 2005.

Είναι γνωστό ότι η επίδραση του ανθρώπινου παράγοντα στην φύση επιφέρει μεταβολές, με δυσμενείς πολλές φορές συνέπειες τόσο στο περιβάλλον αλλά και στην ποιότητα ζωής του. Έτσι είναι πλέον σαφές ότι η οποιαδήποτε παρέμβαση στο περιβάλλον θα πρέπει να γίνεται με ιδιαίτερη προσοχή και δέοντα σεβασμό.

Τα στοιχεία στα οποία στηρίχθηκε η παρούσα εργασία, προέρχονται από παλαιότερη έρευνα γύρω από την χλωρίδα και την βλάστηση της βόρειας περιοχής της Πελοποννήσου. Περιέχει ένα ικανοποιητικό πλήθος δειγματοληψιών που αφορά την φυτοκοινωνιολογία του ορεινού όγκου του Ερυμάνθου.

Ο Ερύμανθος είναι ένα βουνό κοντά στο αστικό κέντρο της Πάτρας και για τον λόγο αυτό έχει υποστεί την ανθρώπινη επίδραση, συχνά και έντονα. Η δασική του βλάστηση είναι εκείνη που έχει δεχτεί τις περισσότερες μεταβολές και υποβαθμίσεις. Έτσι η σημερινή φυτοκοινωνιολογίκη εικόνα του, είναι ένα παζλ μακροχρόνιων αλλαγών και διαδοχών.

Η παρούσα εργασία επιχειρεί από την γνώση της χλωρίδας και των διαφόρων οικολογικών και περιβαλλοντικών παραγόντων (υψόμετρο, κλίση, προσανατολισμός και πέτρωμα εδάφους της δειγματοληψίας) να παρουσιάσει συμπεράσματα για την δυναμική της βλάστησης του συγκεκριμένου όρους. Η κατανόηση της δυναμικής της βλάστησης είναι σημαντική και για τον καθορισμό της κατάλληλης διαχείρισης της περιοχής.

Η δομή της εργασίας περιλαμβάνει αρχικά εκτενή περιγραφική ανάλυση όλων των περιβαλλοντικών παραγόντων που καταγράφονται. Στην συνέχεια, επιχειρείται η εξεύρεση σχέσεων εξάρτησης των παραγόντων αυτών μεταξύ τους. Το τελευταίο και σπουδαιότερο στάδιο, είναι η προσπάθεια να περιγραφεί η δυναμική της βλάστησης μέσω του μοντέλου proportional odds, βάση των παραπάνω περιβαλλοντικών παραγόντων. Καθώς επίσης και η αξιολόγηση του μοντέλου αυτού.





ΚΑΤΑΛΟΓΟΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

ΚΕΦΑΛΑΙΟ 1. ΕΙΣΑΓΩΓΗ	11
ΚΕΦΑΛΑΙΟ 2. ΠΕΡΙΓΡΑΦΗ ΔΕΔΟΜΕΝΩΝ.....	12
2.1. Τίτλοι και τιμές των μεταβλητών.....	12
2.2. Περιγραφική ανάλυση των μεταβλητών.	13
2.2.1 Προσανατολισμός	14
2.2.2 Υψόμετρο	15
2.2.3 Κλίση.	17
ΚΕΦΑΛΑΙΟ 3. ΣΤΑΤΙΣΤΙΚΗ ΑΝΑΛΥΣΗ.....	21
3.1. Σχέσεις και εξαρτήσεις μεταβλητών	21
3.1.1 Συντελεστής συσχέτισης.	21
3.1.2 Σχέση υψομέτρου και κλίσης του εδάφους.	22
3.2. Έλεγχος χ^2 ανεξαρτησίας μεταβλητών.	23
3.2.1 Θεωρητικό υπόβαθρο ελέγχου ανεξαρτησίας	23
3.2.2 Έλεγχος ανεξαρτησίας χ^2 , μεταξύ, υψομέτρου εδάφους και είδους πετρώματος της δειγματοληψίας.	24
3.2.3 Έλεγχος ανεξαρτησίας χ^2 , μεταξύ, κλίσης εδάφους και είδους πετρώματος αυτού. 26	26
3.2.4 Έλεγχος ανεξαρτησίας χ^2 , μεταξύ, είδους πετρώματος και προσανατολισμού δειγματοληψίας.	27
3.2.5 Έλεγχος ανεξαρτησίας χ^2 , μεταξύ, υψομέτρου εδάφους και προσανατολισμού δειγματοληψίας.	28
3.2.6 Έλεγχος ανεξαρτησίας χ^2 , μεταξύ, κλίσης εδάφους και προσανατολισμού δειγματοληψίας.	31
ΚΕΦΑΛΑΙΟ 4. Φυτά.....	33
4.1. Γενικά.....	33
4.2. Anemone Blanda (Anembla) Ποώδες φυτό.....	39
4.3. Quercus coccifera (Quercoco) Δέντρο.	43
4.4. Rosa canina (Rosacan) Θάμνος.....	48
4.5. Campanula statulata (Campspt) Ποώδες φυτό.	53

4.6. Acer mouspesulanum (Acermon) Δέντρο	57
4.7. Festuca jeenpertii (Festjea) Ποώδες αγρωστώδες φυτό.....	62
4.8. Philyrea latifolia (Philat) Δέντρο.....	70
4.9. Brachypodium retusum (Bayret) Ποώδες αγρωστώδες φυτό.....	74
ΚΕΦΑΛΑΙΟ 5. ΣΥΜΠΕΡΑΣΜΑΤΑ	81
ΚΕΦΑΛΑΙΟ 6. ΒΙΒΛΙΟΓΡΑΦΙΑ	85

ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

Πίνακας 1: Κατανομή του είδους τού πετρώματος τού εδάφους.....	13
Πίνακας 2: Κατανομή του προσανατολισμού των δειγματοληψιών.....	14
Πίνακας 3: Κατανομή του υψομέτρου του εδάφους	16
Πίνακας 4: Κατανομή της κλίσης του εδάφους.....	19
Πίνακας 5: Συσχέτιση μεταβλητών, κλίση και υψόμετρο εδάφους.	22
Πίνακας 6: Υψόμετρο-Είδος πετρώματος	26
Πίνακας 7: Κλίση εδάφους - Είδος πετρώματος.....	27
Πίνακας 8: Είδος πετρώματος - Προσανατολισμός εδάφους	28
Πίνακας 9: Υψόμετρο - Προσανατολισμός εδάφους	30
Πίνακας 10: Προσανατολισμός - Κλίση εδάφους.....	31
Πίνακας 11: Κατανομή της πληθοκάλυψης του εδάφους από το φυτό <i>Anemone Blanda</i> . .	40
Πίνακας 12: Κατανομή της πληθοκάλυψης του εδάφους από το φυτό <i>Quercus coccifera</i> . 44	
Πίνακας 13: Κατανομή της πληθοκάλυψης του εδάφους από το φυτό <i>Rosa canina</i>	49
Πίνακας 14: Κατανομή της πληθοκάλυψης του εδάφους από το φυτό <i>Campanula spatulata</i>	54
Πίνακας 15: Κατανομή της πληθοκάλυψης του εδάφους από το φυτό <i>Acer mouspesulanum</i>	58
Πίνακας 16: Κατανομή της πληθοκάλυψης του εδάφους από το φυτό <i>Festuca jeenpertii</i> .	63
Πίνακας 17: Κατανομή της πληθοκάλυψης του εδάφους από το φυτό <i>Philyrea latifolia</i>	71
Πίνακας 18: Κατανομή της πληθοκάλυψης του εδάφους από το φυτό <i>Brachypodium retusum</i>	75

ΚΑΤΑΛΟΓΟΣ ΓΡΑΦΗΜΑΤΩΝ

Γράφημα 1: Είδος πετρώματος του εδάφους	13
Γράφημα 2: Προσανατολισμός δειγματοληψιών	14
Γράφημα 3: Κλίση του εδάφους των δειγματοληψιών.....	19
Γράφημα 4: Διασπορά μεταβλητών, κλίση και υψόμετρο εδάφους.....	22
Γράφημα 5: Ραβδόγραμμα συχνοτήτων της πληθοκάλυψης του εδάφους από το φυτό Anemone Blanda	39
Γράφημα 6: Ραβδόγραμμα συχνοτήτων της πληθοκάλυψης του εδάφους από το φυτό Quercus coccifera	44
Γράφημα 7: Ραβδόγραμμα συχνοτήτων της πληθοκάλυψης του εδάφους από το φυτό Rosa canina	49
Γράφημα 8: Ραβδόγραμμα συχνοτήτων της πληθοκάλυψης του εδάφους από το φυτό Campanula spatulata.....	53
Γράφημα 9: Ραβδόγραμμα συχνοτήτων της πληθοκάλυψης του εδάφους από το φυτό Acer mouspesulanum.....	57
Γράφημα 10: Ραβδόγραμμα συχνοτήτων της πληθοκάλυψης του εδάφους από το φυτό Festuca jeenpertii	62
Γράφημα 11: Ραβδόγραμμα συχνοτήτων της πληθοκάλυψης του εδάφους από το φυτό Philyrea latifolia.....	70
Γράφημα 12: Ραβδόγραμμα συχνοτήτων της πληθοκάλυψης του εδάφους από το φυτό Brachypodium retusum.....	75

ΚΕΦΑΛΑΙΟ 1. ΕΙΣΑΓΩΓΗ

Το αντικείμενο της έρευνας είναι η στατιστική επεξεργασία στοιχείων μιας φυτοκοινωνιολογικής έρευνας της βλάστησης του όρους Ερυμάνθου της βόρειας Πελοποννήσου. Τα στοιχεία αυτά είναι ένα τμήμα μιας χλωριδικής και φυτοκοινωνιολογικής έρευνας που πραγματοποιήθηκε από τον Μαρούλη (2003), στα πλαίσια της διδακτορικής του διατριβής. Η έρευνα διεξήχθη στην ευρύτερη περιοχή του όρους Ερυμάνθου, που περικλείεται στα διοικητικά όρια των νομών Αχαΐας και Ηλείας. Οι δειγματοληψίες διενεργήθηκαν από το 1995 έως το 2000.

Η μέθοδος που χρησιμοποιήθηκε προκειμένου να προσδιορισθούν οι μονάδες βλάστησης είναι η φυτοκοινωνιολογική μέθοδος Braun-Blanquet (1964) και Kent M. & Coker P. (1992). Πραγματοποιήθηκαν 337 δειγματοληψίες με την μέθοδο αυτή. Η δειγματοληπτική τεχνική που ακολουθήθηκε ήταν (proportional to size) ανάλογη του μεγέθους του πληθυσμού των φυτών. Δηλαδή ο αριθμός των δειγμάτων που εκλέχθηκαν να είναι ανάλογος της έκτασης που καταλαμβάνει ο αντίστοιχος τύπος βλάστησης και να εκπροσωπείται το σύνολο των περιβαλλοντικών παραμέτρων (υψόμετρο, κλίση, έκθεση, είδος πετρώματος εδάφους). Επίσης είχε τα εξής στοιχειά:

- A. Κάθε δειγματοληπτική επιφάνεια επιλέχτηκε έτσι ώστε να είναι φυσιογνωμικά ομοιογενής δηλαδή να μην διακόπτεται από μονοπάτια ξέφωτα κλπ.
- B. Εντός της δειγματοληπτικής επιφάνειας καταβλήθηκε προσπάθεια να μην μεταβάλλονται οι οικολογικοί παράγοντες, δηλαδή να μην υπάρχουν αλλαγές στην κλίση την έκθεση το υψόμετρο των φωτισμό κλπ.

Σε κάθε δειγματοληπτική μονάδα καταγράφηκαν τα είδη των φυτών που περιέχονταν, καθώς και η πληθοκάλυψη του εδάφους που αυτά δημιουργούν. Η κλίμακα μέτρησης είναι ordinal από 0 έως 5.

Σκοπός της ανάλυσης που θα πραγματοποιηθεί είναι:

- A. Να εξεταστούν, για τους διάφορους περιβαλλοντικούς παράγοντες, οι πιθανές εξαρτήσεις και συσχετίσεις που υπάρχουν μεταξύ τους.
- B. Να εξεταστεί εάν μπορεί να υπάρξει μοντελοποίηση των διαφόρων φυτοκοινωνιολογικών μονάδων, μέσω των περιβαλλοντικών παραγόντων.
- Γ. Να ελεγχθεί και να αξιολογηθεί κατά πόσο το μοντέλο μπορεί να ερμηνεύσει την κατανομή των ειδών κατά μήκος των περιβαλλοντικών διαβαθμίσεων.
- Δ. Να εξεταστεί η εξάρτηση της πληθοκάλυψης δασικών φυτών από τους ανωτέρω περιβαλλοντικούς παράγοντες.

ΚΕΦΑΛΑΙΟ 2. ΠΕΡΙΓΡΑΦΗ ΔΕΔΟΜΕΝΩΝ.

2.1. Τίτλοι και τιμές των μεταβλητών.

Συνοπτικά, οι μεταβλητές που καταγράφηκαν κατά την έρευνα είναι :

- **Πληθοκάλυψη του εδάφους:** Κατηγορική μεταβλητή με τιμές:

0: απουσία του φυτού από το δείγμα .

1: παρουσία του φυτού στο δείγμα δίνοντας πληθοκάλυψη έως 5% του εδάφους

2: παρουσία του φυτού στο δείγμα, δίνοντας πληθοκάλυψη από 5% έως 25% του εδάφους

3: παρουσία του φυτού στο δείγμα δίνοντας πληθοκάλυψη από 25% έως 50% του εδάφους

4: παρουσία του φυτού στο δείγμα, δίνοντας πληθοκάλυψη από 50% έως 75% του εδάφους

5: παρουσία του φυτού στο δείγμα, δίνοντας πληθοκάλυψη από 75% έως 100% του εδάφους

- **Υψόμετρο της δειγματοληψίας :** Συνεχής μεταβλητή με τιμές των υψόμετρων αυτές που είχαν τα σημεία που έγιναν οι δειγματοληψίες. Οι τιμές κυμαίνονται από 270 έως 1590 μέτρα.

- **Πέτρωμα του εδάφους.** Κατηγορική μεταβλητή με τιμές:

1: Όταν το πέτρωμα αποτελείται από ασβεστόλιθο.

2: Όταν το πέτρωμα αποτελείται από φλύσχη.

3: Όταν το πέτρωμα αποτελείται από ραδιολαρίτη.

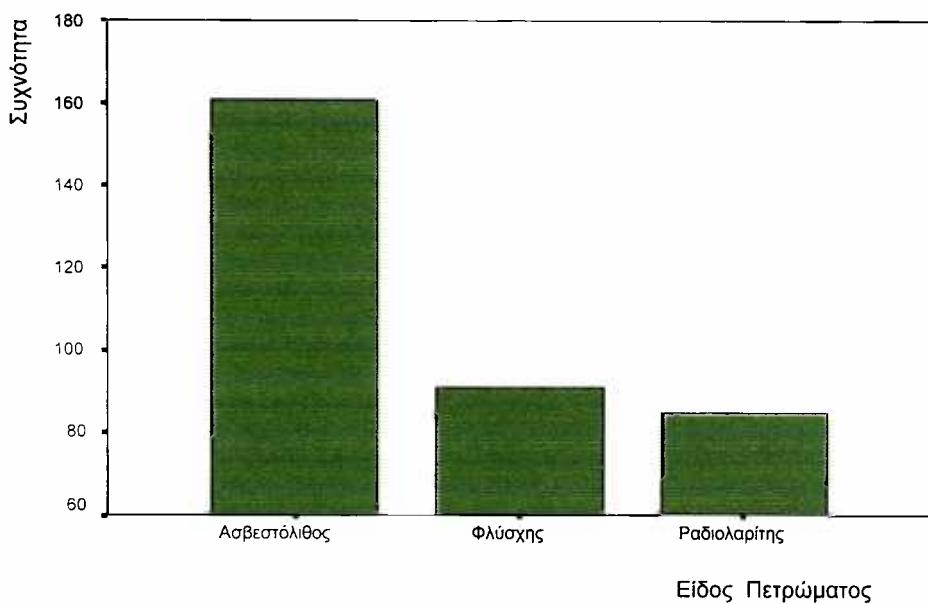
- **Κλίση του εδάφους.** Συνεχής μεταβλητή με τιμές, τις τιμές της κλίσης του εδάφους μετρημένες σε μοίρες, στα σημεία που έγιναν οι δειγματοληψίες. Το εύρος των τιμών είναι από 30 έως 67 μοίρες :

- Προσανατολισμός της δειγματοληψίας :** Κατηγορική μεταβλητή, που περιγράφει το είδος της έκθεσης του δείγματος ως προς τα σημεία του ορίζοντα και έχει τιμές:
1: Βορράς, 2: Νότος, 3: Δύση, 4: Ανατολή .

2.2. Περιγραφική ανάλυση των μεταβλητών.

Είδος πετρώματος

Το Ραβδόγραμμα συχνοτήτων για το είδος του πετρώματος του εδάφους είναι :



Γράφημα 1: Είδος πετρώματος του εδάφους.

Παρατηρείται ότι το ασβεστολιθικό πέτρωμα είναι εμφανώς το συχνότερα εμφανιζόμενο στο όρος Ερύμανθο όπου έγινε η συγκεκριμένη έρευνα.

Ο πίνακας συχνοτήτων των τιμών της μεταβλητής είναι ο εξής :

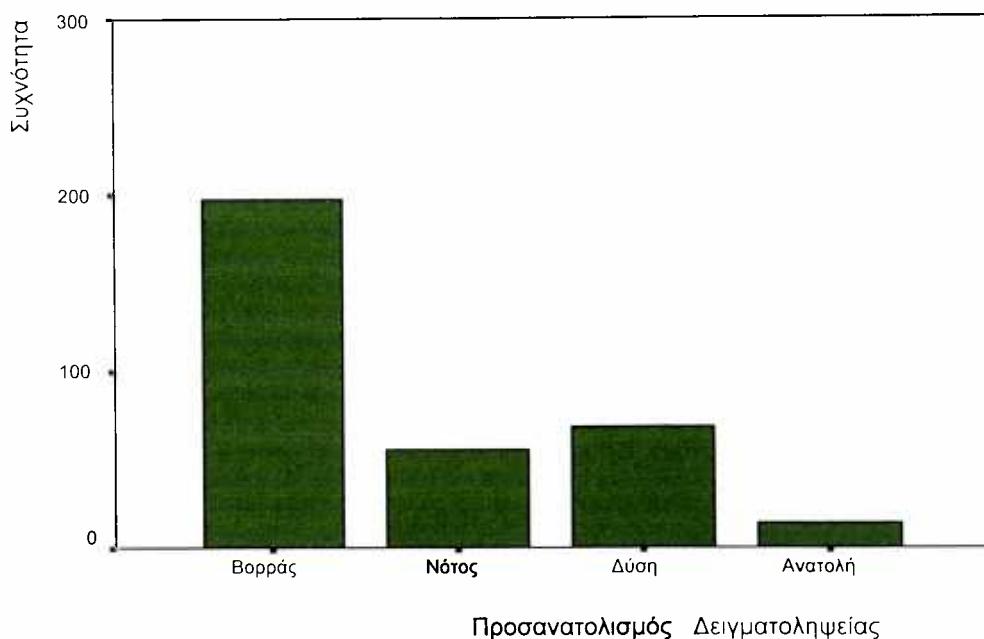
Είδος πετρώματος	Συχνότητα	Σχετική συχνότητα (%)	Αθροιστική Σχετική Συχνότητα (%)
Ασβεστόλιθος	161	47,8	47,8
Φλύσχης	91	27,0	74,8
Ραδιολαρίτης	81	25,2	100,0
Γενικό Σύνολο		100,0	

Πίνακας 1: Κατανομή του είδους του πετρώματος του εδάφους.

Παρατηρείται ότι το 47,8 % των δειγματοληψιών έγιναν σε ασβεστολιθικό πέτρωμα. Αυτό οφείλεται στο ότι το μεγαλύτερο μέρος του Ερυμάνθου έχει τέτοιο πέτρωμα.

2.2.1 Προσανατολισμός

Το Ραβδόγραμμα συχνοτήτων για τον προσανατολισμό των δειγματοληψιών είναι:



Γράφημα 2: Προσανατολισμός δειγματοληψιών

Μια ακριβέστερη εικόνα για την κατανομή των εκθέσεων των δειγματοληψιών προκύπτει από τον πίνακα συχνοτήτων.

Έτσι ο πίνακας συχνοτήτων των τιμών της μεταβλητής είναι ο εξής :

Προσανατολισμός	Συχνότητα	Σχετική συχνότητα (%)	Αθροιστική Συχνότητα (%)
Βορράς	194	57,6	57,6
Νότος	56	16,6	74,2
Ανατολή	14	4,2	78,4
Δύση	73	21,7	100,0
Γενικό Σύνολο	337	100,0	

Πίνακας 2: Κατανομή του προσανατολισμού των δειγματοληψιών.

Παρατηρείται ότι, η επικρατούσα τιμή της μεταβλητής είναι το σημείο Βορράς με ποσοστό 57,6%. Δηλαδή το 57,6% των δειγματοληψιών ελήφθησαν σε σημεία που είχαν έκθεση προς το Βορρά, ενώ το 21,7% έγιναν σε περιοχές που είχαν δυτικό προσανατολισμό.

2.2.2 Υψόμετρο

Για τη μεταβλητή αυτή υπολογίζονται τα εξής :

Μέτρα θέσης

Μέση τιμή : 907,63 μέτρα.

Διάμεσος τιμή : 900 μέτρα.

Παρατηρείται ότι η διάμεσος τιμή είναι περίπου ίση με τη μέση τιμή, κατά συνέπεια αναμένεται μια συμμετρική κατανομή των υψομέτρων που έγιναν οι δειγματοληψίες.

Μέτρα μεταβλητότητας

Ελάχιστη τιμή : 270 μέτρα.

Μέγιστη τιμή : 1590 μέτρα.

Εύρος (έκταση) τιμών μεταβλητής : 1320 μέτρα.

Τυπική απόκλιση (s): 270,01 μέτρα.

Μέτρα σχετικής θέσης

Από τον υπολογισμό των τεταρτημορίων της μεταβλητής, συνάγεται ότι

- το 25% των δειγματοληψιών έγιναν σε υψόμετρο από 270μ. έως 730μ.
- το 50% έγιναν σε υψόμετρο από 270μ έως 900μ.
- το 75% των δειγματοληψιών έγιναν σε υψόμετρο από 270μ. έως 1075μ.

Εξάλλου, με βάση τη μέση τιμή και την τυπική απόκλιση υπολογίζεται το ποσοστό των παρατηρήσεων που βρίσκονται τιμές στα διαστήματα που απέχουν από τη μέση τιμή απόσταση ίση με μία, δύο και τρεις τυπικές αποκλίσεις.

	Απόσταση τιμών από το μέσο	Ποσοστό παρατηρήσεων
Μία τυπική απόκλιση	$x - s = 637,62 \mu$ και $x + s = 1177,64 \mu$.	68,2%
Δύο τυπικές αποκλίσεις	$x - 2s = 367,61 \mu$ και $x + 2s = 1147,65 \mu$	79,2%
Τρεις τυπικές αποκλίσεις	$x - 3s = 97,60 \mu$ και $x + 3s = 1717,66 \mu$	100%

Από τις τιμές των ποσοστών αναμένεται η κατανομή της μεταβλητής να πλησιάσει την κανονική.

Συντελεστής μεταβλητότητας

$$CV = \frac{s}{\bar{x}} = \frac{270}{907,63} = 0,297$$

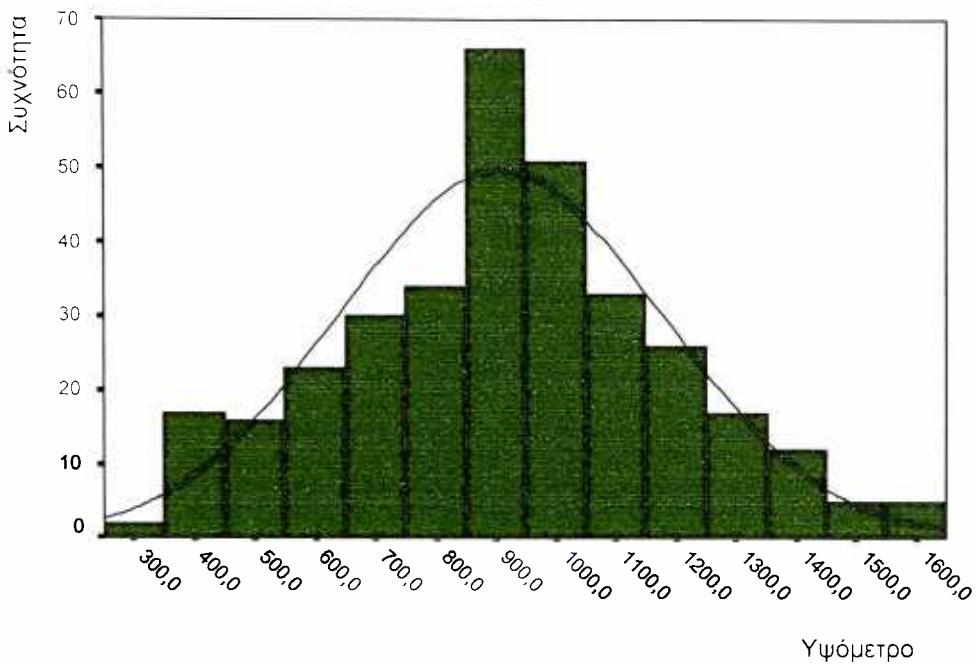
Δηλαδή, το υψόμετρο της δειγματοληψίας παρουσιάζει μικρή μεταβλητότητα. Άρα εμφανίζει σχετική ομοιογένεια.

Στη συνέχεια, χρησιμοποιώντας κλάσεις πλάτους 100 μέτρων, δημιουργείται ο επόμενος πίνακας συγχονοτήτων.

Τάξεις υψομέτρου	Συχνότητα	Σχετική συχνότητα (%)	Αθροιστική σχετική συχνότητα (%)
1. 270 – 370	5	1,5	1,5
2. 370 – 470	22	6,5	8,0
3. 470 – 570	11	3,3	11,3
4. 570 – 670	34	10,1	21,4
5. 670 – 770	25	7,4	28,8
6. 770 – 870	49	14,5	43,3
7. 870 – 970	66	19,6	62,9
8. 970 – 1070	41	12,2	75,1
9. 1070 – 1170	30	8,9	84,0
10. 1170 – 1270	23	6,8	90,8
11. 1270 – 1370	16	4,7	95,5
12. 1370 – 1470	8	2,4	97,9
13. 1470 – 1590	7	2,1	100
Σύνολο	337	100,0	

Πίνακας 3: Κατανομή του υψομέτρου του εδάφους

Διαπιστώνεται από τον πίνακα συγχονοτήτων ότι το 79,5% των δειγματοληψιών έγιναν σε υψόμετρα από 570 έως 1270 μέτρα. Μόνο το 11,3% των δειγματοληψιών ελήφθησαν σε υψόμετρα από 270 έως 570 μέτρα και το 9,2% ελήφθησαν σε υψόμετρα από 1270 έως 1590 μέτρα.



Γράφημα: Υψόμετρο των δειγματοληψιών.

Από το ιστόγραμμα διακρίνεται ότι η κατανομή του υψομέτρου μοιάζει να είναι συμμετρική. Η μέθοδος δειγματοληψίας (Braun-Blauwet 1964), που ακολουθείται έχει ως στόχο την ομοιόμορφη εκπροσώπηση όλων των τύπων βλάστησης του όρους Ερυμάνθου και όχι την ομοιόμορφη λήψη δειγμάτων στα διάφορα επίπεδα του υψομέτρου του όρους αυτού. Για τον λόγο αυτό η κατανομή του υψομέτρου των δειγματοληψιών δεν διαφαίνεται, από το προηγούμενο ιστόγραμμα, να ακολουθεί την ομοιόμορφη κατανομή. Αντίθετα έχει τη μορφή συμμετρικής μονοκόρυφης κατανομής.

2.2.3 Κλίση.

Για τη συνεχή αυτή μεταβλητή ανάλογα υπολογίζονται:

Μέτρα θέσης

Μέση τιμή : 41,81 μοίρες

Διάμεσος τιμή : 40 μοίρες

Επικρατούσα τιμή : 40 μοίρες.

Παρατηρείται ότι τα μέτρα θέσης για την μεταβλητή αυτή είναι πολύ κοντά γεγονός που δίνει την εντύπωση ότι η κατανομή της κλίσης του εδάφους είναι συμμετρική κατανομή.

Μέτρα μεταβλητότητας

Ελάχιστη τιμή : 5 μοίρες

Μέγιστη τιμή : 67 μοίρες

Εύρος (έκταση) τιμών μεταβλητής : 62 μοίρες

Τυπική απόκλιση (s): 12,24 μοίρες

Μέτρα σχετικής θέσης

Από τον υπολογισμό των τεταρτημορίων της μεταβλητής, παρατηρείται ότι:

- το 25% των δειγματοληψιών έγιναν σε κλίσεις από 5 μοίρες έως 35 μοίρες.
- το 50% έγιναν από 5 μοίρες έως 40 μοίρες .
- το 75% των δειγματοληψιών έγιναν σε κλίσεις από 5 μοίρες έως 50 μοίρες.

Εξάλλου, με βάση τη μέση τιμή και την τυπική απόκλιση υπολογίζεται το ποσοστό των παρατηρήσεων που έχουν τιμές στα διαστήματα που απέχουν από τη μέση τιμή απόσταση ίση με μία, δύο και τρεις τυπικές αποκλίσεις.

	Απόσταση τιμών από το μέσο	Ποσοστό παρατηρήσεων
Μία τυπική απόκλιση	$\bar{x} - s = 29,57 \mu$ και $\bar{x} + s = 54,05 \mu$.	69,7%
Δύο τυπικές αποκλίσεις	$\bar{x} - 2s = 17,33 \mu$ και $\bar{x} + 2s = 66,29 \mu$	95,3%
Τρεις τυπικές αποκλίσεις	$\bar{x} - 3s = 5,09 \mu$ και $\bar{x} + 3s = 78,53 \mu$	99,4%

Συντελεστής μεταβλητότητας

$$CV = \frac{s}{\bar{x}} = \frac{12,24}{41,81} = 0,2927.$$

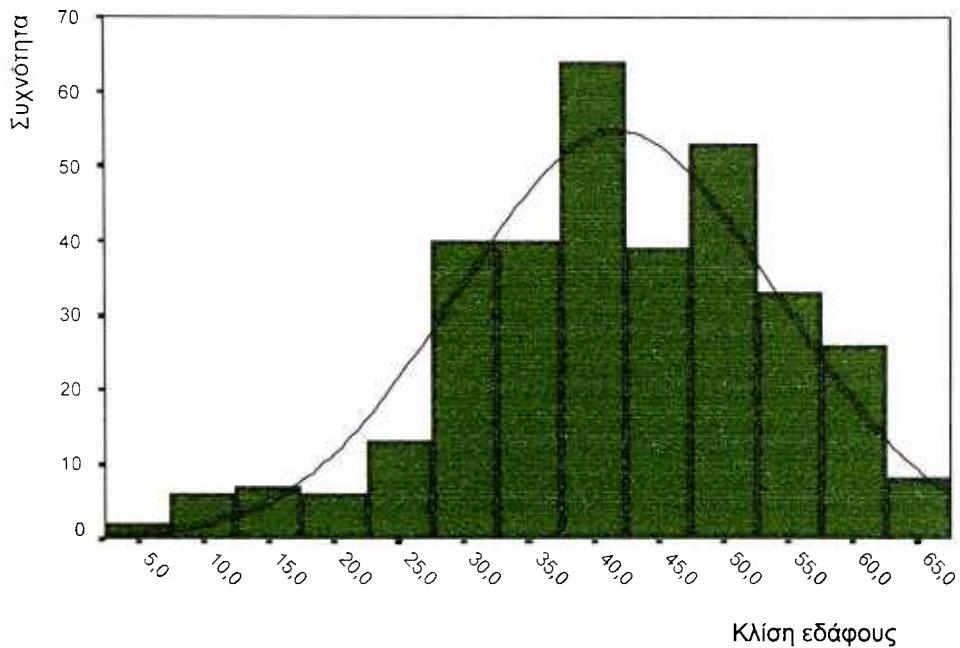
Παρατηρείται ότι η κλίση του εδάφους παρουσιάζει μικρή μεταβλητότητα, γεγονός που σημαίνει ότι η κατανομή παρουσιάζει σχετική ομοιογένεια.

Η ομαδοποίηση της κλίσης του εδάφους σε κλάσεις πλάτους των 10 μοιρών δίνεται στον πίνακα :

Τάξεις κλίσης	Συχνότητα	Σχετική συχνότητα (%)	Αθροιστική σχετική συχνότητα (%)
1.	5 – 15	15	4,5
2.	15 – 25	18	5,3
3.	25 – 35	77	22,8
4.	35 – 45	105	31,2
5.	45 – 55	86	25,5
6.	55 – 67	36	10,7
Σύνολο		337	100,0

Πίνακας 4: Κατανομή της κλίσης του εδάφους.

Στο επόμενο διάγραμμα παρουσιάζεται μια εποπτική εικόνα της κατανομής.



Γράφημα 3: Κλίση του εδάφους των δειγματοληψιών.

Από το προηγούμενο ιστόγραμμα διαφαίνεται ότι η κατανομή της κλίσης εδάφους των δειγματοληψιών είναι μια αριστερά ασύμμετρη κατανομή. Που σημαίνει ότι στις μικρές κλίσεις ελήφθησαν λίγες δειγματοληψίες.

ΚΕΦΑΛΑΙΟ 3. ΣΤΑΤΙΣΤΙΚΗ ΑΝΑΛΥΣΗ

3.1. Σχέσεις και εξαρτήσεις μεταβλητών

3.1.1 Συντελεστής συσχέτισης.

Στην ενότητα αυτή θα εξεταστεί εάν υπάρχουν σχέσεις μεταξύ των μεταβλητών της έρευνας. Οι μεταβλητές υψόμετρο και κλίση εδάφους είναι συνεχείς μεταβλητές.

1. Εδώ θα παρουσιαστεί το μέτρο έντασης της γραμμικής σχέσης δυο μεταβλητών.

Ορισμός: Έστω δυο τυχαίες μεταβλητές X, Y , με μέσες τιμές μ_X, μ_Y και τυπικές αποκλίσεις σ_X, σ_Y αντίστοιχα. Ορίζεται ως **συντελεστής συσχέτισης** (correlation coefficient) των X, Y και συμβολίζεται με $\rho(X, Y)$ ή $\rho_{X,Y}$ το πηλίκο της συνδιακύμανσης των X, Y προς το γινόμενο των τυπικών αποκλίσεων. Δηλαδή:

$$\rho(X, Y) = \text{COV}(X, Y) / \sigma_X \sigma_Y$$

Συνδιακύμανση των μεταβλητών είναι:

$$\text{COV}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)].$$

Όπως δηλώνεται και από την ονομασία, αλλά φαίνεται και από τον τύπο, η συνδιακύμανση είναι ένα μέτρο του τρόπου με τον οποίο οι δυο τυχαίες μεταβλητές μεταβάλλονται από κοινού.

Ο συντελεστής συσχέτισης παίρνει τιμές: $-1 \leq \rho(X, Y) \leq 1$

Όταν οι τιμές του συντελεστή συσχέτισης πλησιάζουν το 1 τότε υπάρχει θετική γραμμική εξάρτηση των μεταβλητών X, Y . Όταν οι τιμές του συντελεστή συσχέτισης είναι κοντά στο -1, τότε υπάρχει αρνητική γραμμική εξάρτηση μεταξύ των μεταβλητών. Όταν οι τιμές του συντελεστή συσχέτισης δεν διαφέρουν πολύ από το μηδέν οδηγούν στο συμπέρασμα ότι δεν υπάρχει γραμμική σχέση των μεταξύ των μεταβλητών, χωρίς αυτό να αποκλείει εξάρτηση άλλης μορφής. Ο συντελεστής συσχέτισης που αναφέρθηκε είναι ο πληθυσμιακός συντελεστής συσχέτισης.

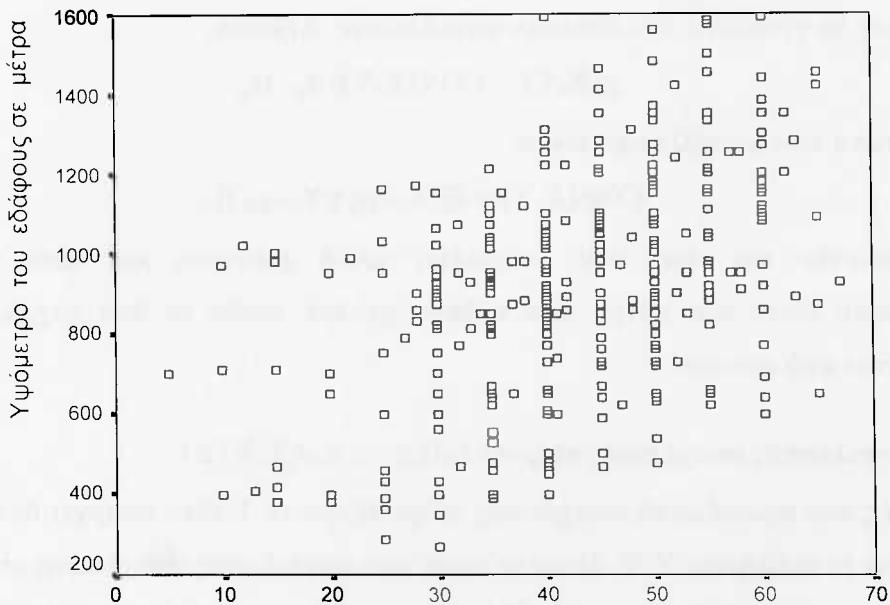
Εάν αντί για ολόκληρο τον πληθυσμό έχουμε δείγμα απ' αυτόν και X_i και Y_i με $i = 0, 1, 2, \dots, n$ οι μεταβλητές που καταγράφονται όπου n το μέγεθος του δείγματος, τότε ορίζεται ο **δειγματικός συντελεστής συσχέτισης** (sample correlation coefficient). Ο συντελεστής αυτός συμβολίζεται R ή $R(X, Y)$ και υπολογίζεται βάση των μετρήσεων από τον εξής τύπο

$$R = \frac{\sum_{i=1}^n (X_i - \bar{X}) \cdot (Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

- Ο δειγματικός συντελεστής συσχέτισης είναι καθαρός αριθμός χωρίς να επηρεάζεται από τις μονάδες των μεταβλητών
- $R(X, Y) = R(Y, X) = R$
- Εάν $R < 0$ τότε υπάρχει αρνητική γραμμική συσχέτιση μεταξύ των μεταβλητών.
- Εάν $R > 0$ τότε υπάρχει θετική γραμμική συσχέτιση μεταξύ των μεταβλητών.
- Εάν R είναι κοντά στο μηδέν τότε δεν υπάρχει γραμμική σχέση.

3.1.2 Σχέση υψομέτρου και κλίσης του εδάφους.

Οι μεταβλητές είναι συνεχείς. Το διάγραμμα διασποράς δίνει μια πρώτη εικόνα για ύπαρξη σχέσης των μεταβλητών.



Κλίση του εδάφους σε μοίρες.

Γράφημα 4: Διασπορά μεταβλητών, κλίση και υψόμετρο εδάφους.

Παρατηρείται στο διάγραμμα διασποράς ότι υπάρχει κάποια συσχέτιση των μεταβλητών. Ο συντελεστής συσχέτισης των μεταβλητών αυτών δίνεται στον πίνακα:

	Υψόμετρο - Κλίση
Pearson-Correlation	0,436
Sig.(2-tailed)	0,000
Σύνολο	337

Πίνακας 5: Συσχέτιση μεταβλητών, κλίση και υψόμετρο εδάφους.

Η τιμή του συντελεστή γραμμικής συσχέτισης του Pearson R=0,436 είναι μια μικρή σχετικά τιμή αλλά από τον έλεγχο για το εάν η τιμή είναι στατιστικά σημαντική προκύπτει από το παρατηρούμενο επίπεδο σημαντικότητας που είναι σχεδόν μηδέν ότι ο συντελεστής γραμμικής συσχέτισης δεν είναι μηδενικός. Έτσι θα μπορούσε να θεωρηθεί ότι υπάρχει μια ασθενής θετική σχέση γραμμικής συσχέτισης μεταξύ των συνεχών μεταβλητών, κλίση και υψομέτρου των δειγματοληπτικών επιφανειών.

3.2. Έλεγχος χ^2 ανεξαρτησίας μεταβλητών.

Τα διαθέσιμα εργαλεία : Chi-Square.

3.2.1 Θεωρητικό υπόβαθρο ελέγχου ανεξαρτησίας

Το στατιστικό εργαλείο που χρησιμοποιείται για τη διαπίστωση εξάρτησης ή ανεξαρτησίας μεταβλητών είναι ο μη παραμετρικός έλεγχος χ^2 . Ο έλεγχος χ^2 χρησιμοποιείται για την επιβεβαίωση ή όχι της στατιστικής σημαντικότητας αποτελεσμάτων που παρουσιάζονται σε δυσδιάστατους πίνακες (πίνακες συνάφειας, crosstabs). Οι πίνακες αυτοί χρησιμοποιούνται για να γίνει αντιληπτή ύπαρξη σχέσης, μεταξύ δυο μεταβλητών. Προϋπόθεση για την εφαρμογή του ελέγχου αυτού είναι στους πίνακες να εμφανίζονται συχνότητες και όχι ποσοστά (Ξεκαλάκη Ε. 2001).

Στους πίνακες αυτούς η μηδενική υπόθεση ορίζεται ως :

H_0 : Δεν υπάρχει συσχέτιση μεταξύ των μεταβλητών

Το ερώτημα που προκύπτει είναι το πόσο πιθανό είναι να έχουμε τις συχνότητες που εμφανίζονται στον πίνακα συνάφειας όταν η μηδενική υπόθεση είναι σωστή. Οι συχνότητες που εμφανίζονται στον πίνακα είναι οι παρατηρούμενες συχνότητες. Στον έλεγχο χρειάζεται να υπολογιστούν και οι αναμενόμενες συχνότητες, όταν ισχύει η μηδενική υπόθεση. Αυτές ορίζονται ως εξής:

(Σύνολο της κατηγορίας i της μεταβλητής 1) x (Σύνολο της κατηγορίας j της μεταβλητής 2) / μέγεθος του δείγματος

Το ενδιαφέρον βρίσκεται στη διαφορά $O_{ij} - E_{ij}$, όπου:

- O_{ij} : η παρατηρούμενη συχνότητα στο κελί (i, j)
- E_{ij} : η αναμενόμενη συχνότητα στο κελί (i, j)

Επειδή όμως το άθροισμα των διαφορών $O_{ij} - E_{ij}$ για όλα τα κελιά του πίνακα συνάφειας είναι μηδενικό, χρησιμοποιείται το $(O_{ij} - E_{ij})^2$. Όμως για να συμπεριλαμβάνεται και η σημαντικότητα της κάθε διαφοράς χρησιμοποιείται τελικά η έκφραση $(O_{ij} - E_{ij})^2/E_{ij}$

Οπότε η στατιστική συνάρτηση του ελέγχου είναι η :

$$X^2 = \sum_{ij} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Παρατηρήσεις

- (α) Η X^2 είναι τυχαία μεταβλητή και αλλάζει τιμή από δείγμα σε δείγμα
- (β) Η κατανομή της X^2 είναι η χ^2 με βαθμούς ελευθερίας $[(C-1) \times (R-1)]$, όπου C είναι ο αριθμός των στηλών του πίνακα και R είναι ο αριθμός των γραμμών.

Αφού υπολογιστεί η τιμή της X^2 , βρίσκονται από τους πίνακες της κατανομής, για τους δεδομένους κάθε φορά οι βαθμοί ελευθερίας και το επίπεδο στατιστικής σημαντικότητας.

Αν $\chi^2 < X^2$, τότε απορρίπτεται η μηδενική υπόθεση, ενώ εάν $\chi^2 > X^2$, τότε δεν υπάρχουν επαρκείς ενδείξεις απόρριψης της μηδενικής υπόθεσης.

3.2.2 Έλεγχος ανεξαρτησίας χ^2 , μεταξύ, υψομέτρου εδάφους και είδους πετρώματος της δειγματοληψίας.

Κατά την πραγματοποίηση των δειγματοληψιών καταγράφτηκαν, το υψόμετρο και το είδος του πετρώματος. Θα ελεγχθεί εάν είναι ανεξάρτητο το είδος του πετρώματος του εδάφους από το υψόμετρο της δειγματοληψίας. Οι υποθέσεις που ελέγχονται είναι:

H₀ : Το είδος του πετρώματος είναι ανεξάρτητο από το υψόμετρο του εδάφους.

H₁ : Το είδος του πετρώματος δεν είναι ανεξάρτητο από το υψόμετρο του εδάφους.

Για τον έλεγχο αυτό χρησιμοποιείται ο επόμενος πίνακα συνάφειας:

Κλάσεις Υψομέτρου		Είδος Πετρώματος			
		Ασβεστόλιθος	Φλύσχης	Ραδιολαρίτης	Σύνολο
270-370	Συχνότητα	2	3	0	5
	Αναμ. Συχν. υπό την Ανεξαρτησία	2,4	1,4	1,3	5
370-470	Συχνότητα	1	21	0	22
	Αναμ. Συχν. υπό την Ανεξαρτησία	10,5	5,9	5,5	22

Κλάσεις Υψομέτρου		Είδος Πετρώματος			
		Ασβεστόλιθος	Φλύσχης	Ραδιολαρίτης	Σύνολο
470-570	Συχνότητα	2	9	0	11
	Αναμ. Συχν. υπό την Ανεξαρτησία	5,3	3	2,8	11
570-670	Συχνότητα	14	20	0	34
	Αναμ. Συχν. υπό την Ανεξαρτησία	16,2	9,2	8,6	34
670-770	Συχνότητα	7	13	5	25
	Αναμ. Συχν. υπό την Ανεξαρτησία	11,9	6,8	6,3	25
770-870	Συχνότητα	17	9	23	49
	Αναμ. Συχν. υπό την Ανεξαρτησία	23,4	13,2	12,4	49
870-970	Συχνότητα	34	12	20	66
	Αναμ. Συχν. υπό την Ανεξαρτησία	31,5	17,8	16,6	66
970-1070	Συχνότητα	25	3	13	41
	Αναμ. Συχν. υπό την Ανεξαρτησία	19,6	11,1	10,3	41
1070-1170	Συχνότητα	19	0	11	30
	Αναμ. Συχν. υπό την Ανεξαρτησία	14,3	8,1	7,6	30
1170-1270	Συχνότητα	14	1	8	23
	Αναμ. Συχν. υπό την Ανεξαρτησία	11	6,2	5,8	23
1270-1370	Συχνότητα	12	0	4	16
	Αναμ. Συχν. υπό την Ανεξαρτησία	7,6	4,3	4	16
1370-1470	Συχνότητα	7	0	1	8
	Αναμ. Συχν. υπό την Ανεξαρτησία	3,3	1,9	1,8	7
1470-1590	Συχνότητα	7	0	0	7
	Αναμ. Συχν. υπό την Ανεξαρτησία	3,8	1,9	1,8	7
Σύνολο	Συχνότητα	161	91	85	337

Κλάσεις Υψομέτρου	Είδος Πετρώματος			
	Ασβεστόλιθος	Φλύσχης	Ραδιολαρίτης	Σύνολο
Αναμ.Συχν. υπό την ανεξαρτησία	161	91	85	337

Πίνακας 6: Υψόμετρο-Είδος πετρώματος

Η τιμή της ελεγχοσυνάρτησης για τον έλεγχο της ανεξαρτησίας των μεταβλητών είναι $X^2=162,522$ με 24 βαθμούς ελευθερίας. Το παρατηρούμενο επίπεδο σημαντικότητας είναι σχεδόν μηδέν. Αυτό σημαίνει ότι απορρίπτεται η αρχική υπόθεση της ανεξαρτησίας των μεταβλητών. Από τον πίνακα διπλής εισόδου παρατηρείται ότι υπάρχει μια τάση να εμφανίζεται το ασβεστολιθικό πέτρωμα στα υψηλότερα υψόμετρα. Επίσης ο φλύσχης να έχει μεγαλύτερη προτίμηση στα χαμηλά υψόμετρα, ενώ ο ραδιολαρήτης είναι φιλικότερος στα ενδιάμεσα υψόμετρα του όρους Ερυμάνθου.

3.2.3 Έλεγχος ανεξαρτησίας χ^2 , μεταξύ, κλίσης εδάφους και είδους πετρώματος αυτού.

Κατά την διεξαγωγή των δειγματοληψιών καταγράφτηκαν η κλίση του εδάφους και το είδος του πετρώματος. Εξετάζεται εάν οι δυο αυτές μεταβλητές είναι ανεξάρτητες. Έτσι προκύπτουν οι υποθέσεις :

H_0 : Το είδος του πετρώματος είναι ανεξάρτητο από την κλίση του εδάφους.

H_1 : Το είδος του πετρώματος δεν είναι ανεξάρτητο από την κλίση του εδάφους.

Ο πίνακας συνάφειας είναι :

Κλίση εδάφους σε μοίρες	Είδος Πετρώματος				
	Ασβεστόλιθος	Φλύσχης	Ραδιολαρίτης	Σύνολο	
5-15	Συχνότητα	10	5	0	15
	Αναμ.Συχν. υπό την ανεξαρτησία	7,2	4,1	3,8	15
15-25	Συχνότητα	8	10	0	18
	Αναμ. Συχν. υπό την ανεξαρτησία	8,6	4,9	4,5	18
25-35	Συχνότητα	37	28	12	77
	Αναμ. Συχν. υπό την ανεξαρτησία	36,8	20,8	19,4	

	Συχνότητα	39	30	36	105
35-45	Αναμ. Συχν. υπό την ανεξαρτησία	50,2	28,4	26,5	105
45-55	Συχνότητα	46	14	26	86
	Αναμ. Συχν. υπό την ανεξαρτησία	41,1	23,3	21,7	86
55-67	Συχνότητα	21	4	11	36
	Αναμ. Συχν. υπό την ανεξαρτησία	17,2	9,7	9,1	36
Σύνολο	Συχνότητα	161	91	85	337
	Αναμ. Συχν. υπό την ανεξαρτησία	161	91	85	337

Πίνακας 7: Κλίση εδάφους - Είδος πετρώματος.

Η τιμή της ελεγχοσυνάρτησης Pearson για τον παραπάνω έλεγχο είναι $X^2 = 36,197$ με 10 βαθμούς ελευθερίας. Το παρατηρούμενο επίπεδο σημαντικότητας είναι σχεδόν μηδέν. Αυτό σημαίνει ότι απορρίπτεται η μηδενική υπόθεση. Δηλαδή απορρίπτεται η υπόθεση της ανεξαρτησίας της κλίσης του εδάφους και το είδος του πετρώματος του. Στον πίνακα συνάφειας παρατηρείται ότι ο ασβεστόλιθος εμφανίζεται περισσότερο στις μεγάλες κλίσεις, ενώ ο ραδιολαρήτης έχει μια προτίμηση στις μεσαίες και υψηλές κλίσεις. Τέλος ο φλύσχης είναι πιο φιλικός στις μικρές κλίσεις.

3.2.4 Έλεγχος ανεξαρτησίας χ^2 , μεταξύ, είδους πετρώματος και προσανατολισμού δειγματοληψίας.

Κατά την πραγματοποίηση της δειγματοληψίας καταγράφτηκαν το είδος του πετρώματος του εδάφους και η θέση του δείγματος στα σημεία του ορίζοντα. Για τον έλεγχο της ανεξαρτησίας των μεταβλητών διατυπώνονται οι υποθέσεις του ελέγχου:

H_0 : Το είδος του πετρώματος είναι ανεξάρτητο από τον προσανατολισμό του εδάφους.

H_1 : Το είδος του πετρώματος δεν είναι ανεξάρτητο από τον προσανατολισμό του εδάφους.

Ο έλεγχος αυτός γίνεται με την βοήθεια του πίνακα συνάφειας :

Προσανατολισμός		Είδος Πετρώματος			
		Ασβεστόλιθος	Φλύσχης	Ραδιολαρίτης	Σύνολο
Βορράς	Συχνότητα	76	52	70	198
	Αναμ. Συχν. υπό την ανεξαρτησία	94,6	53,5	49,9	198
Νότος	Συχνότητα	42	8	6	56
	Αναμ. Συχν. υπό την ανεξαρτησία	26,8	15,1	14,1	56
Δύση	Συχνότητα	34	31	4	69
	Αναμ. Συχν. υπό την ανεξαρτησία	33	18,6	17,4	69
Ανατολή	Συχνότητα	9	0	5	14
	Αναμ. Συχν. υπό την ανεξαρτησία	6,7	3,8	3,5	14
Σύνολο	Συχνότητα	161	91	85	337
	Αναμ. Συχν. υπό την ανεξαρτησία	161	91	85	337

Πίνακας 8: Είδος πετρώματος - Προσανατολισμός εδάφους

Η τιμή της ελεγχοσυνάρτησης Pearson για τον παραπάνω έλεγχο είναι $X^2 = 52,224$ με 6 βαθμούς ελευθερίας. Το παρατηρούμενο επίπεδο σημαντικότητας είναι σχεδόν μηδέν. Έτσι απορρίπτεται η μηδενική υπόθεση. Δηλαδή απορρίπτεται η υπόθεση της ανεξαρτησίας των μεταβλητών είδος πετρώματος και προσανατολισμός του δείγματος. Παρατηρώντας στον πίνακα συνάφειας τις παρατηρούμενες συχνότητες και τις αναμενόμενες, κάτω από την υπόθεση της τέλειας ανεξαρτησίας, προκύπτει ότι ο φλύσχης εμφανίζεται συστηματικά περισσότερο δυτικά και λιγότερο στα ανατολικά από ότι θα συνέβαινε εάν υπήρχε πλήρης ανεξαρτησία. Ο ασβεστόλιθος εμφανίζεται περισσότερο στις περιοχές με νότια έκθεση, ενώ ο ραδιολαρίτης περισσότερο στις βόρειες πλευρές του βουνού και πολύ λιγότερο, από ότι θα αναμενόταν (εάν υπήρχε πλήρης ανεξαρτησία), στις νότιες και δυτικές πλευρές.

3.2.5 Έλεγχος ανεξαρτησίας χ^2 , μεταξύ, υψομέτρου εδάφους και προσανατολισμού δειγματοληψίας.

Στον παρακάτω πίνακα παρουσιάζονται οι παρατηρούμενες συχνότητες και εοι αναμενόμενες συχνότητες για κάθε κελί (i,j) για την i κλάση του υψομέτρου, και τον

προσανατολισμό. Μέσω αυτού ελέγχονται η ανεξαρτησία των δυο χαρακτηριστικών. Οι υποθέσεις είναι:

H₀ : Ο προσανατολισμός του εδάφους είναι ανεξάρτητος από το υψόμετρο που έγινε η δειγματοληψία.

H₁ : Ο προσανατολισμός του εδάφους δεν είναι ανεξάρτητος από το υψόμετρο που έγινε η δειγματοληψία.

Ο πίνακας συνάφειας είναι:

Κλάσεις Υψομέτρου	Προσανατολισμός				
	Βορράς	Νότος	Δύση	Ανατολή	Σύνολο
270-370	Συχνότητα	2	2	1	0
	Αναμ.Συχν. υπό την ανεξαρτησία	2,9	0,8	1	0,2
370-470	Συχνότητα	7	7	8	0
	Αναμ.Συχν. υπό την ανεξαρτησία	12,9	3,7	4,5	0,9
470-570	Συχνότητα	5	1	5	0
	Αναμ.Συχν. υπό την ανεξαρτησία	6,5	1,8	2,3	0,5
570-670	Συχνότητα	28	3	3	0
	Αναμ.Συχν. υπό την ανεξαρτησία	20	5,6	7	1,4
670-770	Συχνότητα	18	2	5	0
	Αναμ.Συχν. υπό την ανεξαρτησία	14,7	4,2	5,1	1
770-870	Συχνότητα	40	0	8	1
	Αναμ.Συχν. υπό την ανεξαρτησία	28,8	8,1	10	2
870-970	Συχνότητα	36	7	10	2
	Αναμ.Συχν. υπό την ανεξαρτησία	38,8	11	13,5	2,7
970-1070	Συχνότητα	22	7	10	2
	Αναμ.Συχν. υπό την ανεξαρτησία	24,1	6,8	8,4	1,7
1070-1170	Συχνότητα	12	8	7	3
	Αναμ. Συχν. υπό την ανεξαρτησία	17,6	5	6,1	1,2
1170-1270	Συχνότητα	14	7	1	1
					23

Κλάσεις Υψομέτρου	Προσανατολισμός				
	Βορράς	Νότος	Δύση	Ανατολή	Σύνολο
Αναμ.Συχν. υπό την ανεξαρτησία	13,5	3,8	4,7	1	23
1270-1370	Συχνότητα	9	5	1	16
	Αναμ.Συχν. υπό την ανεξαρτησία	9,4	2,7	3,3	16
1370-1470	Συχνότητα	3	4	1	8
	Αναμ.Συχν. υπό την ανεξαρτησία	4,7	1,3	1,6	8
1470-1590	Συχνότητα	2	3	2	7
	Αναμ.Συχν. υπό την ανεξαρτησία	4,1	1,2	1,4	7
Σύνολο	Συχνότητα	198	56	69	337
	Αναμ.Συχν. υπό την ανεξαρτησία	198	56	69	337

Πίνακας 9: Υψόμετρο - Προσανατολισμός εδάφους

Η τιμή της ελεγχοσυνάρτησης X^2 του Pearson, για τον παραπάνω έλεγχο είναι $X^2=74,186$ με 36 βαθμούς ελευθερίας. Το παρατηρούμενο επίπεδο σημαντικότητας του παραπάνω ελέγχου είναι σχεδόν μηδέν, μικρότερο από κάθε άλλο σύνηθες επίπεδο σημαντικότητας. Αυτό σημαίνει ότι απορρίπτεται η αρχική υπόθεση της ανεξαρτησίας των δυο μεταβλητών. Εξετάζοντας στον παραπάνω πίνακα, τις παρατηρούμενες συχνότητες και τις αναμενόμενες συχνότητες, παρατηρείται ότι στην ανατολική έκθεση δεν έγιναν δειγματοληψίες σε μικρά υψόμετρα. Αυτό οφείλεται στο ότι ο Ερύμανθος ανατολικά συνδέεται με άλλο ορεινό σύμπλεγμα και έτσι δεν υπάρχουν μικρά υψόμετρα. Ενώ προς τις άλλες εκθέσεις μπορούν να γίνουν δειγματοληψίες σε όλο το φάσμα των υψομέτρων. Στην δυτική έκθεση, λαμβάνονται περισσότερες δειγματοληψίες στα μικρά υψόμετρα και λιγότερες στα μεγάλα, από όσες θα γίνονταν εάν υπήρχε πλήρης ανεξαρτησία. Ακόμη και στην νότια έκθεση παρατηρείται ότι λαμβάνονται περισσότερες δειγματοληψίες στα μεγάλα υψόμετρα από όσες θα γίνονταν εάν υπήρχε ανεξαρτησία.

3.2.6 Έλεγχος ανεξαρτησίας χ^2 , μεταξύ, κλίσης εδάφους και προσανατολισμού δειγματοληψίας.

Ανάλογα με τα παραπάνω, στον επόμενο πίνακα συνάφειας παρουσιάζονται οι παρατηρούμενες συχνότητες και οι αναμενόμενες συχνότητες για κάθε κελί (i,j) για την i κλάση της κλίσης του εδάφους, και τον j προσανατολισμό. Οι υποθέσεις του ελέγχου είναι:

H_0 : Ο προσανατολισμός του εδάφους είναι ανεξάρτητος από την κλίση του εδάφους που έγινε η δειγματοληψία.

H_1 : Ο προσανατολισμός του εδάφους δεν είναι ανεξάρτητος από την κλίση του εδάφους.

Ο πίνακας συνάφειας είναι:

	Κλίση εδάφους σε μοίρες	Προσανατολισμός				
		Βορράς	Νότος	Δύση	Ανατολή	Σύνολο
5-15	Συχνότητα	4	5	2	4	15
	Αναμ. Συχν. υπό την Ανεξαρτησία	8,8	2,5	3,1	0,6	15
15-25	Συχνότητα	7	2	8	1	18
	Αναμ. Συχνότητα υπό την ανεξαρτησία	10,6	3	3,7	0,7	18
25-35	Συχνότητα	34	15	24	4	77
	Αναμ. Συχνότητα υπό την ανεξαρτησία	45,2	12,8	15,8	3,2	77
35-45	Συχνότητα	66	18	20	1	105
	Αναμ. Συχνότητα υπό την ανεξαρτησία	61,7	17,4	21,5	4,4	105
45-55	Συχνότητα	61	13	10	2	86
	Αναμ. Συχνότητα υπό την ανεξαρτησία	50,5	14,3	17,6	3,6	86
55-67	Συχνότητα	26	3	5	2	36
	Αναμ. Συχνότητα υπό την ανεξαρτησία	21,2	6	7,4	1,5	36
Σύνολο	Συχνότητα	198	56	69	14	337
	Αναμ. Συχνότητα υπό την ανεξαρτησία	198	56	69	14	337

Πίνακας 10: Προσανατολισμός - Κλίση εδάφους

Για τον έλεγχο ανεξαρτησίας χ^2 η τιμή της ελεγχοσυνάρτησης είναι $X^2 = 50,983$ με 15 βαθμούς ελευθερίας. Το παρατηρούμενο επίπεδο σημαντικότητας του παραπάνω ελέγχου είναι σχεδόν μηδέν, μικρότερο από κάθε σύνηθες επίπεδο σημαντικότητας. Αυτό σημαίνει ότι απορρίπτεται η αρχική υπόθεση της ανεξαρτησίας των δυο μεταβλητών. Παρατηρώντας τον πίνακα συνάφειας, των μεταβλητών που παρουσιάζουν κάποιο είδος εξάρτησης προκύπτει ότι η δυτική πλευρά έχει μικρές κλίσεις, περισσότερο από ότι θα συνέβαινε εάν οι μεταβλητές ήταν ανεξάρτητες. Στον προηγούμενο έλεγχο παρατηρήθηκε ότι η δυτική πλευρά του Ερυμάνθου εμφανίζει περισσότερα χαμηλά υψόμετρα από ότι αναμενόταν. Ακόμη παρατηρήθηκε ότι στα δυτικά υπάρχει πέτρωμα του είδους φλύσχη, ένα πέτρωμα που θρυμματίζεται και διαβρώνεται εύκολα και συνήθως δεν δημιουργεί μεγάλες κλίσεις. Στον βορρά παρατηρούνται μεγαλύτερες κλίσεις από ότι αναμενόταν, εάν υπήρχε ανεξαρτησία των μεταβλητών, ενώ στην βόρια πλευρά του όρους υπάρχει περισσότερο του αναμενόμενου, ραδιολαρήτης.

ΚΕΦΑΛΑΙΟ 4. Φυτά

4.1. Γενικά

Γραμμική παλινδρόμηση.

Είναι πολύ σημαντικό και χρήσιμο να υπάρχει μια σχέση που να περιγράφει την παρατηρούμενη εξάρτηση δυο μεταβλητών. Η απλούστερη είναι η γραμμική σχέση πρώτου βαθμού. Στο σημείο αυτό θα περιγραφεί πώς η σχέση μιας ευθείας γραμμής μπορεί να βρεθεί με την μέθοδο των ελαχίστων τετραγώνων, όταν υπάρχει ένα σετ δεδομένων (X_i, Y_i) με $i=1,2,3\dots n$. όπου X: η ανεξάρτητη μεταβλητή (independent variable) ενώ Y: η εξαρτημένη μεταβλητή (dependent variable). Οι δυο μεταβλητές είναι συνεχείς. σ' αυτήν την περίπτωση.

Στόχος είναι να βρεθεί η ευθεία που να περιγράφει καλύτερα την σχέση των μεταβλητών αυτών. Υποθέτουμε ότι υπάρχει και είναι $Y = a + \beta X$.

Μια μέθοδος εκτίμησης της ευθείας παλινδρόμησης είναι η μέθοδος ελάχιστων τετραγώνων. Η μέθοδος φαίνεται να ανακαλύφθηκε ανεξάρτητα από τους Carl Friedrich Gauss (1777-1855) και Adrien Marie Legendre (1752-1833). Η λογική της μεθόδου, συμφωνά με Draper & Smith (1996), Panaretto I. (2001) και Kleinbaum D. et al. (1998) είναι η εξής: Έστω δυο στοχαστικά μεγέθη X, Y που θεωρείται ότι συνδέονται με κάποια μορφή εξάρτησης υποθέτουμε ότι αυτή είναι γραμμική, δηλαδή υπάρχει μοντέλο μορφής $Y = a + \beta X$ που την περιγράφει. Για κάποιο δείγμα θα είναι: $Y_i = a + \beta X_i + \varepsilon_i$ αφού για κάθε σταθερή τιμή του X_i η ποσότητα ε_i , είναι η τιμή του σφάλματος του ζεύγους (X_i, Y_i) δηλαδή η τιμή της προσαύξησης μέσω της οποίας κάποια παρατήρηση Y_i πέφτει έξω από την γραμμή της παλινδρόμησης.

Βέβαια οι τιμές ε_i δεν μπορούν να παρατηρηθούν αφού οι συντελεστές a, β είναι άγνωστοι. Για τον λόγο αυτό οι τιμές των παραμέτρων πρέπει να εκτιμηθούν μέσω του δείγματος έστω ότι με \hat{a} , $\hat{\beta}$ συμβολίζονται οι εκτιμήσεις που προκύπτουν από το συγκεκριμένο δείγμα. Τότε οι παρατηρούμενες τιμές των σφαλμάτων e_i θα δίνονται από τον τύπο. $e_i = Y_i - (\hat{a} + \hat{\beta} X_i)$, $i=1,2,3\dots n$

Θα πρέπει όμως να προσδιορισθούν οι τιμές \hat{a} και $\hat{\beta}$ έτσι ώστε να θεωρούνται ικανοποιητικές προσεγγίσεις των a και β . Αυτό θα γίνει με την μέθοδο των ελαχίστων

τετραγώνων. Η μέθοδος αυτή στηρίζεται στην ελαχιστοποίηση των σφαλμάτων, δηλαδή την ελαχιστοποίηση της απόκλισης της παρατηρούμενης τιμής Y_i και της προβλεπόμενης, από το μοντέλο τιμής \hat{Y}_i . Το άθροισμα των τετραγώνων των αποκλίσεων είναι $S = \sum (Y_i - \hat{a} - \hat{\beta} \cdot X_i)^2 = f(\hat{a}, \hat{\beta})$. Βρίσκοντας τα ακρότατα της παραπάνω συνάρτησης των δυο μεταβλητών, και μέσω των κανονικών εξισώσεων προκύπτουν τα εξής:

$$\hat{a} = \bar{Y} - \hat{\beta} \cdot \bar{X} \quad \text{και} \quad \hat{\beta} = \frac{\sum (X_i - \bar{X}) \cdot (Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$

Τα \hat{a} και $\hat{\beta}$ ονομάζονται εκτιμήσεις ελαχίστων τετραγώνων των συντελεστών γραμμικής παλινδρόμησης. Οι εκτιμητές \hat{a} και $\hat{\beta}$ που υπολογίζονται με τους παραπάνω τύπους δίνουν σημειακές εκτιμήσεις των α και β . Για να υπάρχει μεγαλύτερη ακρίβεια των εκτιμήσεων θα πρέπει να κατασκευαστούν διαστήματα εμπιστοσύνης αυτών. Προκειμένω να γίνει οποιουδήποτε είδους συμπερασματολογία όμως θα πρέπει να είναι γνωστές οι κατανομές των εκτιμητών των μεταβλητών α , β . Η συνήθης υπόθεση για την κατανομή των σφαλμάτων είναι η κανονική.

Υποθέτουμε λοιπόν ότι τα ε_i ακολουθούν την $N(0, \sigma^2)$ για $i=1,2,\dots,n$, δηλαδή κανονική κατανομή γύρω από το μηδέν με σταθερή διακύμανση.

Η βασική υπόθεση για την ισχύ της μεθόδου είναι η ανεξαρτησία των ε_i . Κάποιες από τις παραπάνω υποθέσεις είναι αναγκαίες (όπως η ανεξαρτησία) και κάποιες είναι λιγότερο περιοριστικές, (όπως η κανονικότητα) με την έννοια ότι μπορούν είτε να αντικαταστούν, είτε να αντιμετωπιστούν μετασχηματίζοντας τα δεδομένα.

Εάν πληρούνται τα παραπάνω αποδεικνύεται ότι οι κατανομές είναι:

$$\hat{\alpha} \sim N(\alpha, \sigma^2 \left(\frac{1}{n} + \frac{\bar{X}^2}{S_{xx}} \right)) \quad \hat{\beta} \sim N(\beta, \frac{\sigma^2}{S_{xx}}) \quad \text{όπου} \quad S_{xx} = \sum_{i=1}^n (X_i - \bar{X})^2$$

Τώρα είναι δυνατό να κατασκευαστούν διαστήματα εμπιστοσύνης καθώς και έλεγχοι υποθέσεων των παραμέτρων α , β της παλινδρόμησης.

Proportional Odds

Στο συγκεκριμένο κεφάλαιο υπάρχει ενδιαφέρον να εντοπιστεί και να διερευνηθεί η σχέση της πληθοκάλυψης των φυτών του όρους Ερυμάνθου με κάποιες από τις περιβαλλοντικές παραμέτρους που έχουν καταγράφει στην έρευνα. Καθώς και η ισχύ μιας τέτοιας σχέσης. Ως εξαρτημένη μεταβλητή θεωρείται η πληθοκάλυψη των φυτών, η οποία είναι κατηγορική μεταβλητή, με τιμές 0,1,2,3,4,5. Στην περίπτωση αυτή το μοντέλο της

γραμμικής παλινδρόμησης δεν προσαρμόζεται ικανοποιητικά εξαιτίας του ότι η γραμμική παλινδρόμηση παρουσιάζεται ευαίσθητη στην κατηγοριοποίηση. Αυτό σημαίνει ότι εάν κάποιο μοντέλο γραμμικής παλινδρόμηση με μια κατηγοριοποίηση, συγκριθεί με ένα μοντέλο που προήλθε από ενοποίηση δυο γειτονικών και όμοιων κατηγοριών, θα διαφέρει σημαντικά απ' αυτό, ενώ δεν θα έπρεπε. Για τον λόγο αυτό χρησιμοποιούνται εναλλακτικοί τρόποι διασύνδεσης μιας κατηγορικής εξαρτημένης μεταβλητής και κάποιων ανεξάρτητων συνεχών, ή διακριτών, ή και τα δυο ειδών.

Τα μοντέλα που ικανοποιούν τις ανάγκες της έρευνας σε αυτήν την περίπτωση είναι τα γενικευμένα γραμμικά μοντέλα. Είναι μια μεγάλη κατηγορία μοντέλων και προσφέρονται για ευρεία χρήση στατιστικών ερευνών.

- Όταν η εξαρτημένη μεταβλητή είναι δίτιμη τότε το γενικευμένο γραμμικό μοντέλο που χρησιμοποιείται είναι το γνωστό μοντέλο της λογιστικής παλινδρόμησης σύμφωνα με Hosmer & Lemeshow, (1989).
- Όταν η εξαρτημένη μεταβλητή παίρνει πεπερασμένο πλήθος διακριτών τιμών, όπως συμβαίνει στην περίπτωση αυτή. Τότε τα μοντέλα σύνδεσης είναι γνωστά ως Proportional Odds κατά τους Mc Cullagh (1980), A. Agresti,(1996) και Guisan & Harrell,(2000) και Scott S., et al. (1997). Αυτά τα μοντέλα θα εξεταστούν και θα αναλυθούν παρακάτω. Η μορφή τους είναι:

$$LinkP(Y \leq j / X = x_i) = a_j + B \cdot x_i.$$

$$P(Y \leq j / X = x_i) = \pi_1(x_i) + \pi_2(x_i) + \dots + \pi_j(x_i)$$

$$\pi_j(x_i) = P(Y = j / X = x_i), \quad j=0,1,2,\dots$$



Διευκρινήσεις:

Όπου link είναι η συνάρτηση σύνδεσης και είναι κάποια συνάρτηση των αθροιστικών πιθανοτήτων κάθε κατηγορίας έκβασης. Οι δυνατοί τρόποι σύνδεσης περιγράφονται παρακάτω. (Βιβλιογραφία: S.P.S.S. 11.0)

- Η **logit** με τύπο: $\log\left[\frac{P(Y \leq j)}{1 - P(Y \leq j)}\right]$, χρησιμοποιείται όταν η εξαρτημένη εμφανίζεται να έχει μια ομαλή κατανομή κατηγοριών.
- Η **Complementary log-log** με τύπο: $\log[-\log(1 - P(Y \leq j))]$, χρησιμοποιείται όταν η κατανομή της εξαρτημένης μεταβλητής εμφανίζει να έχει περισσότερες παρατηρήσεις στις υψηλότερες κατηγορίες.

- Η **Negative log-log** με τύπο: $-\log[-\log(P(Y \leq j))]$ χρησιμοποιείται όταν η κατανομή της εξαρτημένης μεταβλητής εμφανίζει να έχει περισσότερες παρατηρήσεις στις χαμηλότερες κατηγορίες.
- Η **probit** με τύπο: $\Phi^{-1}(P(Y \leq j))$, είναι η αντίστροφη συνάρτηση της πυκνότητας πιθανότητας της κανονικής κατανομής. Χρησιμοποιείται όταν η κατανομή του Y είναι περίπου κανονική.
- Η **Caucht** με τύπο: $\tan[\pi \cdot (P(Y \leq j))]$, χρησιμοποιείται όταν η κατανομή του Y είναι έχει ακραίες τιμές.

Εκτός από τις γενικές αυτές αρχές, δεν υπάρχει κάποιος θεωρητικός τρόπος επιλογής του link. Αλλά επιλέγεται εκείνο που δίνει την καλύτερη προσαρμογή στα δεδομένα (με δοκιμές).

a_j : είναι η βάση της j κατηγορίας.

$B = (\beta_1, \beta_2, \dots, \beta_p)$ οι συντελεστές παλινδρόμησης.

$x_i = (x_{1i}, x_{2i}, x_{3i}, \dots, x_{pi})$ οι τιμές πρόβλεψης του i διανύσματος

Το μοντέλο βασίζεται στο ότι υπάρχει κάποια αφανής συνεχής μεταβλητή, (στην συγκεκριμένη περίπτωση το ποσοστό επί τοις εκατό της πληθοκάλυψης των ειδών) της οποίας η ομαδοποίηση μας οδηγεί στην κατηγορική μεταβλητή. Σε πολλές περιπτώσεις έχει φυσική έννοια κάποια λανθάνουσα συνεχής μεταβλητή, αλλά ακόμη και όταν δεν έχει, μπορεί να χρησιμοποιηθούν τα μοντέλα αυτά και να έχουν ικανοποιητική προσαρμογή στα δεδομένα.

Όπως παρατηρείται στην βασική μορφή σύνδεσης η σταθερά a , του μοντέλου εξαρτάται μόνο από την κατηγορία πρόβλεψης και όχι από τις τιμές πρόβλεψης του κάθε διανύσματος. Ακόμη, προϋπόθεση του μοντέλου είναι ότι το δεύτερο τμήμα $B \cdot x$, εξαρτάται μόνο από τις τιμές του x , και όχι από την ίδια την κατηγορία. Έτσι σε κάθε κατηγορία έκβασης j το διάνυσμα B είναι το ίδιο (υπόθεση που πρέπει να ελέγχεται).

Γραφικά η αθροιστική συνάρτηση πιθανότητας έχει μια σιγμοειδή μορφή, και το διάνυσμα B επιδρά στην ένταση της μεταβολής της σιγμοειδούς καμπύλης. Επειδή η τιμή του B είναι ίδια για κάθε κατηγορία έκβασης, οι σιγμοειδείς αυτές καμπύλες δημιουργούν ένα σύνολο "παραλλήλων" σιγμοειδών γραμμών.

Προσδιορισμός των συντελεστών. Maximum Likelihood Estimation

Ο υπολογισμός των a_j και του διανύσματος $B = (\beta_1, \beta_2, \dots, \beta_p)$ γίνεται με την μέθοδο μέγιστης πιθανοφάνειας και τον επαναληπτικό αλγόριθμο του Fisher ή Newton-Raphson.

Έλεγχος σημαντικότητας των συντελεστών. Wald test.

Η κατανομή του εκτιμητή μέγιστης πιθανοφάνειας \hat{B} για μεγάλο μέγεθος δείγματος ακολουθεί την πολυμεταβλητή κανονική κατανομή

$$\hat{B} \sim N(B, \Sigma_B) \text{ όπου } \Sigma_B \text{ ο πίνακας διακύμανσης συνδιακύμανσης των } \beta_i$$

Από αυτό προκύπτει ότι $\hat{\beta}_i \sim N(\beta_i, \sigma_i^2)$ για κάθε $i=1,2,3,\dots,p$. Από την γνώση της κατανομής των $\hat{\beta}_i$ μπορεί να γίνει ο έλεγχος:

$$H_0 : \beta_i = 0$$

$$H_1 : \beta_i \neq 0$$

Η συνάρτηση ελέγχου, κάτω από την μηδενική υπόθεση είναι: $\left(\frac{\hat{\beta}_i}{ASE(\hat{\beta}_i)} \right)^2 \sim \chi^2_1$.

Όπου $\hat{\beta}_i$ η εκτίμηση της πραγματικής τιμής β_i . Με $ASE(\hat{\beta}_i)$ συμβολίζεται το ασυμπτωτικό τυπικό σφάλμα του β_i . Η φυσική σημασία του ελέγχου είναι εάν η μεταβλητή X_i είναι στατιστικά σημαντική για το μοντέλο.

Ακόμη ελέγχεται εάν το μοντέλο που περιέχει τις ανεξάρτητες μεταβλητές, έχει καλύτερη προσαρμογή στα δεδομένα, από το μοντέλο που δεν περιέχει καμία επεξηγηματική μεταβλητή. Ουσιαστικά ελέγχεται εάν οι ανεξάρτητες μεταβλητές είναι στατιστικά σημαντικές. Η μηδενική υπόθεση του ελέγχου είναι: $H_0 : B=0$ με εναλλακτική υπόθεση: $H_1 : B \neq 0$.

Ο έλεγχος αυτός γίνεται υπολογίζοντας τις τιμές των Πιθανοφανειών των δύο συγκρινόμενων μοντέλων, ή για την ακρίβεια την $-2\log(\text{likelihood})$ των δύο μοντέλων. Η συνάρτηση του παραπάνω ελέγχου είναι:

$$G = -2\log(\text{likelihood}) \quad m_1 - 2\log(\text{likelihood}) \quad m_2$$

m_1 : Το μοντέλο το οποίο δεν έχει ανεξάρτητες μεταβλητές.

m_2 : Το μοντέλο που χρησιμοποιείται.

Αυτή ακολουθεί την χ^2_v κατανομή με $v=p-q$ βαθμούς ελευθερίας. Όπου p ο αριθμός των παραμέτρων το m_2 μοντέλου και q ο αριθμός των παραμέτρων του m_1 μοντέλου.

Έτσι, ελέγχεται εάν το μοντέλο δίνει στατιστικά καλύτερη πρόβλεψη από το μοντέλο που δεν περιέχει ανεξάρτητες μεταβλητές.

Πρέπει να αναφερθεί ότι η ελεγχοσυνάρτηση βασίζεται στην χ^2 κατανομή και οι έλεγχοι με την κατανομή αυτή επηρεάζονται πολύ από τα άδεια κελιά. Τέοια κελιά δημιουργούνται πολύ συχνά όταν οι ανεξάρτητες μεταβλητές είναι συνεχείς. Για τον λόγο αυτόν γενικά συμπεράσματα τέτοιου ελέγχου πρέπει να εξάγονται με προσοχή.

Έλεγχος καλής προσαρμογής του μοντέλου. Deviance test

Θα πρέπει να ελεγχθεί εάν οι προβλέψεις του μοντέλου είναι ικανοποιητικές, δηλαδή εάν συμβαδίζουν με τα δεδομένα σύμφωνα με τους Venables W. N. & Ripley B.D. (1994) και Jongman R. H. G. et al (1995).

Οι υποθέσεις που ελέγχονται είναι:

H_0 : Το μοντέλο έχει καλή προσαρμογή.

H_1 : Αντίθετα.

Η ελεγχοσυνάρτηση Deviance του παραπάνω ελέγχου, αποδεικνύεται ότι είναι: $D=-2[L_M-L_S]=-2\log(\text{likelihood})_M+2\log(\text{likelihood})_S$ με L_M , L_S οι τιμές των μέγιστων πιθανοφανειών του εκτιμώμενου μοντέλου και του κορεσμένου μοντέλου αντίστοιχα. Η D ακολουθεί την χ^2 κατανομή με p-q βαθμούς ελευθερίας. Όπου p ο αριθμός των παραμέτρων του κορεσμένου μοντέλου, στο οποίο αντιστοιχεί μια παράμετρος σε κάθε παρατήρηση και q ο αριθμός των παραμέτρων του υπό εξέταση μοντέλου. Ο έλεγχος καλής προσαρμογής είναι πολύ χρήσιμος όταν υπάρχει μικρός αριθμός ανεξάρτητων μεταβλητών και αυτό επειδή η κατανομή της παραπάνω ελεγχοσυνάρτησης είναι χ^2 , η οποία όπως προαναφέρθηκε επηρεάζεται πολύ από τα άδεια κελιά, έτσι και αυτός ο έλεγχος θα πρέπει να αντιμετωπίζεται ιδιαίτερα προσεκτικά.

Ένας άλλος τρόπος για να ελεγχθεί η προσαρμογή του μοντέλου στα δεδομένα, είναι μέσω των Pseudo R-square που είναι τα ανάλογα του R-square της γραμμική παλινδρόμησης. Αυτά μετρούν το ποσοστό της διασποράς της εξαρτημένης μεταβλητής που οφείλεται στις ανεξάρτητες μεταβλητές. Στα κατηγορικά (ordinal) δεδομένα οι μετρήσεις αυτές βασίζονται σε σχέσεις πιθανοφάνειας και όχι σε κατάλοιπα. Υπάρχουν τρεις μέθοδοι υπολογισμού των Pseudo R-Square

- Cox and Snell, η μέγιστη τιμή της όμως είναι μικρότερη από 1.
- Nagelkerke, είναι κάποια τροποποίηση της προηγούμενης ώστε να δίνει μέγιστη τιμή 1.
- McFadden.

Τεστ παραλλήλων γραμμών.

Η μεθοδολογία που αναπτύχθηκε στηρίζεται στην υπόθεση ότι οι καμπύλες των αθροιστικών πιθανοτήτων είναι «παράλληλες», δηλαδή το εκτιμώμενο μοντέλο έχει το ίδιο σετ συντελεστών για κάθε κατηγορία έκβασης. Αυτή η υπόθεση όμως πρέπει να ελέγχεται. Οι υποθέσεις του ελέγχου αυτού είναι:

H₀ : Οι γραμμές είναι παράλληλες. Όλες οι κατηγορίες έχουν το ίδιο σετ συντελεστών των ανεξάρτητων μεταβλητών.

H₁ : Αντίθετα.

Η ελεγχοσυνάρτηση είναι: $G = -2 \log(\text{likelihood})m_1 + 2 \log(\text{likelihood})m_2$

m_1 : Το μοντέλο του οποίου το διάνυσμα B είναι ίδιο σε κάθε κατηγορία.

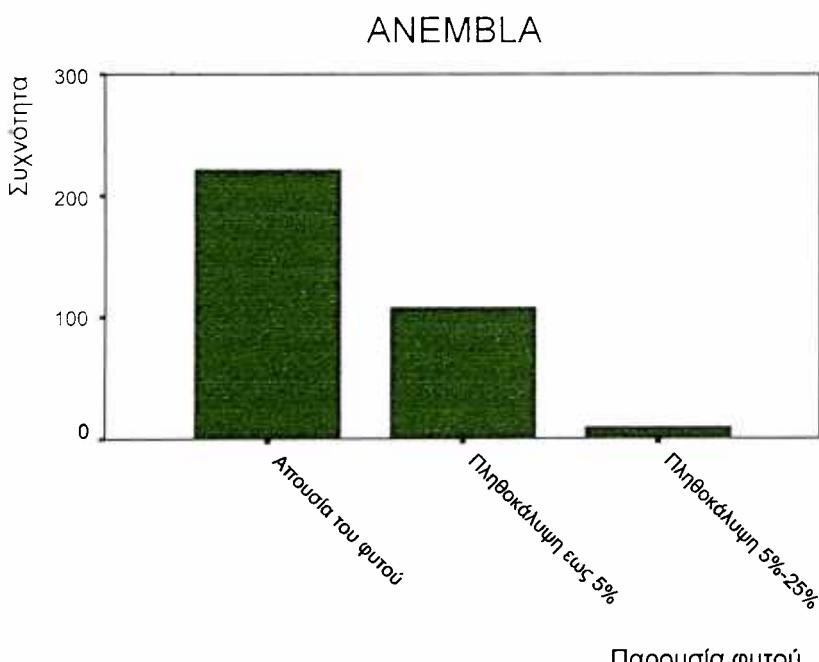
m_2 : Το μοντέλο που εκτιμήθηκε με βάση τα δεδομένα.

Η οποία ακολουθεί την χ^2 κατανομή με $v = p - q$ βαθμούς ελευθερίας. Όπου p ο αριθμός των παραμέτρων το m_2 μοντέλου και q ο αριθμός των παραμέτρων του m_1 μοντέλου.

Όλες οι αναλύσεις πραγματοποιούνται με το στατιστικό πρόγραμμα SPSS 10.0

4.2. Anemone Blanda (Anembla) Ποώδες φυτό.

Προκειμένου να βρεθεί η κατανομή της πληθοκάλυψης του εδάφους από το συγκεκριμένο φυτό, κατασκευάζεται το παρακάτω διάγραμμα.



Γράφημα 5: Ραβδόγραμμα συχνοτήτων της πληθοκάλυψης του εδάφους από το φυτό Anemone Blanda

Στον παρακάτω πίνακα υπάρχει μια σαφέστερη εικόνα της κατανομής της μεταβλητής.

Εμφάνιση φυτού	Συχνότητα	Σχετική συχνότητα (%)	Αθροιστική Σχετική Συχνότητα (%)
Απουσία του φυτού από το δείγμα	221	65,6	65,6
Παρουσία του φυτού στη Δειγματοληψία και πληθοκάλυψη του εδάφους έως 5%	107	31,8	97,3
Παρουσία του φυτού στη Δειγματοληψία και πληθοκάλυψη του εδάφους από 5% έως 25%	9	2,7	100
Σύνολο	337	100,0	

Πίνακας 11: Κατανομή της πληθοκάλυψης του εδάφους από το φυτό *Anemone Blanda*.

Από τα παραπάνω παρατηρείται ότι στο 65,6% των δειγματοληψιών δεν παρατηρήθηκε το συγκεκριμένο φυτό, ενώ στο 32,8% των δειγματοληψιών που βρέθηκε το φυτό, χωρίς όμως να δημιουργεί μεγάλες πληθοκαλύψεις στο έδαφος.

Θα αναζητηθεί ένας τρόπος σύνδεσης της μεταβλητής «πληθοκάλυψη του εδάφους», που θεωρείται ως εξαρτημένη μεταβλητή και του υψομέτρου. Ως καταλληλότερη συνάρτηση σύνδεσης εδώ είναι η **Complementary log-log**.

Κατά την επεξεργασία των δεδομένων στο SPSS λαμβάνεται ο πίνακας των εκτιμήσεων των παραμέτρων του μοντέλου.

$$\log(-\log(1 - P(Y \leq j))) = \hat{\alpha}_j + \hat{\beta} \cdot x \text{ με } j=0,1.$$

	Εκτίμηση	Τυπική απόκλιση	Τιμή (Wald)	Βαθμοί Ελευθερίας	Επίπεδο Σημαντικότητας
$\hat{\alpha}_0$	0,843	0,223	14,322	1	0,000
$\hat{\alpha}_1$	2,123	0,253	70,576	1	0,000
$\hat{\beta}$	$8,539 \cdot 10^{-4}$	$0.238 \cdot 10^{-3}$	$\cdot 12,862$	1	0,000

Τα μοντέλα είναι: $\log(-\log(1 - P(Y \leq 0))) = 0,843 + 8,539 \cdot 10^{-4} \cdot x$

$$\log(-\log(1 - P(Y \leq 1))) = 2,123 + 8,539 \cdot 10^{-4} \cdot x$$

Από αυτά τα δυο μοντέλα υπολογίζονται οι πιθανότητες $P(Y=0/X)$, $P(Y=1/X)$, $P(Y=2/X)$.

Στον παραπάνω πίνακα βλέπουμε τις τιμές για το Wald test που ελέγχει:

$$H_0 : \beta_j = 0$$

$$H_1 : \beta_j \neq 0 \quad \text{με } j=0,1.$$

Η τιμή της ελεγχοσυνάρτησης είναι $W=12,862$ με παρατηρούμενο επίπεδο σημαντικότητας σχεδόν μηδέν. Αυτό σημαίνει ότι απορρίπτεται η μηδενική υπόθεση. Δηλαδή, απορρίπτεται ότι η ανεξάρτητη μεταβλητή (υψόμετρο) δεν είναι στατιστικά σημαντική.

Ακόμη για τον έλεγχο:

$$H_0 : a_j = 0$$

$$H_1 : a_j \neq 0 \quad \text{με } j=0,1.$$

Οι αντίστοιχες Wald τιμές είναι, $W=14,322$ και $W=70,567$ και τα παρατηρούμενα επίπεδα σημαντικότητας και για τους δύο ελέγχους είναι σχεδόν μηδέν. Ετσι, απορρίπτονται οι μηδενικές υποθέσεις. Η φυσική σημασία αυτών των ελέγχων είναι ότι ο σταθερός όρος του μοντέλου για τις δύο κατηγορίες δεν είναι μηδέν.

Αξιολόγηση του μοντέλου.

Τα Pseudo R-Square των μοντέλων αυτών είναι:

Pseudo R-square	Τιμές
Cox and Snell	0,036
Nagelkerke	0,047
McFadden	0,025

Οι τιμές των Pseudo R-Square είναι ιδιαίτερα μικρές.

Αξιολόγηση του μοντέλου με βάση την προσαρμογή του στα δεδομένα:

Οι υποθέσεις που ελέγχονται είναι:

$$H_0 : \text{Το μοντέλο έχει καλή προσαρμογή.}$$

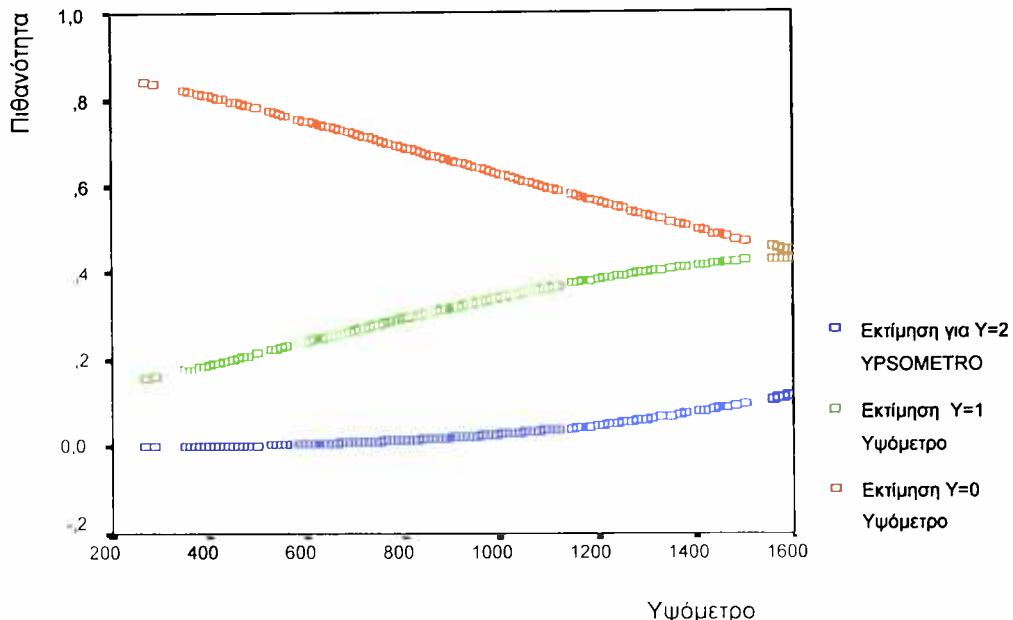
$$H_1 : \text{Αντίθετα.}$$

Τα αποτελέσματα του ελέγχου αυτού περιέχονται στον πίνακα:

Goodness-of-fit	Τιμή χ^2	Βαθμοί Ελευθερίας	Επίπεδο Σημαντικότητας
Deviance	159,014	195	0,972

Ο έλεγχος αυτός κατά Deviance έχει τιμή ελεγχοσυνάρτησης $G= 159.014$, που όπως έχει αναφερθεί η ελεγχοσυνάρτηση αυτή ακολουθεί την χ^2 κατανομή με 195 βαθμούς ελευθερίας. Το παρατηρούμενο επίπεδο σημαντικότητας του παραπάνω ελέγχου είναι $p\text{-value}=0,972$ μεγαλύτερο από κάθε σύνηθες επίπεδο σημαντικότητας, αυτό σημαίνει ότι

δεν υπάρχουν αρκετές ενδείξεις ώστε να απορρίπτεται η μηδενική υπόθεση. Άρα, δεν μπορεί να μην γίνει δεκτό ότι το μοντέλο έχει καλή προσαρμογή στα δεδομένα. Το μοντέλο αυτό λοιπόν μπορεί να χρησιμοποιηθεί για προβλέψεις. Οι προβλέψεις του μοντέλου αυτού όμως είναι όλες μηδέν, δηλαδή το μοντέλο προβλέπει απουσία του φυτού. Έχει όμως σημασία και η πιθανότητα με την οποία δίνει αυτές τις προβλέψεις. Αυτές οι πιθανότητες πρόβλεψης των κατηγοριών, ανάλογα με το υψόμετρο, μπορούν να παρασταθούν σχηματικά στο παρακάτω διάγραμμα:



Παρατηρείται ότι το μοντέλο αυτό προβλέπει πάντα ότι το φυτό (*Anemone Blanda*) θα απουσιάζει. Αυτό το προβλέπει με μεγάλη πιθανότητα στα μικρά υψόμετρα. Αλλά καθώς τα υψόμετρα αυξάνουν διατηρείται η ίδια πρόβλεψη με μικρότερη όμως πιθανότητα. Το μοντέλο δεν προκρίνει ποτέ την πρόβλεψη $Y=1$, δηλαδή την παρουσία του φυτού με πληθοκάλυψη έως 5%, αλλά ενώ στα χαμηλά υψόμετρα αυτό μπορεί να συμβεί με πιθανότητες μικρές γύρω στο 0,2 στα υψηλά υψόμετρα υπάρχει μεγαλύτερη πιθανότητα που πλησιάζει το 0,5. Η αύξηση αυτή της πιθανότητας είναι ανάλογη της αύξησης του υψομέτρου. Εξαιτίας αυτού το φυτό θα μπορούσε να χαρακτηριστεί ως υψόφιλο. Η πρόβλεψη για δημιουργία πληθοκάλυψης από 5% έως 25% στα χαμηλά υψόμετρα έχει πιθανότητα σχεδόν μηδέν ενώ στα υψηλά είναι λίγο μεγαλύτερη, διατηρώντας όμως αντικειμενικά πολύ μικρές τιμές. Αυτό σημαίνει ότι το φυτό αυτό προβλέπεται να συναντάται σπάνια και αυτό θα συμβαίνει σε υψηλά υψόμετρα.

Παρατηρείται επίσης ότι η $P(Y=1/X) + P(Y=2/X)$, δηλαδή η πιθανότητα παρουσίας του φυτού σε οποιαδήποτε μορφή είναι μεγαλύτερη από την $P(Y=0/X)$ την πιθανότητα της

απουσίας, σε ύψη μεγαλύτερα των 800 μέτρων. Δηλαδή σε αυτά τα ύψη αναμένεται να μην απουσιάζει το φυτό, ενώ το μοντέλο προκρίνει απουσία.

Ελεγχος υποθέσεων.

Το παραπάνω μοντέλο εκτιμήθηκε με την προϋπόθεση ότι η επίδραση της ανεξάρτητης μεταβλητής είναι ίδια σε κάθε επίπεδο παρατήρησης της εξαρτημένης μεταβλητής. Αυτή η υπόθεση όμως πρέπει να ελεγχθεί. Έτσι οι υποθέσεις του ελέγχου είναι:

H₀ : Όλες οι κατηγορίες έχουν το ίδιο σετ συντελεστών.

H₁ : Αντίθετα.

Οι τιμές του ελέγχου παρουσιάζονται στον επόμενο πίνακα.

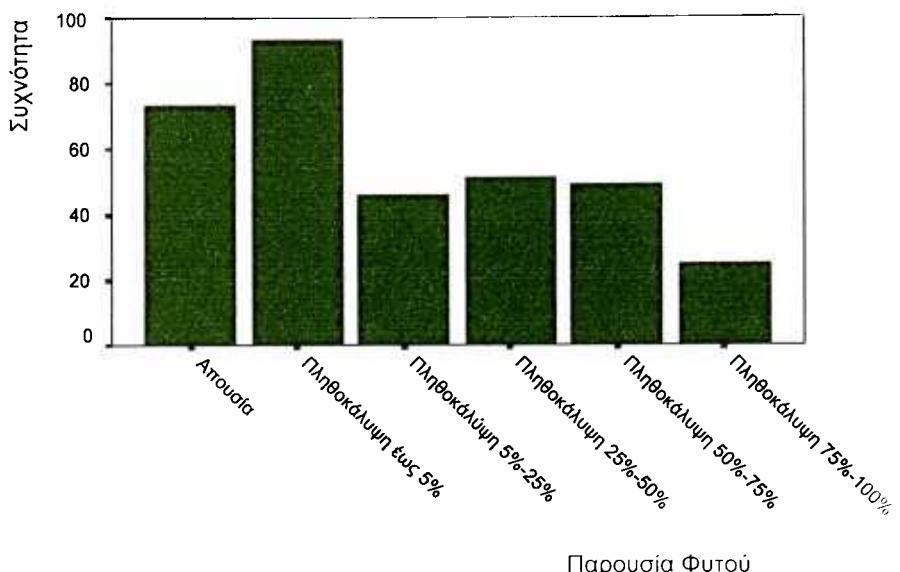
Μοντέλο	-2log-likelihood	Τιμή χ^2	Βαθμοί Ελευθερίας	Επίπεδο Σημαντικότητας
Null Hypothesis	263,614			
General	263,523	0,090	1	0,764

Η ελεγχοσυνάρτηση του ελέγχου αυτού όπως έχει αναφερθεί υπολογίζεται από την διαφορά των λογαρίθμων των Πιθανοφανειών. Η κατανομή της ελεγχοσυνάρτησης αυτής είναι η χ^2 με έναν βαθμό ελευθερίας στην περίπτωση αυτή. Η τιμή της ελεγχοσυνάρτησης είναι $G=0,090$. Το παρατηρούμενο επίπεδο σημαντικότητας είναι $p\text{-value}=0,764$ μεγαλύτερο από κάθε σύνηθες επίπεδο σημαντικότητας. Αυτό σημαίνει ότι δεν υπάρχουν αρκετές ενδείξεις ώστε να απορρίπτεται η μηδενική υπόθεση. Έτσι δεν μπορεί να μην γίνει δεκτό ότι οι σιγμοειδείς καμπύλες των αθροιστικών πιθανοτήτων είναι παράλληλες, δηλαδή ότι δεν υπάρχει ο ίδιος συντελεστής στην ανεξάρτητη μεταβλητή για κάθε επίπεδο της εξαρτημένης μεταβλητής.

4.3. *Quercus coccifera* (Quercoco) Δέντρο.

Κατασκευάζεται το διάγραμμα της μεταβλητής πληθοκάλυψη εδάφους από το φυτό *Quercus coccifera*. (Πρόκειται για το γνωστό φυτό πουρνάρι.)

QUERCOCO



Παρουσία Φυτού

Γράφημα 6: Ραβδόγραμμα συχνοτήτων της πληθοκάλυψης του εδάφους από το φυτό *Quercus coccifera*

Η κατανομή της παραπάνω μεταβλητής, διακρίνεται καλύτερα στον παρακάτω πίνακα συχνοτήτων.

Εμφάνιση φυτού	Συχνότητα	Σχετική συχνότητα (%)	Αθροιστική Σχετική Συχνότητα (%)
Απουσία του φυτού από το δείγμα	73	21,7	21,7
Παρουσία του φυτού στην δειγματοληψία και πληθοκάλυψη του εδάφους έως 5%	93	27,6	49,3
Παρουσία του φυτού στην δειγματοληψία και πληθοκάλυψη του εδάφους από 5% έως 25%	46	13,6	62,9
Παρουσία του φυτού στην δειγματοληψία και πληθοκάλυψη του εδάφους από 25% έως 50%	51	15,1	78,0
Παρουσία του φυτού στην δειγματοληψία και πληθοκάλυψη του εδάφους από 50% έως 75%	49	14,5	92,6
Παρουσία του φυτού στην δειγματοληψία και πληθοκάλυψη του εδάφους από 75% έως 100%	25	7,4	100
Σύνολο	337	100,0	

Πίνακας 12: Κατανομή της πληθοκάλυψης του εδάφους από το φυτό *Quercus coccifera*.

Παρατηρείται ότι στο 79% των δειγματοληψιών περιέχεται το συγκεκριμένο φυτό, ενώ στο 27% των δειγματοληψιών δίνει πληθοκάλυψη έως 5%. Ήα αναζητηθεί μοντέλο πρόβλεψης της πληθοκάλυψης του εδάφους από το συγκεκριμένο φυτό, χρησιμοποιώντας με ανεξάρτητη μεταβλητή μόνο το υψόμετρο. Η κατάλληλη συνάρτηση είναι η **Negative log-log**, γιατί είναι η μόνη που καλύπτει την προϋπόθεση της παραλλήλων γραμμών.

Οι εκτιμήσεις των συντελεστών του μοντέλου:

$$-\log(-\log(P(Y \leq j))) = \hat{\alpha}_j + \hat{\beta} \cdot x \text{ με } j=0,1,2,3,4.$$

υπάρχουν στον επόμενο πίνακα.

	Εκτίμηση	Τυπική απόκλιση	Τιμή (Wald)	Βαθμοί Ελευθερίας	Επίπεδο Σημαντικότητας
$\hat{\alpha}_0$	-2,062	0,234	77,644	1	0,000
$\hat{\alpha}_1$	-1,190	0,222	28,771	1	0,000
$\hat{\alpha}_2$	-0,736	0,222	10,941	1	0,001
$\hat{\alpha}_3$	$-8,52 \cdot 10^{-2}$	0,231	0,136	1	0,712
$\hat{\alpha}_4$	1,096	0,280	15,288	1	0,000
$\hat{\beta}$	$-1,72 \cdot 10^{-3}$	$0,248 \cdot 10^{-3}$	48,362	1	0,000

Έτσι το μοντέλο για κάθε κατηγορία της εξαρτημένης μεταβλητής είναι:

$$-\log(-\log(P(Y \leq 0))) = -2,062 - 1,72 \cdot 10^{-3} \cdot x$$

$$-\log(-\log(P(Y \leq 1))) = -1,190 - 1,72 \cdot 10^{-3} \cdot x$$

$$-\log(-\log(P(Y \leq 2))) = -0,736 - 1,72 \cdot 10^{-3} \cdot x$$

$$-\log(-\log(P(Y \leq 3))) = -8,52 \cdot 10^{-2} - 1,72 \cdot 10^{-3} \cdot x$$

$$-\log(-\log(P(Y \leq 4))) = 1,096 - 1,72 \cdot 10^{-3} \cdot x$$

Από τις εξισώσεις αυτές υπολογίζονται οι πιθανότητες κάθε κατηγορίας.

Στον πίνακα υπάρχουν οι έλεγχοι για την στατιστική σημαντικότητα των α_j και β με τα γνωστά Wald τεστ.

Δηλαδή για τον έλεγχο

$$H_0 : \alpha_j = 0$$

$$H_1 : \alpha_j \neq 0 \text{ με } j=0,1,2,3,4.$$

Οι αντίστοιχες Wald τιμές και τα παρατηρούμενα επίπεδα σημαντικότητας συμπεραίνουν ότι απορρίπτεται η μηδενική υπόθεση, εκτός από την $j=3$.

Ανάλογα για τον έλεγχο:

$$H_0 : \beta = 0$$

$$H_1 : \beta \neq 0$$

η φυσική σημασία του οποίου είναι εάν η επίδραση της ανεξάρτητης μεταβλητής είναι στατιστικά σημαντική. Όμοια και για τον έλεγχο αυτό, απορρίπτεται η μηδενική υπόθεση. Έτσι κατά την μοντελοποίηση του συγκεκριμένου φυτού με την μέθοδο που εξετάζεται, η μεταβλητή υψόμετρο είναι στατιστικά σημαντική.

Αξιολόγηση του μοντέλου.

Τα Pseudo R-square για το μοντέλο είναι:

Pseudo R-square	Τιμές
Cox and Snell	0,179
Negelkerke	0,185
McFadden	0,057

Οι τιμές των μέτρων αυτών είναι αρκετά μικρές.

Ο έλεγχος καλής προσαρμογής κατά Deviance ελέγχει τις υποθέσεις:

H_0 : Το μοντέλο έχει καλή προσαρμογή στα δεδομένα.

H_1 : Αντίθετα.

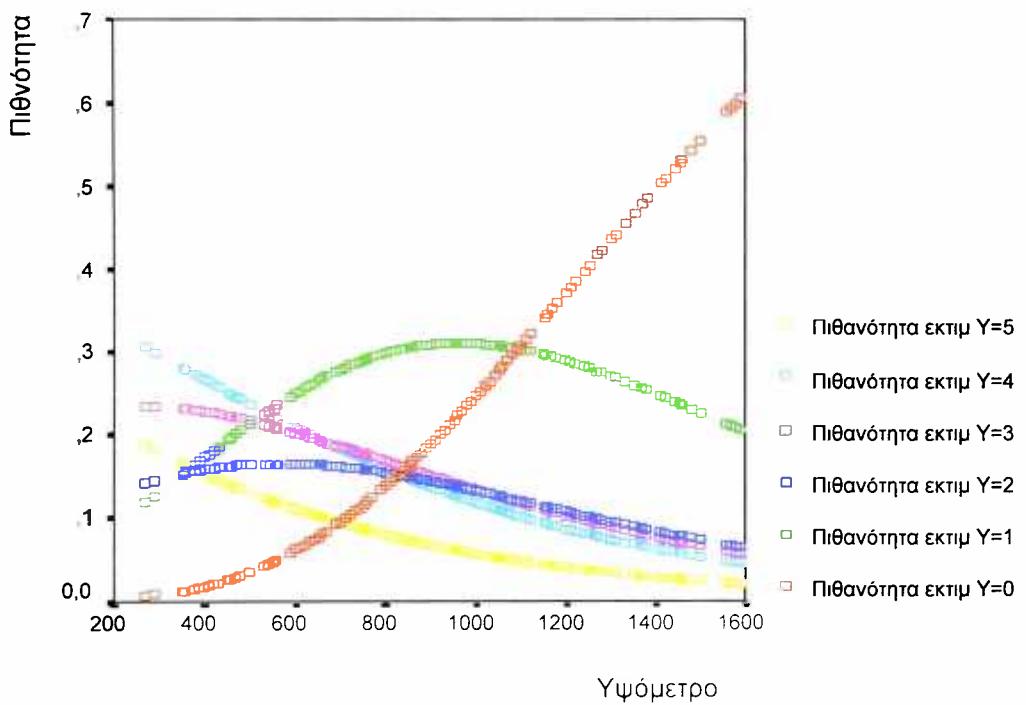
Goodness-of-fit	Τιμή χ^2	Βαθμοί Ελευθερίας	Επίπεδο Σημαντικότητας
Deviance	460,830	489	0,815

Η τιμή της συνάρτησης ελέγχου είναι $G=460,830$ με 489 βαθμούς ελευθερίας. Το παρατηρούμενο επίπεδο σημαντικότητας είναι $p\text{-value}=0,815$, μεγαλύτερο από κάθε σύνηθες επίπεδο σημαντικότητας. Άρα δεν υπάρχουν αρκετές ενδείξεις απόρριψης της μηδενικής υπόθεσης. Έτσι δεν μπορεί να μην γίνει δεκτό ότι το μοντέλο έχει καλή προσαρμογή. Το μοντέλο αυτό μπορεί να χρησιμοποιηθεί για προβλέψεις.

Δημιουργείται ο επόμενος πίνακας συνάφειας που δίνει τις παρατηρούμενες πληθοκαλύψεις και τις προβλεπόμενες από το μοντέλο πληθοκαλύψεις.

Προβλεπόμενες τιμές της πληθοκάλυψης από το μοντέλο	Παρατηρούμενες τιμές της πληθοκάλυψης στο δείγμα						Σύνολο
	0	1	2	3	4	5	
0	44	23	6	2	2	1	78
1	24	59	37	39	43	23	225
4	5	11	3	10	4	1	34
Σύνολο	73	93	46	51	49	25	337

Το μοντέλο προβλέπει μόνο τις τιμές 0, 1, 4 για την εξαρτημένη μεταβλητή. Έχει όμως ουσιαστική σημασία και η πιθανότητα με την οποία τις προκρίνει και η πιθανότητα με την οποία απορρίπτει τις άλλες κατηγορίες της εξαρτημένης μεταβλητής. Αυτές οι πιθανότητες παρατηρούνται στο επόμενο διάγραμμα.



Στο διάγραμμα αυτό διακρίνεται ότι υπάρχει μεγάλη πιθανότητα το φυτό (*Quercus coccifera*) να εμφανιστεί στα χαμηλά υψόμετρα ενώ στα υψηλά (μεγαλύτερα των 1000-1100 μέτρων) υπάρχει μεγάλη πιθανότητα να απουσιάζει. Σε υψόμετρα από 200-600 μέτρα υπάρχει πιθανότητα 0,2-0,3 να παρατηρηθούν πληθοκαλύψεις από 50% έως 75%. Σε μεγαλύτερα ύψη το φυτό φαίνεται να μην εξαπλώνεται αρκετά. Με αποτέλεσμα να είναι πιθανότερο, σε ύψη από 600 έως 1100 μέτρα, να υπάρχουν πληθοκαλύψεις έως 5%. Άρα το *Quercus coccifera* είναι φιλικότερο και αναπτύσσεται περισσότερο στα μικρά υψόμετρα.

Ελεγχος υποθέσεων.

Το παραπάνω μοντέλο εκτιμήθηκε με την προϋπόθεση ότι η επίδραση της ανεξάρτητης μεταβλητής είναι ίδια σε κάθε επίπεδο παρατήρησης της εξαρτημένης μεταβλητής. Δηλαδή τα μοντέλα αυτά έχουν σταθερή τιμή για το β σε όλα τα επίπεδα της μεταβλητής Y. Αυτή η υπόθεση όμως πρέπει να ελεγχθεί, έτσι οι υποθέσεις του ελέγχου είναι:

H_0 : Όλες οι κατηγορίες έχουν το ίδιο σετ συντελεστών.

H_1 : Αντίθετα.

Τα αποτελέσματα του ελέγχου αυτού είναι:

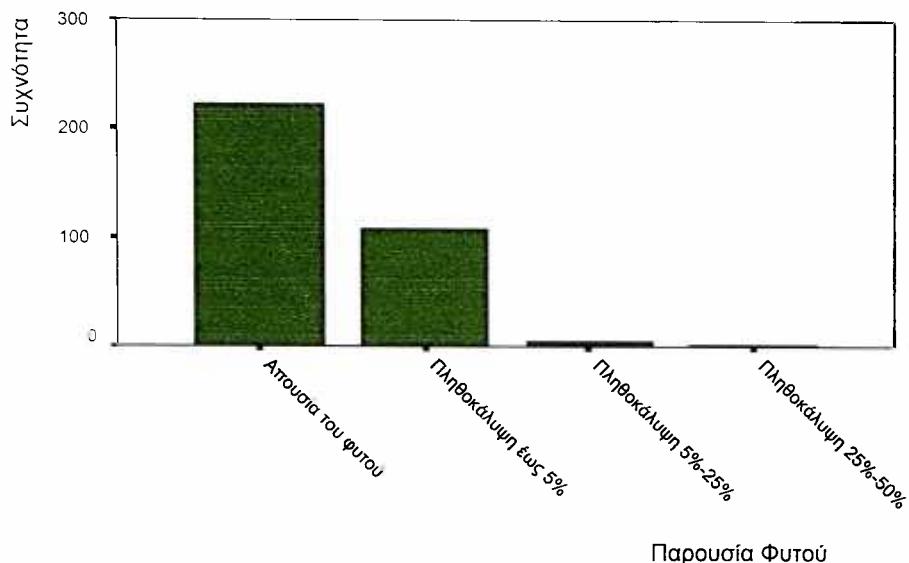
Μοντέλο	-2log-likelihood	Τιμή χ^2	Βαθμοί Ελευθερίας	Επίπεδο Σημαντικότητας
Null Hypothesis	678,207			
General	675,668	2,539	1	0,638

Η ελεγχοσυνάρτηση του ελέγχου αυτού όπως έχει αναφερθεί υπολογίζεται από την διαφορά των λογαρίθμων των πιθανοφανειών. Η κατανομή της ελεγχοσυνάρτησης αυτής είναι η χ^2 με έναν βαθμό ελευθερίας στην περίπτωση αυτή. Η τιμή της ελεγχοσυνάρτησης είναι G=2,539. Το παρατηρούμενο επίπεδο σημαντικότητας είναι p-value=0,638 μεγαλύτερο από κάθε σύνηθες επίπεδο σημαντικότητας. Αυτό σημαίνει ότι δεν υπάρχουν αρκετές ενδείξεις ώστε να απορρίπτεται η μηδενική υπόθεση. Έτσι δεν μπορεί να μην γίνει δεκτό ότι οι σιγμοειδείς καμπύλες των αθροιστικών πιθανοτήτων είναι παράλληλες. Δηλαδή ότι δεν υπάρχει ο ίδιος συντελεστής της ανεξάρτητης μεταβλητής για κάθε επίπεδο της εξαρτημένης.

4.4. Rosa canina (Rosacan) Θάμνος.

Μια πρώτη εικόνα της πληθοκάλυψης του εδάφους από το φυτό Rosa canina λαμβάνεται από το επόμενο διάγραμμα:

ROSACAN



Γράφημα 7: Ραβδόγραμμα συχνοτήτων της πληθοκάλυψης του εδάφους από το φυτό Rosa canina.

Ο επόμενος πίνακας συχνοτήτων περιγράφει ακριβέστερα την κατανομή αυτή.

Εμφάνιση φυτού	Συχνότητα	Σχετική συχνότητα (%)	Αθροιστική Σχετική Συχνότητα (%)
Απουσία του φυτού από το δείγμα	222	65,9	65,9
Παρουσία του φυτού στην δειγματοληψία και πληθοκάλυψη του εδάφους έως 5%	109	32,3	98,2
Παρουσία του φυτού στην δειγματοληψία και πληθοκάλυψη του εδάφους από 5% έως 25%	5	1,5	99,7
Παρουσία του φυτού στην δειγματοληψία και πληθοκάλυψη του εδάφους από 25% έως 50%	1	0,3	100
Σύνολο	337	100,0	

Πίνακας 13: Κατανομή της πληθοκάλυψης του εδάφους από το φυτό Rosa canina

Από τον πίνακα διακρίνεται ότι στο 65,9% των δειγματοληψιών δεν υπήρχε το συγκεκριμένο φυτό, στο υπόλοιπο 34,1% που παρατηρήθηκε, το 32,3% δημιουργούσε μικρές πληθοκαλύψεις έως 5%.

Το καταλληλότερο μοντέλο σύνδεσης από όσα κατασκευάστηκαν είναι το **logit**. Εξαρτημένη μεταβλητή είναι η πληθοκάλυψη του εδάφους και ανεξάρτητη μόνο το υψόμετρο αφού οι μεταβλητές κλίση εδάφους, είδος πετρώματος και έκθεση εδάφους ήταν

στατιστικά ασήμαντες κατά την δοκιμή των διαφόρων μοντέλων. Από την επεξεργασία των δεδομένων προκύπτουν οι εκτιμήσεις των παραμέτρων του μοντέλου:

$$\log\left[\frac{P(Y \leq j)}{1-P(Y \leq j)}\right] = \widehat{\alpha}_j + \widehat{\beta} \cdot x \text{ για } j=0,1,2.$$

	Εκτίμηση	Τυπική απόκλιση	Τιμή (Wald)	Βαθμοί Ελευθερίας	Επίπεδο Σημαντικότητας
$\widehat{\alpha}_0$	1,798	0,426	17,834	1	0,000
$\widehat{\alpha}_1$	5,178	0,593	76,325	1	0,000
$\widehat{\alpha}_2$	6,985	1,085	41,424	1	0,000
$\widehat{\beta}$	$1,23 \cdot 10^{-3}$	$0,438 \cdot 10^{-3}$	7,960	1	0,005

Τα μοντέλα είναι:

$$\log\left[\frac{P(Y \leq 0)}{1-P(Y \leq 0)}\right] = 1,796 + 1,23 \cdot 10^{-3} \cdot x$$

$$\log\left[\frac{P(Y \leq 1)}{1-P(Y \leq 1)}\right] = 5,178 + 1,23 \cdot 10^{-3} \cdot x$$

$$\log\left[\frac{P(Y \leq 2)}{1-P(Y \leq 2)}\right] = 6,985 + 1,23 \cdot 10^{-3} \cdot x$$

Από τα μοντέλα αυτά προκύπτουν οι πιθανότητες $P(Y=0/X)$, $P(Y=1/X)$, $P(Y=2/X)$ και $P(Y=3/X)$. Στον πίνακα αυτόν υπάρχουν τα αποτελέσματα του ελέγχου:

$$H_0 : \beta = 0$$

$$H_1 : \beta \neq 0$$

Η φυσική σημασία του οποίου είναι, εάν η ανεξάρτητη μεταβλητή είναι στατιστικά σημαντική. Η τιμή της ελεγχοσυνάρτησης είναι $W=7,960$ που ακολουθεί την χ^2 κατανομή με 1 βαθμό ελευθερίας. Το παρατηρούμενο επίπεδο σημαντικότητας είναι $p-value=0,005$ που είναι μικρότερο από κάθε σύνηθες επίπεδο σημαντικότητας. Αυτό σημαίνει ότι απορρίπτεται η μηδενική υπόθεση, δηλαδή απορρίπτεται η υπόθεση ότι το υψόμετρο δεν είναι σημαντικός παράγοντας για την εμφάνιση και ανάπτυξη του φυτού.

Για τους ελέγχους:

$$H_0 : \alpha_j = 0$$

$$H_1 : \alpha_j \neq 0 \text{ με } j=0,1,2,3$$

Οι αντίστοιχες Wald τιμές είναι $W=17,834$, $W=76,325$, $W=41,424$ όλες ακολουθούν την χ^2 κατανομή με έναν B.E. Οι p-value τιμές είναι σχεδόν μηδέν για όλες της τιμές του j. Άρα οι τιμές των α_j για κάθε $j=0,1,2,3$, δεν μπορεί να θεωρηθεί ότι είναι μηδενικές.

Αξιολόγηση του μοντέλου.

Τα Pseudo R-square για το μοντέλο είναι:

Pseudo R-square	Τιμές
Cox and Snell	0.024
Negelkerke	0.032
McFadden	0.017

Οι τιμές των μέτρων αυτών είναι ιδιαίτερα μικρές.

Αξιολόγηση του μοντέλου με βάση την προσαρμογή στα δεδομένα.

Οι υποθέσεις που ελέγχονται είναι:

H_0 : Το μοντέλο προσαρμόζεται ικανοποιητικά στα δεδομένα.

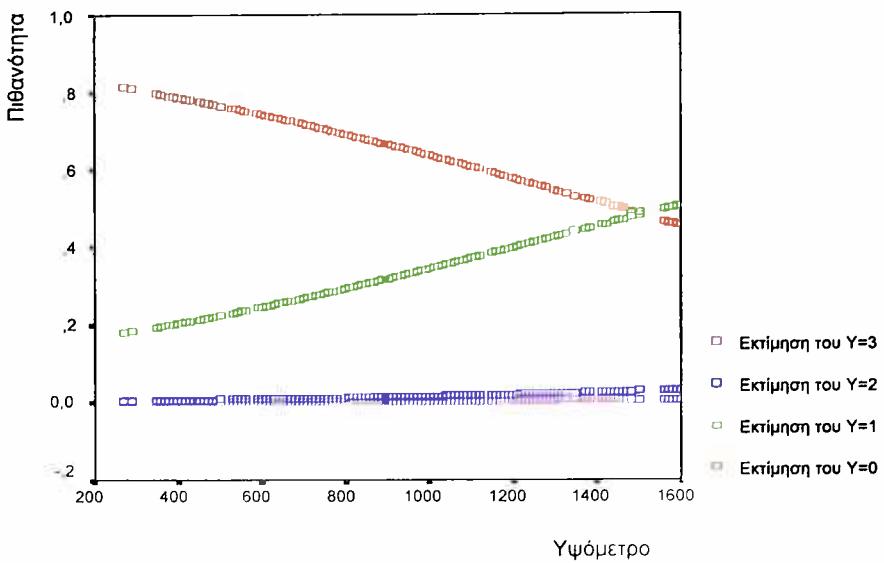
H_1 : Αντίθετα.

Στον επόμενο πίνακα υπάρχουν τα αποτελέσματα του ελέγχου κατά Deviance.

Goodness-of-fit	Τιμή χ^2	Βαθμοί Ελευθερίας	Επίπεδο Σημαντικότητας
Deviance	187,290	293	1

Η τιμή της συνάρτησης ελέγχου είναι $G=187,290$ με 293 βαθμούς ελευθερίας. Το παρατηρούμενο επίπεδο σημαντικότητας είναι σχεδόν 1, μεγαλύτερο από κάθε σύνηθες επίπεδο σημαντικότητας. Άρα δεν υπάρχουν αρκετές ενδείξεις απόρριψης της μηδενικής υπόθεσης. Ετσι δεν μπορεί να μην γίνει δεκτό ότι το μοντέλο έχει καλή προσαρμογή στα δεδομένα. Ετσι το μοντέλο μπορεί να χρησιμοποιηθεί για προβλέψεις.

Οι προβλέψεις που δίνει το μοντέλο σε σχέση με τις πιθανότητες πραγματοποίησης τους εμφανίζονται στο παρακάτω διάγραμμα.



Εδώ παρατηρείται ότι το μοντέλο έχει πολύ μεγάλες πιθανότητες να απουσιάζει στα χαμηλά υψόμετρα του βουνού. Καθώς αυξάνεται το υψόμετρο όμως η πιθανότητα της απουσίας μειώνεται και η πιθανότητα της παρουσίας αυξάνεται με τρόπο αντιστρόφως ανάλογο. Από το υψόμετρο των 1400 μέτρων περίπου και πάνω το μοντέλο προβλέπει ότι θα υπάρχει το φυτό δημιουργώντας πληθοκαλύψεις έως 5%. Εξετάζοντας τις πιθανότητες εμφάνισης του φυτού με μεγαλύτερη κάλυψη, παρατηρείται ότι αυτό έχει πολύ μικρή πιθανότητα να συμβεί γενικά. Το συγκεκριμένο φυτό θα μπορούσε να χαρακτηριστεί ως υψόφιλο.

Έλεγχος υποθέσεων.

Το παραπάνω μοντέλο εκτιμήθηκε με την προϋπόθεση ότι η επίδραση της ανεξάρτητης μεταβλητής είναι ίδια σε κάθε επίπεδο παρατήρησης της εξαρτημένης μεταβλητής. Τα μοντέλα αυτά προέκυψαν με την προϋπόθεση ότι το β είναι σταθερό σε όλα τα επίπεδα της μεταβλητής Y . Αυτή η υπόθεση όμως πρέπει να ελεγχθεί, έτσι οι υποθέσεις του ελέγχου είναι:

H_0 : Όλες οι κατηγορίες έχουν το ίδιο σετ συντελεστών.

H_1 : Αντίθετα.

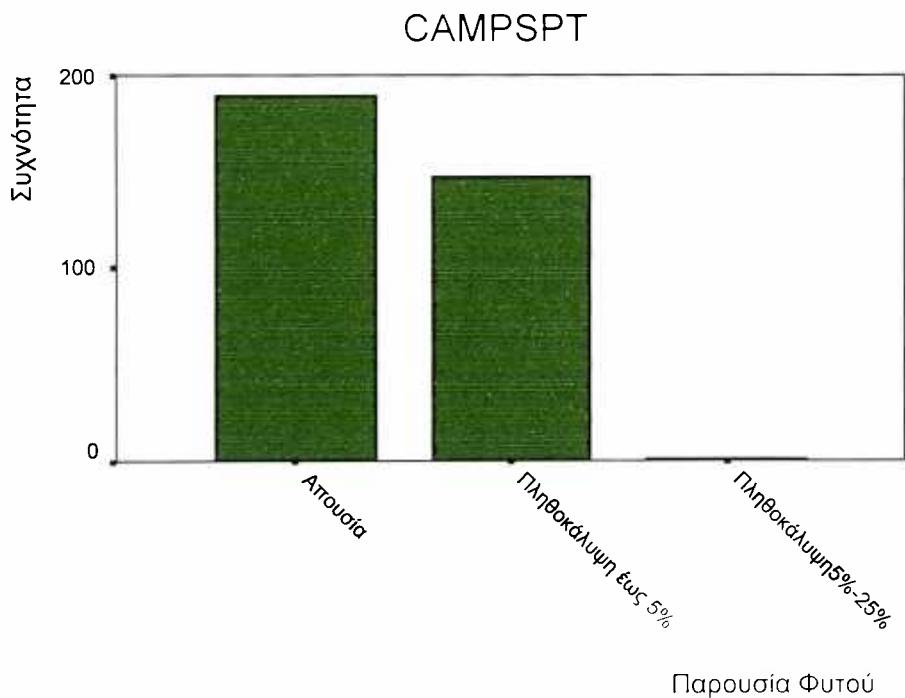
Η ελεγχοσυνάρτηση του ελέγχου αυτού όπως έχει αναφερθεί υπολογίζεται από την διαφορά των λογαρίθμων των πιθανοφανειών των μοντέλων. Η κατανομή της ελεγχοσυνάρτησης αυτής είναι η χ^2 με δύο βαθμούς ελευθερίας. Τα αποτελέσματα του ελέγχου αυτού δίνονται στον επόμενο πίνακα:

Μοντέλο	-2log-likelihood	Τιμή χ^2	Βαθμοί Ελευθερίας	Επίπεδο Σημαντικότητας
Null Hypothesis	287,810			
General	277,807	1,003	2	0,606

Η τιμή της ελεγχοσυνάρτησης είναι $G=1,003$. Το παρατηρούμενο επίπεδο σημαντικότητας είναι $p\text{-value}=0,606$ μεγαλύτερο από κάθε σύνηθες επίπεδο σημαντικότητας. Αυτό σημαίνει ότι δεν υπάρχουν αρκετές ενδείξεις ώστε να απορρίπτεται η μηδενική υπόθεση. Έτσι δεν μπορεί να μην γίνει δεκτό ότι οι σιγμοειδείς καμπύλες των αθροιστικών πιθανοτήτων είναι παράλληλες. Δηλαδή, ότι δεν υπάρχει ο ίδιος συντελεστής της ανεξάρτητη μεταβλητή για κάθε επίπεδο της εξαρτημένης.

4.5. *Campanula statulata* (Campspt) Ποώδες φυτό.

Το διάγραμμα συχνοτήτων της πληθοκάλυψης του εδάφους από το συγκεκριμένο φυτό είναι:



Γράφημα 8: Ραβδόγραμμα συχνοτήτων της πληθοκάλυψης του εδάφους από το φυτό *Campanula spatulata*.

Η κατανομή συχνοτήτων της πληθοκάλυψης ακριβέστερα δίνεται στον επόμενο πίνακα:

Εμφάνιση φυτού	Συχνότητα	Σχετική συχνότητα (%)	Αθροιστική Σχετική Συχνότητα (%)
Απουσία του φυτού από το δείγμα	189	56,1	56,1
Παρουσία του φυτού στην δειγματοληψία και πληθοκάλυψη του εδάφους έως 5%	147	43,6	99,7
Παρουσία του φυτού στην δειγματοληψία και πληθοκάλυψη του εδάφους από 5% έως 25%	1	0,3	100
Σύνολο	337	100,0	

Πίνακας 14: Κατανομή της πληθοκάλυψης του εδάφους από το φυτό *Campanula spatulata*

Στον πίνακα αυτόν παρατηρείται ότι στο 56,1% των δειγματοληψιών δεν υπήρχε το συγκεκριμένο φυτό. Από το υπόλοιπο 43,9% που υπήρχε το φυτό, το 43,6% δημιουργούσε μικρές πληθοκαλύψεις έως 5%. Ήα αναζητηθεί και εδώ μοντέλο περιγραφής και πρόβλεψης, μεταξύ της πληθοκάλυψης του εδάφους και των περιβαλλοντικών μεταβλητών, που εχουν καταγραφεί. Η καταλληλότερη συνάρτηση σύνδεσης από όσες χρησιμοποιήθηκαν είναι η **Cauchit**, γιατί ικανοποιεί της προϋποθέσεις που απαιτούνται, και δίνει τα υψηλότερα Pseudo R-Square. Η μορφή της είναι:

$$\tan[\pi \cdot (P(Y \leq j) - 0,5)] = \hat{\alpha} + \hat{\beta} \cdot x \text{ με } j=0,1$$

Η μόνη στατιστικά σημαντική μεταβλητή που προέκυψε κατά την διαδικασία των δοκίμων, είναι το υψόμετρο. Για τον λόγο αυτό, λαμβάνεται ως η μοναδική ανεξάρτητη μεταβλητή του μοντέλου.

Οι εκτιμήσεις των συντελεστών του μοντέλου αυτού υπολογίζονται με την μέθοδο μέγιστης πιθανοφάνειας και δίνονται στον επόμενο πίνακα::

	Εκτίμηση	Τυπική απόκλιση	Τιμή (Wald)	Βαθμοί Ελευθερίας	Επίπεδο Σημαντικότητας
$\hat{\alpha}_0$	2,005	0,458	19,208	1	0,000
$\hat{\alpha}_1$	108,447	106,214	1,042	1	0,307
$\hat{\beta}$	$19,56 \cdot 10^{-4}$	$0,47 \cdot 10^{-3}$	17,334	1	0,000

Τα μοντέλα για κάθε κατηγορία έκβασης είναι:

$$\tan[\pi \cdot (P(Y \leq 0) - 0,5)] = 2,005 + 19,56 \cdot 10^{-4} \cdot x$$

$$\tan[\pi \cdot (P(Y \leq 1) - 0,5)] = 108,447 + 19,56 \cdot 10^{-4} \cdot x$$

Από τους τύπους αυτούς υπολογίζονται οι πιθανότητες κάθε κατηγορίας έκβασης.

Στον πίνακα υπάρχουν οι τιμές για τον έλεγχο:

$$H_0 : \beta = 0$$

$$H_1 : \beta \neq 0$$

Η τιμή της ελεγχοσυνάρτησης είναι $W=17,334$ και ακολουθεί την χ^2 με έναν βαθμό ελευθερίας. Το παρατηρούμενο επίπεδο σημαντικότητας είναι σχεδόν μηδέν, άρα απορρίπτεται η μηδενική υπόθεση. Η ερμηνεία αυτού του ελέγχου είναι ότι δεν μπορούμε να μην δεχτούμε ότι η μεταβλητή υψόμετρο είναι σημαντική για την παρουσία και ανάπτυξη του συγκεκριμένου φυτού.

Για τους ελέγχους:

$$H_0 : a_j = 0$$

$$H_1 : a_j \neq 0 \text{ με } j=0,1,$$

Η τιμή της ελεγχοσυνάρτησης για $j=0$, είναι $W=19,208$ και ακολουθεί την με έναν βαθμό ελευθερίας. Το παρατηρούμενο επίπεδο σημαντικότητας είναι σχεδόν μηδέν, άρα απορρίπτεται η μηδενική υπόθεση. Για $j=1$ τιμή της ελεγχοσυνάρτησης είναι $W=1,042$ και ακολουθεί την χ^2 με έναν βαθμό ελευθερίας. Το παρατηρούμενο επίπεδο σημαντικότητας είναι $p\text{-value}=0,307$ μεγαλύτερο από κάθε σύνηθες επίπεδο σημαντικότητας (1% έως 5%). Έτσι δεν απορρίπτεται η μηδενική υπόθεση. Δηλαδή δεν απορρίπτεται η υπόθεση ότι η τιμή του a_1 είναι μηδενική.

Αξιολόγηση του μοντέλου.

Τα Pseudo R-square για το μοντέλο είναι:

Pseudo R-square	Τιμές
Cox and Snell	0.080
Negelkerke	0.106
McFadden	0.059

Οι τιμές των μέτρων αυτών είναι ιδιαίτερα μικρές. Η πραγματοποιηθεί αξιολόγηση του μοντέλου με βάση την προσαρμογή στα δεδομένα. Η αξιολόγηση θα γίνει με έλεγχο οι υποθέσεις του οποίου είναι:

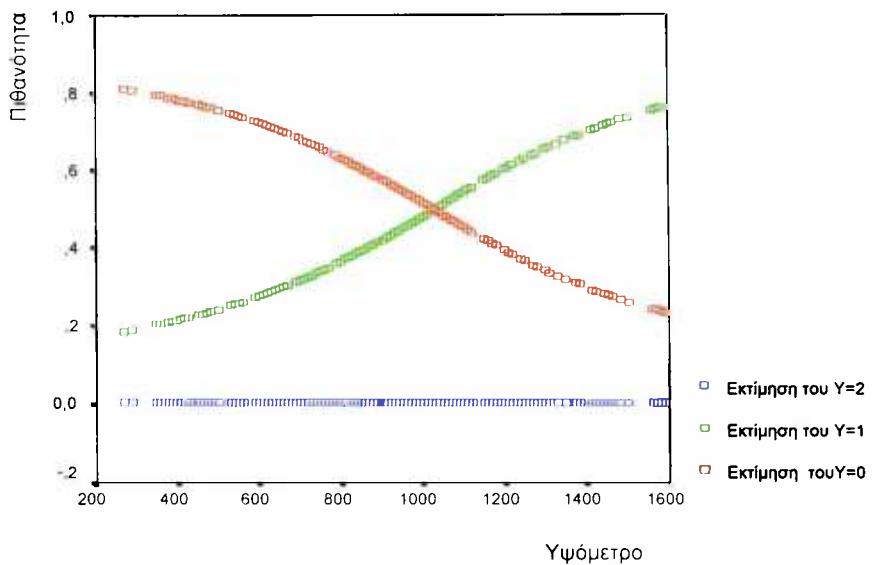
H_0 : Το μοντέλο προσαρμόζεται ικανοποιητικά στα δεδομένα.

H_1 : Αντίθετα.

Τα αποτελέσματα του ελέγχου δίνονται στον επόμενο πίνακα.

Goodness-of-fit	Τιμή χ^2	Βαθμοί Ελευθερίας	Επίπεδο Σημαντικότητας
Deviance	143,202	159	0,998

Η τιμή της συνάρτησης ελέγχου κατά Deviance είναι $G=143,202$ ακολουθεί χ^2 κατανομή με 159 βαθμούς ελευθερίας. Το παρατηρούμενο επίπεδο σημαντικότητας είναι $p\text{-value}=0,998$, μεγαλύτερο από κάθε σύνηθες επίπεδο σημαντικότητας. Άρα δεν υπάρχουν αρκετές ενδείξεις απόρριψης της μηδενικής υπόθεσης. Έτσι δεν μπορεί να μην γίνει δεκτό ότι το μοντέλο έχει καλή προσαρμογή στα δεδομένα. Το μοντέλο αυτό μπορεί να χρησιμοποιηθεί για προβλέψεις. Οι προβλέψεις που δίνει το μοντέλο σε σχέση με τις πιθανότητες πραγματοποίησης τους εμφανίζονται στο παρακάτω σχήμα.



Παρατηρείται ότι το μοντέλο μέχρι τα 1000 μέτρα περίπου προβλέπει απουσία του φυτού, ενώ πάνω από εκεί προβλέπει παρουσία. Η πιθανότητα απουσίας είναι πολύ μεγάλη αρχικά και σταδιακά μειώνεται με την αύξηση του υψομέτρου. Η παρουσία του φυτού δημιουργώντας πληθοκαλύψεις έως 5%, αρχικά έχει πολύ μικρή πιθανότητα να συμβεί, η οποία όμως αυξάνεται με την αύξηση του υψομέτρου. Στα πολύ ψηλά του βουνού προβλέπεται ότι θα υπάρχει το φυτό, με αυτή την κάλυψη, με πολύ μεγάλη πιθανότητα η οποία πλησιάζει το 0,8. Ακόμη είναι σχεδόν απίθανο, να δημιουργεί το συγκεκριμένο φυτό, πληθοκαλύψεις μεγαλύτερες του 5% στο όρος Ερύμανθο.

Έλεγχος υποθέσεων.

Το παραπάνω μοντέλο εκτιμήθηκε με την προϋπόθεση ότι η επίδραση της ανεξάρτητης μεταβλητής είναι ίδια σε κάθε επίπεδο παρατήρησης της εξαρτημένης

μεταβλητής. Τα μοντέλα αυτά προέκυψαν με την προϋπόθεση ότι το β είναι σταθερό σε όλα τα επίπεδα της μεταβλητής Y. Αυτή η υπόθεση όμως πρέπει να ελεγχθεί, έτσι οι υποθέσεις του ελέγχου είναι:

H_0 : Όλες οι κατηγορίες έχουν το ίδιο σετ συντελεστών.

H_1 : Αντίθετα.

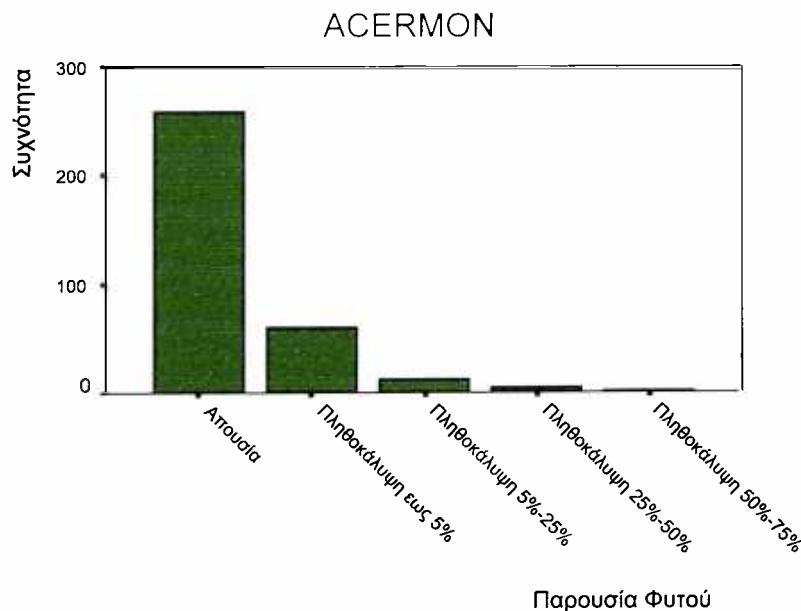
Τα αποτελέσματα αυτού του ελέγχου δίνονται στον πίνακα:

Μοντέλο	-2log-likelihood	Τιμή χ^2	Βαθμοί Ελευθερίας	Επίπεδο Σημαντικότητας
Null Hypothesis	237,713			
General	237,606	1,107	1	0,743

Η κατανομή της ελεγχοσυνάρτησης αυτής είναι η χ^2 με έναν βαθμό ελευθερίας. Η τιμή της ελεγχοσυνάρτησης είναι $G=1,107$. Το παρατηρούμενο επίπεδο σημαντικότητας είναι $p\text{-value}=0,743$ μεγαλύτερο από κάθε σύνηθες επίπεδο σημαντικότητας. Αυτό σημαίνει ότι δεν υπάρχουν αρκετές ενδείξεις ώστε να απορρίπτεται η μηδενική υπόθεση. Έτσι δεν μπορεί να μην γίνει δεκτό ότι οι σιγμοειδείς καμπύλες των αθροιστικών πιθανοτήτων είναι παράλληλες. Δηλαδή ότι δεν υπάρχει ο ίδιος συντελεστής της ανεξάρτητης μεταβλητής για κάθε επίπεδο της εξαρτημένης.

4.6. Acer mouspesulanum (Acermon) Δέντρο

Το διάγραμμα συχνοτήτων της πληθοκάλυψης του φυτού Acer mouspesulanum (το γνωστό δέντρο σφενδάμι) είναι:



Παρουσία Φυτού

Γράφημα 9: Ραβδόγραμμα συχνοτήτων της πληθοκάλυψης των εδάφους από το φυτό *mouspesulanum*

Ο επόμενος πίνακας περιγράφει ακριβέστερα την κατανομή της μεταβλητής αυτής.

Εμφάνιση φυτού	Συχνότητα	Σχετική συχνότητα (%)	Αθροιστική Σχετική Συχνότητα (%)
Απουσία του φυτού από το δείγμα	258	76,6	76,6
Παρουσία του φυτού στην δειγματοληψία και πληθοκάλυψη του εδάφους έως 5%	61	18,1	94,7
Παρουσία του φυτού στην δειγματοληψία και πληθοκάλυψη του εδάφους από 5% έως 25%	13	3,9	98,5
Παρουσία του φυτού στην δειγματοληψία και πληθοκάλυψη του εδάφους από 25% έως 50%	4	1,2	99,7
Παρουσία του φυτού στην δειγματοληψία και πληθοκάλυψη του εδάφους από 50% έως 75%	1	0,3	100
Σύνολο	337	100,0	

Πίνακας 15: Κατανομή της πληθοκάλυψης του εδάφους από το φυτό Acer mouspesulanum

Από τον πίνακα παρατηρείται ότι από το 76,6% των δειγματοληψιών απουσίαζε το συγκεκριμένο φυτό. Στο υπόλοιπο 23,4% που παρατηρήθηκε το φυτό, το 18,1% δημιουργούσε πληθοκαλύψεις του εδάφους έως 5%, και το 3,9% δημιουργούσε πληθοκαλύψεις από 5% έως 25%. Η αναζήτηση τρόπου σύνδεσης μεταξύ της μεταβλητής πληθοκάλυψη του εδάφους και των περιβαλλοντικών παραγόντων (υψόμετρο, κλίση, πέτρωμα, έκθεση) οδήγησε στο συμπέρασμα ότι η καταλληλότερη συνάρτηση σύνδεσης είναι η Probit, με ανεξάρτητες μεταβλητές το υψόμετρο και το υψόμετρο στην δευτέρα. Το μοντέλο αυτό είναι το καλύτερο (μεταξύ εκείνων που δοκιμάστηκαν) τόσο ως προς την κάλυψη των προϋποθέσεων, αλλά και ως προς τις υψηλότερες τιμές των Pseudo R-Square. Η μορφή του είναι:

$$\Phi^{-1}(P(Y \leq j)) = \hat{\alpha}_j + \hat{\beta}_1 \cdot x + \hat{\beta}_2 \cdot x^2 \quad \text{με } j=0,1,2,3$$

Θα ελεγχθεί εάν οι ανεξάρτητες μεταβλητές βελτιώνουν την προβλεπτική ικανότητα του μοντέλου σε σχέση με το μοντέλο που δεν περιέχει ανεξάρτητες μεταβλητές. Οι υποθέσεις που ελέγχονται είναι:

$$H_0 : B=0 \text{ δηλαδή } (\beta_1 = \beta_2 = 0)$$

$$H_1 : B \neq 0 \text{ Διαφορετικά}$$

Ο έλεγχος αυτός γίνεται με την σύγκριση των πιθανοφανειών των δυο μοντέλων, του μοντέλου με τις δυο ανεξάρτητες μεταβλητές (υψόμετρο και υψόμετρο στην δευτέρα) και

του μοντέλου που δεν περιέχει ανεξάρτητες μεταβλητές. Τα αποτελέσματα του ελέγχου αυτού δίνονται στον επόμενο πίνακα.

Model Fitting Information	-2log-likelihood	Τιμή χ^2	Βαθμοί Ελευθερίας	Επίπεδο Σημαντικότητας
Intercept Only	282,843			
Final	246,655	36,188	2	0,000

Η ελεγχοσυνάρτηση έχει τιμή $G=36,188$ και ακολουθεί την χ^2 κατανομή με δυο βαθμούς ελευθερίας. Το παρατηρούμενο επίπεδο σημαντικότητας είναι σχεδόν μηδέν. Αυτό σημαίνει ότι απορρίπτεται η μηδενική υπόθεση, άρα απορρίπτεται η υπόθεση ότι οι ανεξάρτητες μεταβλητές δεν προσφέρουν στην προβλεπτική ικανότητα του μοντέλου. Για τον λόγο αυτό χρησιμοποιούνται και οι δυο μεταβλητές στην κατασκευή του μοντέλου. Οι τιμές των παραμέτρων του παραπάνω μοντέλου εκτιμούνται και δίνονται στον επόμενο πίνακα:

	Εκτίμηση	Τυπική απόκλιση	Τιμή (Wald)	Βαθμοί Ελευθερίας	Επίπεδο Σημαντικότητας
$\widehat{\alpha}_0$	7,006	1,389	25,437	1	0,000
$\widehat{\alpha}_1$	7,965	1,401	32,341	1	0,000
$\widehat{\alpha}_2$	8,549	1,411	36,697	1	0,000
$\widehat{\alpha}_3$	9,138	1,441	40,211	1	0,000
$\widehat{\beta}_1$	$1,27 \cdot 10^{-2}$	$2,825 \cdot 10^{-3}$	20,221	1	0,000
$\widehat{\beta}_2$	$6,06 \cdot 10^{-6}$	$1,408 \cdot 10^{-3}$	18,526	1	0,000

Το μοντέλο για κάθε κατηγορία έκβασης είναι:

$$\Phi^{-1}(P(Y \leq 0)) = 7,006 + 1,27 \cdot 10^{-2} \cdot x - 6,06 \cdot 10^{-6} \cdot x^2$$

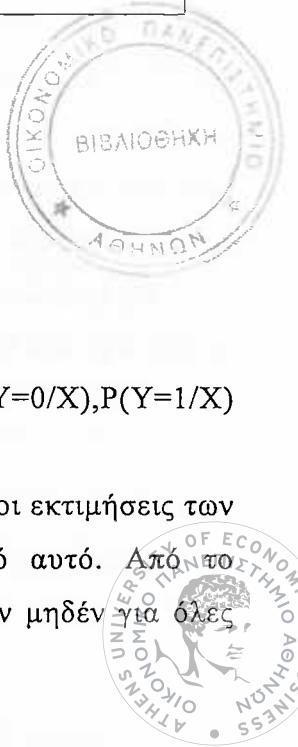
$$\Phi^{-1}(P(Y \leq 1)) = 7,956 + 1,27 \cdot 10^{-2} \cdot x - 6,06 \cdot 10^{-6} \cdot x^2$$

$$\Phi^{-1}(P(Y \leq 2)) = 8,549 + 1,27 \cdot 10^{-2} \cdot x - 6,06 \cdot 10^{-6} \cdot x^2$$

$$\Phi^{-1}(P(Y \leq 3)) = 9,138 + 1,27 \cdot 10^{-2} \cdot x - 6,06 \cdot 10^{-6} \cdot x^2$$

Από τις παραπάνω μορφές υπολογίζονται οι πιθανότητες $P(Y=0/X), P(Y=1/X)$, $P(Y=2/X), P(Y=3/X)$

Στον πίνακα ακόμη υπάρχουν οι τιμές του Wald ελέγχου για το εάν οι εκτιμήσεις των παραμέτρων του μοντέλου είναι μηδέν ή διαφέρουν σημαντικά από αυτό. Από το παρατηρούμενο επίπεδο σημαντικότητας, κάθε ελέγχου που είναι σχεδόν μηδέν για όλες



τις παραμέτρους σημαίνει ότι απορρίπτεται η υπόθεση ότι η συγκεκριμένη παράμετρος είναι μηδέν. Το σημαντικότερο όμως είναι ότι κατά τον έλεγχο για το εάν οι μεταβλητές υψόμετρο και υψόμετρο στην δευτέρα είναι σημαντικές ως επεξηγηματικές μεταβλητές του μοντέλου. Προκύπτει ότι δεν μπορεί να μην θεωρηθούν στατιστικά σημαντικές για τον συγκεκριμένο τρόπο μοντελοποίησης.

Αξιολόγηση του μοντέλου.

Τα Pseudo R-square για το μοντέλο είναι:

Pseudo R-square	Τιμές
Cox and Snell	0,102
Negelkerke	0,134
McFadden	0,076

Οι τιμές των μέτρών αυτών είναι σχετικά μικρές.

Θα πραγματοποιηθεί αξιολόγηση του μοντέλου με βάση την προσαρμογή στα δεδομένα. Η αξιολόγηση θα γίνει και με έλεγχο οι υποθέσεις του οποίου είναι:

H_0 : Το μοντέλο προσαρμόζεται ικανοποιητικά στα δεδομένα.

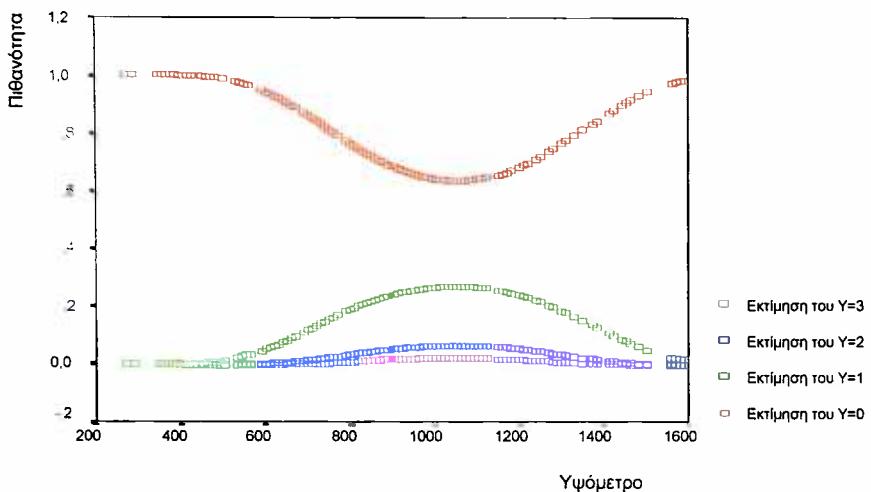
H_1 : Αντίθετα.

Στον επόμενο πίνακα υπάρχουν τα αποτελέσματα του ελέγχου αυτού.

Goodness-of-fit	Τιμή χ^2	Βαθμοί Ελευθερίας	Επίπεδο Σημαντικότητας
Deviance	153	390	1

Η τιμή της συνάρτησης ελέγχου κατά Deviance είναι $G=153$ που ακολουθεί την χ^2 με 390 βαθμούς ελευθερίας. Το παρατηρούμενο επίπεδο σημαντικότητας είναι περίπου 1, μεγαλύτερο από κάθε σύνηθες επίπεδο σημαντικότητας. Άρα δεν υπάρχουν αρκετές ενδείξεις απόρριψης της μηδενικής υπόθεσης. Έτσι δεν μπορεί να μην γίνει δεκτό ότι το μοντέλο έχει καλή προσαρμογή στα δεδομένα.

Το μοντέλο αυτό είναι δυνατό να χρησιμοποιηθεί για προβλέψεις, οι πιθανότητες πρόβλεψης των διαφόρων κατηγοριών έκβασης παριστάνονται στο επόμενο διάγραμμα.



Είναι εμφανές ότι το μοντέλο προβλέπει πάντα ότι τα φυτό θα απουσιάζει από το όρος Ερύμανθο. Στα υψόμετρα έως 600 μέτρα και πάνω από 1400, δίνει την πρόβλεψη αυτή ως βέβαιο ενδεχόμενο. Στα ενδιάμεσα υψόμετρα από 600 έως 1100 μέτρα η πιθανότητα απουσίας του φυτού μειώνεται σταδιακά και γύρω στα 1000 μέτρα η πιθανότητα αυτή είναι περίπου 0,6. Από τα 1000 μέτρα και πάνω η πιθανότητα απουσίας του φυτού αυξάνεται και πάλι, μέχρι που στα μεγάλα υψόμετρα να θεωρείται απίθανο να υπάρχει το φυτό. Έτσι το φυτό Acer mouspesulanum θα μπορούσε να χαρακτηριστεί ως σπάνιο φυτό της χλωρίδας του Ερυμάνθου. Όμως η πιθανότητα εμφάνισης του είναι μεγαλύτερη σε υψόμετρα από 600 εως 1400, με αισθητή κορύφωση στα 1000 μέτρα.

Έλεγχος υποθέσεων.

Το συγκεκριμένο μοντέλο προέκυψε με την προϋπόθεση ότι το διάνυσμα B είναι σταθερό σε όλα τα επίπεδα της μεταβλητής Y. Η υπόθεση αυτή όμως πρέπει να ελεγχθεί. Αυτές διατυπώνονται:

H₀ : Όλες οι κατηγορίες έχουν το ίδιο σετ συντελεστών ανεξάρτητων μεταβλητών.

H₁ : Αντίθετα.

Τα αποτελέσματα του ελέγχου αυτού, δίνονται στον πίνακα:

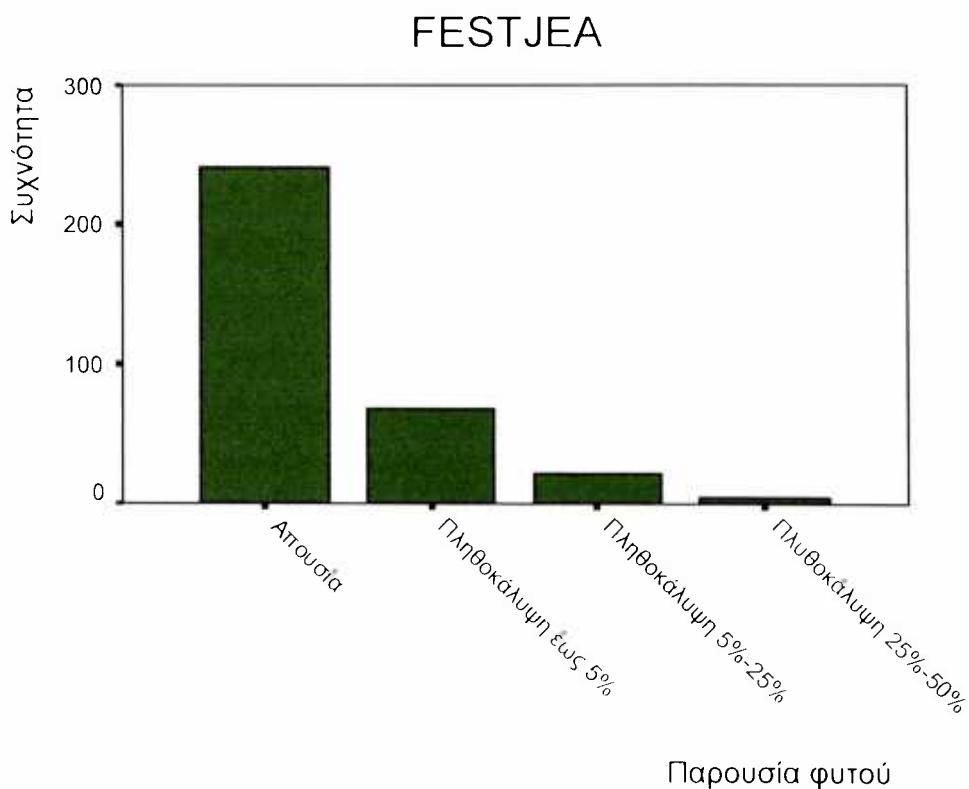
Μοντέλο	-2log-likelihood	Τιμή X^2	Βαθμοί Ελευθερίας	Επίπεδο Σημαντικότητας
Null Hypothesis	246.655			
General	264,340	-	6	-

Η ελεγχοσυνάρτηση του ελέγχου αυτού όπως έχει αναφερθεί υπολογίζεται από την διαφορά των λογαρίθμων των πιθανοφανειών μεταξύ του μοντέλου που δημιουργήθηκε

και ενός θεωρητικού μοντέλου που έχει σταθερό το διάνυσμα Β για κάθε κατηγορία έκβασης της εξαρτημένης μεταβλητής. Η τιμή της ελεγχοσυνάρτησης είναι θετική από κατασκευής, για τον λόγο αυτό δεν έχει νόημα τιμής ελεγχοσυνάρτησης η διαφορά των πιθανοφανειών των παραπάνω μοντέλων. Έτσι δεν μπορεί να γίνει ο έλεγχος της παραλληλίας των γραμμών για το συγκεκριμένο μοντέλο και για κάθε άλλο εναλλακτικό που επιχειρήθηκε να βρεθεί.

4.7. *Festuca jeenpertii* (Festjea) Ποώδες αγρωστώδες φυτό.

Το διάγραμμα συχνοτήτων της πληθοκάλυψης του εδάφους από το φυτό *Festuca jeenpertii* είναι:



Γράφημα 10: Ραβδόγραμμα συχνοτήτων της πληθοκάλυψης του εδάφους από το φυτό *Festuca jeenpertii*

Ο επόμενος πίνακας δίνει μια πιο ακριβή εικόνα της κατανομή της μεταβλητής πληθοκάλυψη του εδάφους από το συγκεκριμένο φυτό.

Εμφάνιση φυτού	Συχνότητα	Σχετική συχνότητα (%)	Αθροιστική Σχετική Συχνότητα (%)
Απουσία του φυτού από το δείγμα	242	71,8	71,8
Παρουσία του φυτού στην δειγματοληψία και πληθοκάλυψη του εδάφους έως 5%	68	20,2	92,0
Παρουσία του φυτού στην δειγματοληψία και πληθοκάλυψη του εδάφους από 5% έως 25%	25	6,5	98,5
Παρουσία του φυτού στην δειγματοληψία και πληθοκάλυψη του εδάφους από 25% έως 50%	5	1,5	100
Σύνολο	337	100,0	

Πίνακας 16: Κατανομή της πληθοκάλυψης του εδάφους από το φυτό Festuca jeenpertii

Από τον πίνακα συχνοτήτων φαίνεται ότι στο 71,8% των δειγματοληψιών δεν παρατηρήθηκε το συγκεκριμένο είδος. Για το υπόλοιπο 28,2% που υπήρχε το φυτό, στο 20,2% βρέθηκε να δημιουργεί πληθοκαλύψεις μέχρι 5%.

Μοντέλο 1

Θα αναζητηθεί συνάρτηση σύνδεσης της εξαρτημένης μεταβλητής, πληθοκάλυψη του εδάφους και ανεξάρτητης, το υψόμετρο του εδάφους Η καταλληλότερη συνάρτηση σύνδεσης είναι η Complementary log-log

$$\log[-\log(1 - P(Y \leq j))] = \hat{\alpha}_j + \hat{\beta} \cdot x \text{ με } j=0,1.$$

Ο πίνακας με τις εκτιμήσεις των παραμέτρων του μοντέλου είναι:

Εκτίμηση	Τυπική απόκλιση	Τιμή (Wald)	Βαθμοί Ελευθερίας	Επίπεδο Σημαντικότητας
$\hat{\alpha}_0$	1,288	0,221	33,894	1 0,000
$\hat{\alpha}_1$	2,017	0,235	73,844	1 0,000
$\hat{\alpha}_2$	2,540	0,260	95,496	1 0,000
$\hat{\beta}$	$1,132 \cdot 10^{-3}$	$0,235 \cdot 10^{-3}$	23,250	1 0,000

Η μορφή του μοντέλου για κάθε επίπεδο της κατηγορίας έκβασης της εξαρτημένης μεταβλητής είναι:

$$\log[-\log(1 - P(Y \leq 0))] = 1,288 + 1,132 \cdot 10^{-3} \cdot x$$

$$\log[-\log(1 - P(Y \leq 1))] = 2,017 + 1,132 \cdot 10^{-3} \cdot x$$

$$\log[-\log(1 - P(Y \leq 2))] = 2,540 + 1,132 \cdot 10^{-3} \cdot x$$

Στον πίνακα αυτό υπάρχουν και τα αποτελέσματα του Wald test που ελέγχουν εάν οι παράμετροι του μοντέλου διαφέρουν στατιστικά σημαντικά από το μηδέν. Για όλους τους ελέγχους αυτούς τα παρατηρούμενα επίπεδα σημαντικότητας είναι σχεδόν μηδέν, αυτό σημαίνει ότι όλοι οι πραγματικές τιμές των συντελεστών του μοντέλου ($\alpha_1, \alpha_2, \alpha_3, \beta$) διαφέρουν σημαντικά από το μηδέν. Το συμπέρασμα ότι ο συντελεστής της ανεξάρτητης μεταβλητής διαφέρει σημαντικά από το μηδέν σημαίνει ότι η μεταβλητή υψόμετρο είναι στατιστικά σημαντική για το μοντέλο.

Αξιολόγηση του μοντέλου.

Τα Pseudo R-square για το μοντέλο είναι:

Pseudo R-square	Τιμές
Cox and Snell	0,056
Negelkerke	0,071
McFadden	0,036

Οι τιμές των μέτρων αυτών είναι πολύ μικρές.

Η αξιολόγηση του μοντέλου θα γίνει και με βάση την προσαρμογή του στα δεδομένα. Η αξιολόγηση αυτή θα γίνει με έλεγχο, οι υποθέσεις του οποίου είναι:

H_0 : Το μοντέλο προσαρμόζεται ικανοποιητικά στα δεδομένα.

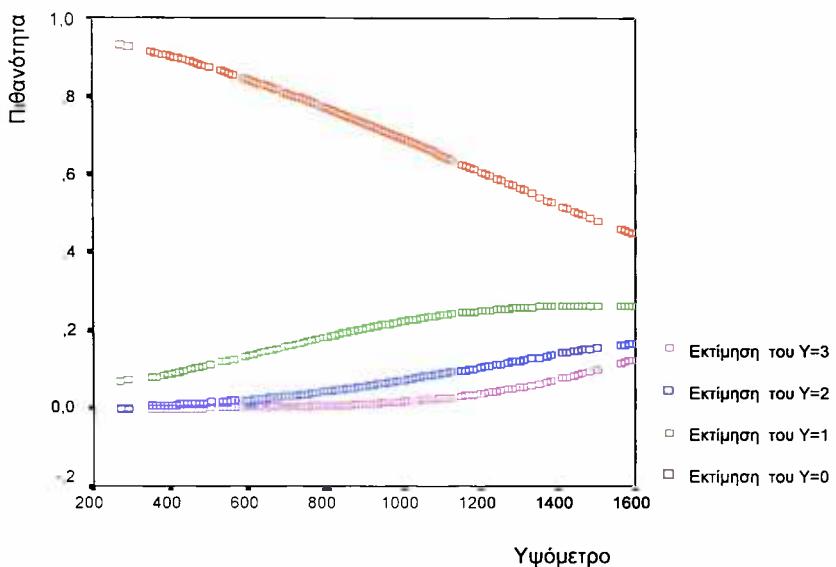
H_1 : Αντίθετα.

Στον επόμενο πίνακα περιέχονται τα αποτελέσματα του παραπάνω ελέγχου.

Goodness-of-fit	Τιμή χ^2	Βαθμοί Ελευθερίας	Επίπεδο Σημαντικότητας
Deviance	175,290	293	1

Η τιμή της ελεγχοσυνάρτησης Deviance είναι $G=175,29$ και ακολουθεί την χ^2 κατανομή με 293 βαθμούς ελευθερίας. Το παρατηρούμενο επίπεδο(p-value) σημαντικότητας είναι 1, μεγαλύτερο από κάθε σύνηθες επίπεδο σημαντικότητας. Άρα δεν υπάρχουν

αρκετές ενδείξεις απόρριψης της μηδενικής υπόθεσης. Αυτό ερμηνεύεται ως, ότι το μοντέλο δεν μπορεί να θεωρηθεί ότι δεν περιγράφει σωστά τα δεδομένα. Έτσι θα χρησιμοποιηθεί για προβλέψεις των κατηγοριών έκβασης της εξαρτημένης μεταβλητής, καθώς και των πιθανοτήτων με τις οποίες δίνει τις προβλέψεις αυτές. Οι πιθανότητες των προβλέψεων αυτών παριστάνονται στο επόμενο διάγραμμα συναρτήσει του υψομέτρου.



Παρατηρείται ότι το μοντέλο προβλέπει πάντα την απουσία του φυτού. Στα χαμηλά υψόμετρα η πιθανότητα της απουσίας είναι αρκετά μεγάλη και μειώνεται σταδιακά έως που σε υψόμετρα, κοντά στα 1600μ γίνεται περίπου 0,5. Αυτό σημαίνει ότι και η πιθανότητα του να υπάρχει το φυτό είναι κοντά στο 0,5. Έτσι το φυτό έχει μεγαλύτερη πιθανότητα να υπάρχει στα υψηλά σημεία του όρους Ερύμανθου. Η πιθανότητα του να υπάρχει το φυτό και να δημιουργεί πληθοκαλύψεις μέχρι 5% είναι η μεγαλύτερη από την πιθανότητα κάθε άλλης κατηγορίας έκβασης. Ακόμη παρατηρείται ότι αυτή η πιθανότητα αυξάνει καθώς αυξάνεται το υψόμετρο. Λόγο των παραπάνω το φυτό Festuca jeenpertii θα μπορούσε να χαρακτηριστεί ως υψόφιλο.

Έλεγχος υποθέσεων.

Το παραπάνω μοντέλο εκτιμήθηκε με την προϋπόθεση ότι η επίδραση της ανεξάρτητης μεταβλητής είναι ίδια σε κάθε επίπεδο παρατήρησης της εξαρτημένης μεταβλητής. Τα μοντέλα αυτά προέκυψαν με την προϋπόθεση ότι το β είναι σταθερό σε όλα τα επίπεδα της μεταβλητής Y . Αυτή η υπόθεση όμως πρέπει να ελεγχθεί, έτσι οι υποθέσεις του ελέγχου είναι:

H_0 : Όλες οι κατηγορίες έχουν το ίδιο σετ συντελεστών της μεταβλητής του υψομέτρου.

H_1 : Αντίθετα.

Η ελεγχοσυνάρτηση του ελέγχου αυτού, όπως έχει αναφερθεί υπολογίζεται από την διαφορά των λογαρίθμων των πιθανοφανειών των δυο μοντέλων. Του μοντέλου που υπάρχει κάτω από την μηδενική υπόθεση και του μοντέλου που προέκυψε από τα δεδομένα. Τα αποτελέσματα του ελέγχου αυτού δίνονται στον επόμενο πίνακα:

Μοντέλο	-2log-likelihood	Τιμή χ^2	Βαθμοί Ελευθερίας	Επίπεδο Σημαντικότητας
Null Hypothesis	290,396			
General	284,731	5,665	2	0,059

Η τιμή της ελεγχοσυνάρτησης είναι $G=5,665$ και ακολουθεί την χ^2 με 2 βαθμούς ελευθερίας. Το παρατηρούμενο επίπεδο σημαντικότητας είναι $p-value=0,059$ μεγαλύτερο από κάθε σύνηθες επίπεδο σημαντικότητας. Αυτό σημαίνει ότι δεν υπάρχουν αρκετές ενδείξεις ώστε να απορρίπτεται η μηδενική υπόθεση. Έτσι δεν μπορεί να μην γίνει δεκτό ότι οι στιγμοειδείς καμπύλες των αθροιστικών πιθανοτήτων είναι παράλληλες. Δηλαδή ότι δεν υπάρχει ο ίδιος συντελεστής της ανεξάρτητη μεταβλητή (υψόμετρο εδάφους) για κάθε επίπεδο της εξαρτημένης.

Μοντέλο 2

Θα χρησιμοποιηθεί ένα δεύτερο μοντέλο σύνδεσης που έχει την ίδια συνάρτηση Complementary log-log, και ως επεξηγηματικές (ανεξάρτητες) μεταβλητές τις μεταβλητές υψόμετρο και η κλίση. Η μορφή του μοντέλου είναι :

$$\log[-\log(1-P(Y \leq j))] = \hat{a}_j + \hat{B} \cdot x = \hat{a}_j + \hat{\beta}_1 \cdot x_1 + \hat{\beta}_2 \cdot x_2 \text{ με } j=0,1,2 \quad x_1: \text{υψόμετρο} \quad x_2: \text{κλίση}$$

Τα αποτελέσματα της επεξεργασίας των δεδομένων παρατίθενται παρακάτω:

Αρχικά ελέγχονται οι υποθέσεις:

$$H_0 : B=0 \text{ Δηλαδή } (\beta_1 = \beta_2 = 0)$$

$$H_1 : B \neq 0 \text{ Διαφορετικά}$$

Η φυσική σημασία του οποίου είναι εάν, οι δυο ανεξάρτητες μεταβλητές είναι χρήσιμες στο μοντέλο, δηλαδή το μοντέλο που περιέχει και της δυο μεταβλητές έχει καλύτερη προσαρμογή και προβλεπτική ικανότητα από το μοντέλο που δεν περιέχει ανεξάρτητες μεταβλητές.

Έτσι στον επόμενο πίνακα παρατηρούνται τα αποτελέσματα του ελέγχου αυτού.

Model Fitting Information	-2log-likelihood	Τιμή χ^2	Βαθμοί Ελευθερίας	Επίπεδο Σημαντικότητας
Intercept Only	507,781			
Final	476,360	31,421	2	0,000

Η τιμή της ελεγχούσυνάρτησης του ελέγχου αυτού είναι $G=31,421$ και ακολουθεί την χ^2 κατανομή με δύο βαθμούς ελευθερίας. Το παρατηρούμενο επίπεδο σημαντικότητας είναι σχεδόν μηδέν, αυτό σημαίνει ότι απορρίπτεται η μηδενική υπόθεση. Δηλαδή απορρίπτεται η υπόθεση ότι οι μεταβλητές υψόμετρο και κλίση εδάφους είναι στατιστικά ασήμαντες για το μοντέλο. Οι εκτιμήσεις των συντελεστών του μοντέλου βρίσκονται στον επόμενο πίνακα.

	Εκτίμηση	Τυπική Απόκλιση	Τιμή (Wald)	Βαθμοί Ελευθερίας	Επίπεδο Σημαντικότητας
$\widehat{\alpha}_0$	0,922	0,254	13,156	1	0,000
$\widehat{\alpha}_1$	1,674	0,263	40,391	1	0,000
$\widehat{\alpha}_2$	2,262	0,286	62,667	1	0,000
$\widehat{\beta}_1$	$1,588 \cdot 10^{-3}$	$0,271 \cdot 10^{-3}$	34,322	1	0,000
$\widehat{\beta}_2$	$-1,88 \cdot 10^{-2}$	0,006	11,131	1	0,001

Η μορφή των μοντέλων για κάθε κατηγόρια έκβασης είναι :

$$\log[-\log(1-P(Y \leq 0))] = 0,922 + 1,588 \cdot 10^{-3} \cdot x_1 - 1,88 \cdot 10^{-2} \cdot x_2 \quad x_1: \text{υψόμετρο} \quad x_2: \text{κλίση}$$

$$\log[-\log(1-P(Y \leq 1))] = 1,674 + 1,588 \cdot 10^{-3} \cdot x_1 - 1,88 \cdot 10^{-2} \cdot x_2$$

$$\log[-\log(1-P(Y \leq 2))] = 1,674 + 1,588 \cdot 10^{-3} \cdot x_1 - 1,88 \cdot 10^{-2} \cdot x_2$$

Από τους τύπους αυτούς υπολογίζονται οι πιθανότητες κάθε κατηγορίας. Στον προηγούμενο πίνακα περιέχονται οι χαρακτηριστικές τιμές των Wald ελέγχων, για το εάν οι συντελεστές των μοντέλων διαφέρουν σημαντικά από το μηδέν. Από τα παρατηρούμενα επίπεδα των ελέγχων αυτών εξάγεται το συμπέρασμα ότι όλοι οι συντελεστές διαφέρουν σημαντικά από το μηδέν. Άρα και οι ανεξάρτητες μεταβλητές που χρησιμοποιούνται στο μοντέλο είναι στατιστικά σημαντικές.

Αξιολόγηση του μοντέλου.

Τα Pseudo R-square του μοντέλου είναι:

Pseudo R-square	Τιμές
Cox and Snell	0,089
Negelkerke	0,111
McFadden	0,058

Οι τιμές τους είναι μικρές όμως αισθητά μεγαλύτερες από τις αντίστοιχες τιμές του πρώτου μοντέλου (με μοναδική ανεξάρτητη μεταβλητή το υψόμετρο).

Η αξιολόγηση του μοντέλου με βάση την προσαρμογή του στα δεδομένα, γίνεται με έλεγχο, οι υποθέσεις του οποίου είναι:

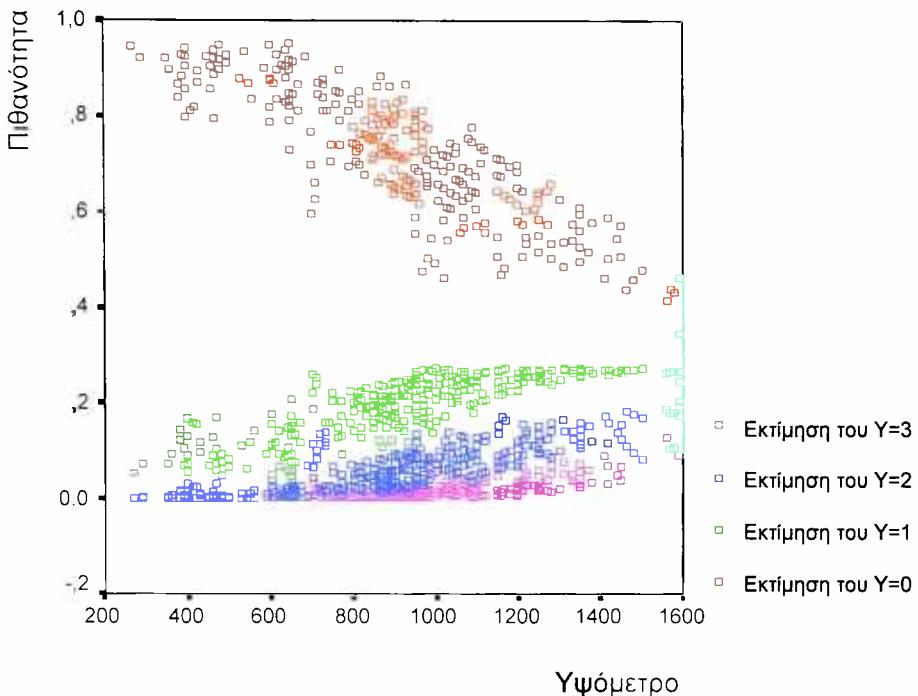
H_0 : Το μοντέλο προσαρμόζεται ικανοποιητικά στα δεδομένα.

H_1 : Αντίθετα.

Στον επόμενο πίνακα περιέχονται τα αποτελέσματα του ελέγχου αυτού.

Goodness-of-fit	Τιμή χ^2	Βαθμοί Ελευθερίας	Επίπεδο Σημαντικότητας
Deviance	447,352	823	1

Ο έλεγχος κατά Deviance που συνήθως χρησιμοποιείται δίνει τιμή της ελεγχοσυνάρτησης $G=447,352$ και ακολουθεί την χ^2 κατανομή με 823 βαθμούς ελευθερίας. Το παρατηρούμενο επίπεδο σημαντικότητας είναι σχεδόν ένα, αυτό σημαίνει ότι δεν απορρίπτεται η μηδενική υπόθεση, έτσι δεν μπορεί να μην γίνει δεκτό ότι το μοντέλο προσαρμόζεται ικανοποιητικά στα δεδομένα. Έτσι αυτό μπορεί να χρησιμοποιηθεί για προβλέψεις. Οι κατηγορίες που προκρίνει η συγκεκριμένη μοντελοποίηση, καθώς και οι πιθανότητες κάθε κατηγορίας έκβασης, παριστάνονται στο επόμενο διάγραμμα..



Παρατηρείται ότι οι προβλέψεις των πιθανοτήτων των κατηγοριών, για κάθε τιμή του υψομέτρου δεν δίνονται μονότιμα και αυτό διότι υπάρχει και δεύτερη ανεξάρτητη μεταβλητή (κλίση του εδάφους). Έτσι οι τιμές των πιθανοτήτων για κάθε κατηγορία έκβασης βρίσκονται σε ζώνες. Είναι ευκρινές όμως ότι η απουσία του φυτού Festuca jeenpertii έχει την μεγαλύτερη πιθανότητα από κάθε άλλη κατηγόρια, αλλά η πιθανότητα αυτή στα χαμηλά υψόμετρα είναι ιδιαίτερα μεγάλη, με την αύξηση του υψομέτρου όμως μειώνεται. Στα υψόμετρα γύρω στα 1600 μέτρα η πιθανότητα του (ενδεχομένου) να μην υπάρχει το φυτό είναι περίπου όση και του (ενδεχομένου) να υπάρχει. Επίσης και η πληθοκάλυψη του φυτού πιθανολογείται ότι θα είναι μεγαλύτερη στα υψηλά υψόμετρα παρά στα χαμηλά. Από τα παραπάνω θα μπορούσε να εξαχθεί το συμπέρασμα ότι το φυτό Festuca jeenpertii παρουσιάζεται φιλικότερο προς τα μεγάλα υψόμετρα, τόσο για την παρουσία του όσο και για την ανάπτυξη του.

Ελεγχος υποθέσεων.

Το παραπάνω μοντέλο εκτιμήθηκε με την προϋπόθεση ότι η επίδραση των ανεξάρτητων μεταβλητών (υψόμετρο και κλίση του εδάφους) είναι ίδια σε κάθε επίπεδο παρατήρησης της εξαρτημένης μεταβλητής. Δηλαδή ότι το διάνυσμα $B=(\beta_1, \beta_2)$ των συντελεστών των ανεξαρτήτων μεταβλητών είναι ίδιο για όλα τα επίπεδα της μεταβλητής Y . Η μηδενική και η εναλλακτική υπόθεση του ελέγχου αυτού είναι:

H_0 : Όλες οι κατηγορίες της εξαρτημένης μεταβλητής έχουν το ίδιο σετ συντελεστών στις ανεξάρτητες μεταβλητές.

H_1 : Αντίθετα.

Τα αποτελέσματα αυτού του ελέγχου δίνονται στον πίνακα:

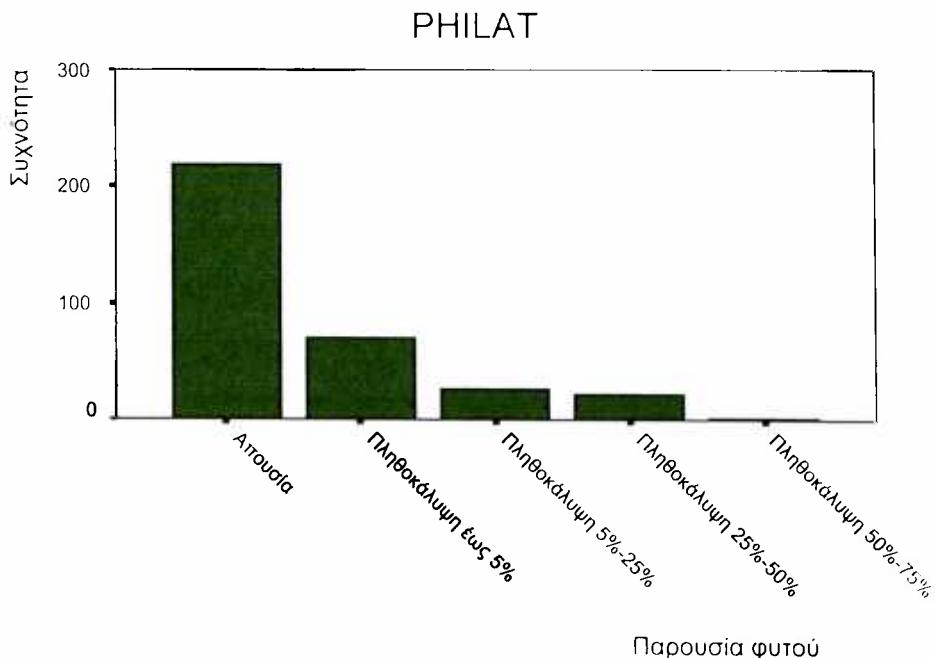
Μοντέλο	-2log-likelihood	Τιμή χ^2	Βαθμοί Ελευθερίας	Επίπεδο Σημαντικότητας
Null Hypothesis	476,360			
General	466,887	9,472	4	0,05

Η τιμή της ελεγχοσυνάρτησης είναι $G=9,472$, κατανομή της είναι η χ^2 με τέσσερις βαθμούς ελευθερίας. Το παρατηρούμενο επίπεδο σημαντικότητας είναι $p\text{-value}=0,05$ μεγαλύτερο από τα συνήθη επίπεδα σημαντικότητας. Αυτό σημαίνει ότι δεν υπάρχουν αρκετές ενδείξεις ώστε να απορρίπτεται η μηδενική υπόθεση. Έτσι δεν μπορεί να μην γίνει δεκτό ότι οι σιγμοειδείς καμπύλες των αθροιστικών πιθανοτήτων είναι παράλληλες. Δηλαδή ότι δεν υπάρχουν οι ίδιοι συντελεστές των ανεξάρτητων μεταβλητών για κάθε επίπεδο της εξαρτημένης.

Παρατήρηση: Το δεύτερο μοντέλο είναι πιο σύνθετο αλλά αυξάνει τις τιμές των Pseudo R-square.

4.8. *Philyrea latifolia* (Philat) Δέντρο.

Το διάγραμμα συχνοτήτων της πληθοκάλυψης του εδάφους από το φυτό (γνωστό και φυλίκι) είναι:



Γράφημα 11: Ραβδόγραμμα συχνοτήτων της πληθοκάλυψης του εδάφους από το φυτό *Philyrea latifolia*.

Στον επόμενο πίνακα παρουσιάζεται μια ακριβέστερη εικόνα της κατανομής της πληθοκάλυψης του εδάφους από το συγκεκριμένο φυτό.

Εμφάνιση φυτού	Συχνότητα	Σχετική συχνότητα (%)	Αθροιστική Σχετική Συχνότητα (%)
Απουσία του φυτού από το δείγμα	218	64,7	64,7
Παρουσία του φυτού στην δειγματοληψία και πληθοκάλυψη του εδάφους έως 5%	69	20,5	85,2
Παρουσία του φυτού στην δειγματοληψία και πληθοκάλυψη του εδάφους από 5% έως 25%	26	7,7	92,9
Παρουσία του φυτού στην δειγματοληψία και πληθοκάλυψη του εδάφους από 25% έως 50%	22	6,5	99,4
Παρουσία του φυτού στην δειγματοληψία και πληθοκάλυψη του εδάφους από 50% έως 75%	2	0,6	100
Σύνολο	337	100,0	

Πίνακας 17: Κατανομή της πληθοκάλυψης του εδάφους από το φυτό *Philyrea latifolia*.

Στον πίνακα παρατηρείται ότι στο 64,7% των δειγματοληψιών δεν παρατηρήθηκε το φυτό. Στο υπόλοιπο 35,3% που υπήρχε το φυτό, στο 20,5% (των δειγματοληψιών) δημιουργούσε πολύ μικρές πληθοκαλύψεις του εδάφους, μέχρι 5%.

Θα αναζητηθεί μοντέλο σύνδεσης μεταξύ της εξαρτημένης μεταβλητής (πληθοκάλυψη του εδάφους από το φυτό *Philyrea latifolia*) και των ανεξάρτητων μεταβλητών (περιβαλλοντικοί παράγοντες). Για τον σκοπό αυτό βρέθηκαν δύο μοντέλα που αναλύονται πιο κάτω.

Με δοκιμές διαπιστώθηκε ότι καταλληλότερη συνάρτηση σύνδεση είναι η **Complementary log-log** με τύπο $\log[-\log(1-P(Y \leq j))] = \hat{\alpha}_j + \hat{\beta} \cdot x$, $j=0,1,2,3$. Ανεξάρτητη ή επεξηγηματική μεταβλητή θα χρησιμοποιηθεί μόνο το υψόμετρο του εδάφους. Οι εκτιμήσεις των συντελεστών του μοντέλου παρουσιάζονται στον επόμενο πίνακα.

	Εκτίμηση	Τυπική Απόκλιση	Τιμή (Wald)	Βαθμοί Ελευθερίας	Επίπεδο Σημαντικότητας
$\hat{\alpha}_0$	-2,535	0,294	74,218	1	0,000
$\hat{\alpha}_1$	-1,727	0,272	40,150	1	0,000
$\hat{\alpha}_2$	-1,310	0,261	25,245	1	0,000
$\hat{\alpha}_3$	-0,542	0,249	4,744	1	0,029
$\hat{\beta}$	$2,92 \cdot 10^{-3}$	$0,307 \cdot 10^{-3}$	90,118	1	0,000

Για κάθε κατηγορία έκβασης οι μορφές του μοντέλου είναι:

$$\log[-\log(1 - P(Y \leq 0))] = -2,535 - 2,92 \cdot 10^{-2} \cdot x$$

$$\log[-\log(1 - P(Y \leq 1))] = -1,727 - 2,92 \cdot 10^{-2} \cdot x$$

$$\log[-\log(1 - P(Y \leq 2))] = -1,310 - 2,92 \cdot 10^{-2} \cdot x$$

$$\log[-\log(1 - P(Y \leq 3))] = -0,542 - 2,92 \cdot 10^{-2} \cdot x$$

Στον προηγούμενο πίνακα βρίσκονται και τα αποτελέσματα του Wald ελέγχου, τα παρατηρούμενα επίπεδα σημαντικότητας όλων των ελέγχων είναι πολύ μικρά. Μικρότερα από τα συνήθη επίπεδα σημαντικότητας. Από αυτά εξάγεται το συμπέρασμα ότι οι πραγματικοί συντελεστές του μοντέλου διαφέρουν σημαντικά από το μηδέν. Η φυσική έννοια του συμπεράσματος ότι $\beta \neq 0$ είναι ότι η μεταβλητή υψόμετρο έχει ερμηνευτική σημασία στην εμφάνιση και ανάπτυξη του συγκεκριμένου φυτού.

Αξιολόγηση του μοντέλου.

Οι τιμές των Pseudo R-square του μοντέλου είναι:

Pseudo R-square	Τιμές
Cox and Snell	0,757
Negelkerke	0,872
McFadden	0,699

Οι τιμές αυτές είναι ικανοποιητικές αυτό σημαίνει ότι το μοντέλο προσαρμόζεται ικανοποιητικά στα δεδομένα.

Η αξιολόγηση του μοντέλου με βάση την προσαρμογή του στα δεδομένα, γίνεται και με έλεγχο, οι υποθέσεις του οποίου είναι:

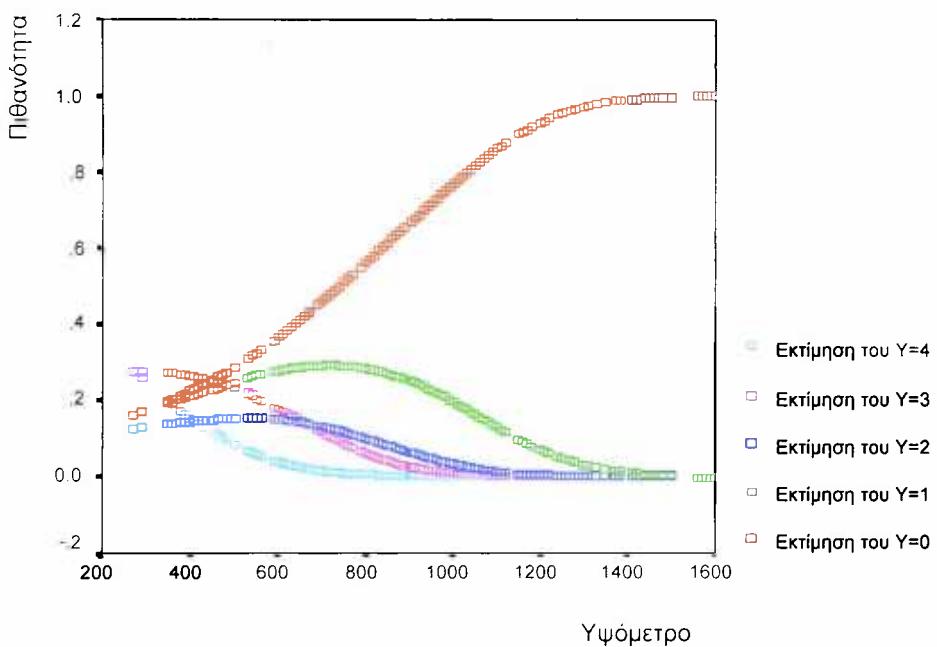
H_0 : Το μοντέλο προσαρμόζεται ικανοποιητικά στα δεδομένα.

H_1 : Αντίθετα.

Τα αποτελέσματα του ελέγχου βρίσκονται στον επόμενο πίνακα.

Goodness-of-fit	Τιμή χ^2	Βαθμοί Ελευθερίας	Επίπεδο Σημαντικότητας
Deviance	225,042	391	1

Το παρατηρούμενο επίπεδο σημαντικότητας του ελέγχου αυτού είναι σχεδόν 1. Μεγαλύτερο από κάθε σύνηθες επίπεδο σημαντικότητας. Αυτό σημαίνει ότι δεν υπάρχουν αρκετές ενδείξεις για απόρριψη της μηδενικής υπόθεσης, δηλαδή δεν μπορεί να θεωρηθεί ότι το μοντέλο δεν έχει καλή προσαρμογή στα δεδομένα. Εξαιτίας αυτού μπορεί να χρησιμοποιηθεί το μοντέλο αυτό, για περιγραφή και πρόβλεψη της πληθοκάλυψης του εδάφους μέσω του υψομέτρου. Οι πιθανότητες πρόβλεψης κάθε κατηγορίας του Y εμφανίζονται στο επόμενο διάγραμμα.



Το μοντέλο προβλέπει ότι στα υψόμετρα έως 400 μέτρα μεγαλύτερη πιθανότητα έχει η παρουσία του φυτού με αρκετά αναπτυγμένη μορφή. Έτσι που να δημιουργεί πληθοκαλύψεις από 25% έως 50% του εδάφους. Στα χαμηλά αυτά υψόμετρα το να απουσιάζει το φυτό είναι αρκετά σπάνιο με πιθανότητα 0,2 – 0,3. Έτσι στα υψόμετρα αυτά πιθανολογείται ότι θα συναντάται το φυτό αρκετά συχνά. Καθώς αυξάνεται το υψόμετρο και ιδιαίτερα πάνω από τα 600- 900 μέτρα, η πιθανότητα να απουσιάζει το φυτό είναι η μεγαλύτερη αλλά όχι μεγαλύτερη από 0,5. Εκεί η κατηγορία να απουσιάζει το φυτό προκρίνεται, αλλά η πιθανότητα της απουσίας δεν είναι μεγαλύτερη από την πιθανότητα της παρουσίας. Στα υψόμετρα αυτά αναμένεται λοιπόν να υπάρχει το φυτό. Η πιθανότερη κάλυψη από το συγκεκριμένο φυτό είναι έως 5%. Καθώς αυξάνεται το υψόμετρο πάνω

από 1000 μέτρα η πιθανότητα απουσίας αυξάνεται με γοργό ρυθμό και πλησιάζει το 1. Από όλα αυτά, το φυτό *Philyrea latifolia* θα μπορούσε να χαρακτηριστεί ως φιλικό σε υψόμετρα έως 500 μέτρα.

Έλεγχος υποθέσεων.

Το παραπάνω μοντέλο εκτιμήθηκε με την προϋπόθεση ότι η επίδραση της ανεξάρτητης μεταβλητής είναι ίδια σε κάθε επίπεδο παρατήρησης της εξαρτημένης μεταβλητής. Δηλαδή ότι η τιμή του β είναι σταθερή για όλα τα επίπεδα της μεταβλητής Y. Η προϋπόθεση αυτή όμως πρέπει να ελεγχθεί. Έτσι οι υποθέσεις που ελέγχονται είναι:

H_0 : Όλες οι κατηγορίες έχουν το ίδιο σετ συντελεστών του υψομέτρου.

H_1 : Αντίθετα.

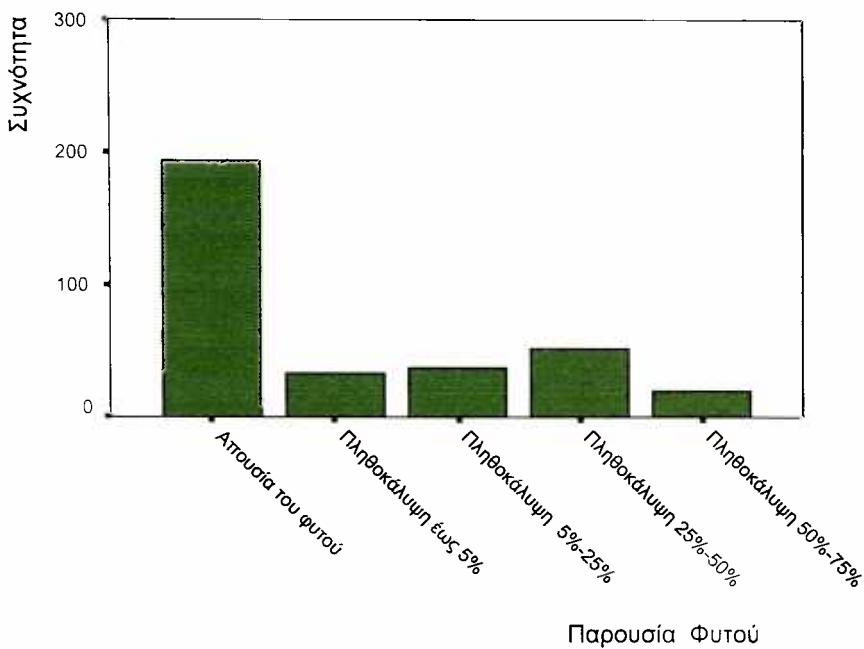
Τα αποτελέσματα δίνονται στον πίνακα:

Μοντέλο	-2log-likelihood	Τιμή χ^2	Βαθμοί Ελευθερίας	Επίπεδο Σημαντικότητας
Null Hypothesis	0,000			
General	0,000	0,000	3	1

Παρατηρείται ότι οι τιμές των πιθανοφανειών των υπό σύγκριση μοντέλων, είναι μηδέν γεγονός που δεν συνάδει με τον ορισμό της πιθανοφάνειας ενός μοντέλου. Εξαιτίας αυτού είναι αδύνατος ο έλεγχος των παραλλήλων γραμμών. Η αδυναμία αυτή δεν ξεπεράστηκε με κανένα εναλλακτικό τρόπο σύνδεσης, ούτε με κάποιον διαφορετικό συνδυασμό επεξηγηματικών μεταβλητών.

4.9. *Brachypodium retusum* (Bayret) Ποώδες αγρωστώδες φυτό.

Το διάγραμμα συχνοτήτων που ακολουθεί δίνει μια πρώτη εικόνα της κατανομής της πληθοκάλυψης του φυτού *Brachypodium retusum*.



Γράφημα 12: Ραβδόγραμμα συχνοτήτων της πληθοκάλυψης του εδάφους από το φυτό *Brachypodium retusum*.

Στον επόμενο πίνακα υπάρχει ακριβέστερη εικόνα της κατανομής της μεταβλητής.

Εμφάνιση φυτού	Συχνότητα	Σχετική συχνότητα (%)	Αθροιστική Σχετική Συχνότητα (%)
Απουσία του φυτού από το δείγμα	194	54,6	54,6
Παρουσία του φυτού στην δειγματοληψία και πληθοκάλυψη του εδάφους έως 5%	33	9,8	67,4
Παρουσία του φυτού στην δειγματοληψία και πληθοκάλυψη του εδάφους από 5% έως 25%	38	11,3	78,6
Παρουσία του φυτού στην δειγματοληψία και πληθοκάλυψη του εδάφους από 25% έως 50%	52	15,4	94,1
Παρουσία του φυτού στην δειγματοληψία και πληθοκάλυψη του εδάφους από 50% έως 75%	20	5,9	100
Σύνολο	337	100,0	

Πίνακας 18: Κατανομή της πληθοκάλυψης του εδάφους από το φυτό *Brachypodium retusum*.

Από στον πίνακα παρατηρείται, ότι στις μισές σχεδόν δειγματοληψίες (54,6%) δεν βρέθηκε το συγκεκριμένο φυτό. Ενώ στις υπόλοιπες που υπήρχε, παρατηρήθηκαν όλες οι

δυνατές μορφές ανάπτυξης, σε κάποια περίπου ομοιομορφία εμφανίσεων. Η αναζήτηση τρόπου σύνδεσης, της πληθοκάλυψης του εδάφους και των περιβαλλοντικών μεταβλητών, οδήγησε στην **Logit** συνάρτηση σύνδεσης και ανεξαρτήτων μεταβλητών, του υψομέτρου και της κλίσης του εδάφους. Το μοντέλο αυτό επιλέγηκε γιατί ήταν εκείνο που ικανοποιούσε την προϋπόθεση της παραλληλίας των σιγμοειδών καμπύλων. Η μορφή του είναι: $\log\left[\frac{P(Y \leq j)}{1 - P(Y \leq j)}\right] = \hat{\alpha}_j + \hat{B} \cdot x = \hat{\alpha}_j + \hat{\beta}_1 \cdot x_1 + \hat{\beta}_2 \cdot x_2$

με $j=0,1,2,3$. x_1 : υψόμετρο σε μέτρα, x_2 : κλίση του εδάφους σε μοίρες.

Αρχικά θα ελεγχθεί εάν οι ανεξάρτητες μεταβλητές βελτιώνουν την προβλεπτική ικανότητα του μοντέλου. Δηλαδή εάν το μοντέλο με τις δυο επεξηγηματικές μεταβλητές έχει καλύτερη προσαρμογή και προβλεπτική ικανότητα από το μοντέλο που δεν περιέχει ανεξάρτητες μεταβλητές. Ο έλεγχος αυτός έχει υποθέσεις :

$$H_0 : B=0 \text{ Δηλαδή } (\beta_1 = \beta_2 = 0)$$

$$H_1 : B \neq 0 \text{ Διαφορετικά}$$

Τα αποτελέσματα του δίνονται στον πίνακα.

Model Fitting Information	-2log-likelihood	Tιμή χ^2	Βαθμοί Ελευθερίας	Επίπεδο Σημαντικότητας
Intercept Only	786,625			
Final	685,700	100,925	2	0,000

Η τιμή της στατιστικής συνάρτησης ελέγχου που ισούται με την διαφορά των πιθανοφανειών των δυο μοντέλων είναι $G=100,925$ Ακολουθεί την χ^2 κατανομή με 2 βαθμούς ελευθερίας. Το παρατηρούμενο επίπεδο σημαντικότητας είναι σχεδόν μηδέν μικρότερο από κάθε σύνηθες επίπεδο σημαντικότητας, έτσι απορρίπτεται η μηδενική υπόθεση. Άρα δεν μπορεί να θεωρηθεί ότι το μοντέλο με τις δυο επεξηγηματικές μεταβλητές δεν είναι καλύτερο (ως προς την προσαρμογή) από το μοντέλο που δεν περιέχει ανεξάρτητες μεταβλητές. Οι συντελεστές του εκτιμώνται όπως έχει αναφερθεί, με την μέθοδο μέγιστης πιθανοφάνειας και παρουσιάζονται στον επόμενο πίνακα.

	Εκτίμηση	Τυπική Απόκλιση	Τιμή (Wald)	Βαθμοί Ελευθερίας	Επίπεδο Σημαντικότητας
$\widehat{\alpha}_0$	-4,133	0,527	61,559	1	0,000
$\widehat{\alpha}_1$	-3,569	0,514	48,214	1	0,000
$\widehat{\alpha}_2$	-2,872	0,501	32,900	1	0,000
$\widehat{\alpha}_3$	-1,287	0,501	6,608	1	0,010
$\widehat{\beta}_1$	$-3,55 \cdot 10^{-3}$	0,001	46,588	1	0,000
$\widehat{\beta}_2$	$-3,27 \cdot 10^{-2}$	0,010	10,608	1	0,001

Το μοντέλο αυτό για κάθε κατηγόρια έκβασης εκτιμάται ότι είναι:

$$\log\left[\frac{P(Y \leq 0)}{1 - P(Y \leq 0)}\right] = -4,133 - 3,55 \cdot 10^{-3} \cdot x_1 - 3,27 \cdot 10^{-2} \cdot x_2$$

$$\log\left[\frac{P(Y \leq 1)}{1 - P(Y \leq 1)}\right] = -3,569 - 3,55 \cdot 10^{-3} \cdot x_1 - 3,27 \cdot 10^{-2} \cdot x_2$$

$$\log\left[\frac{P(Y \leq 2)}{1 - P(Y \leq 2)}\right] = -2,872 - 3,55 \cdot 10^{-3} \cdot x_1 - 3,27 \cdot 10^{-2} \cdot x_2$$

$$\log\left[\frac{P(Y \leq 3)}{1 - P(Y \leq 3)}\right] = -1,287 - 3,55 \cdot 10^{-3} \cdot x_1 - 3,27 \cdot 10^{-2} \cdot x_2$$

Στον παραπάνω πίνακα υπάρχουν και τα αποτελέσματα των Wald ελέγχων. Το μεγαλύτερο ενδιαφέρον έχουν οι έλεγχοι για την σημαντικότητα των ανεξάρτητων μεταβλητών. Οι υποθέσεις των οποίων είναι:

$$H_0 : \beta_j = 0 \quad \text{Me } j=0,1,2.$$

$$H_1 : \beta_j \neq 0$$

Όταν $j=1$ ελέγχεται η σημαντικότητα της μεταβλητής του υψομέτρου. Η τιμή της ελεγχοσυνάρτησης είναι $W= 46,588$ και ακολουθεί την κατανομή χ^2 με έναν βαθμό ελευθερίας. Το παρατηρούμενο επίπεδο σημαντικότητας του ελέγχου σχεδόν μηδέν. Αυτό σημαίνει ότι απορρίπτεται η υπόθεση H_0 , ότι η μεταβλητή υψόμετρο δεν είναι στατιστικά σημαντική για το μοντέλο.

Όταν $j=2$ ελέγχεται η σημαντικότητα της μεταβλητής κλίση του εδάφους. Η τιμή της ελεγχοσυνάρτησης είναι $W=10,608$ και ακολουθεί την κατανομή χ^2 με έναν βαθμό ελευθερίας. Το παρατηρούμενο επίπεδο σημαντικότητας σχεδόν μηδέν. Έτσι απορρίπτεται η μηδενική υπόθεση, ότι η μεταβλητή κλίση του εδάφους δεν έχει ερμηνευτική αξία για το μοντέλο.

Αξιολόγηση του μοντέλου.

Ο βαθμός προσαρμογής του κάθε μοντέλου στα δεδομένα μετράται πρωτίστως με τα Pseudo R-square. Οι τιμές τους για το συγκεκριμένο μοντέλο είναι:

Pseudo R-square	Τιμές
Cox and Snell	0,259
Negelkerke	0,282
McFadden	0,120

Η αξιολόγηση της προσαρμογής του μοντέλου στα δεδομένα γίνεται και με έλεγχο κατά Deviance. Οι υποθέσεις του οποίου είναι:

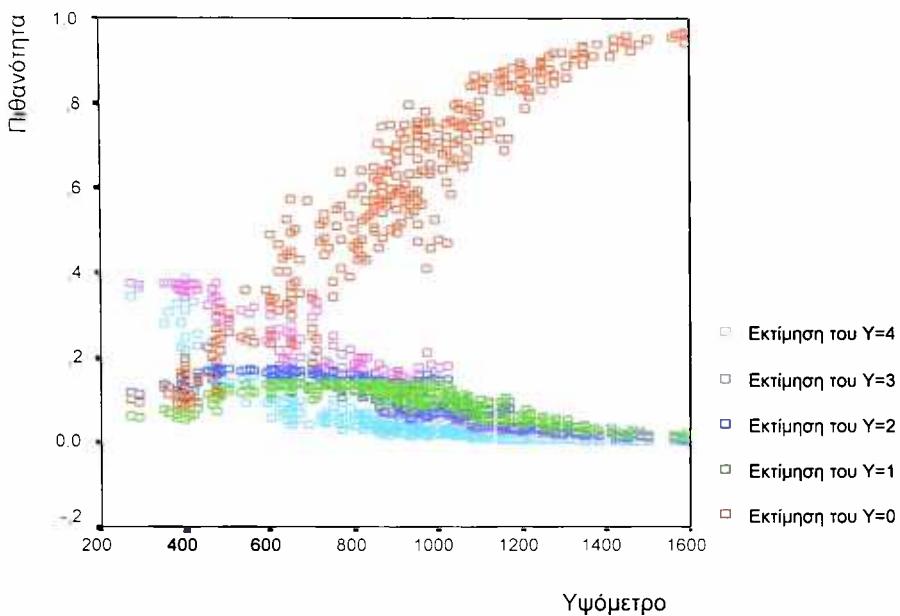
H₀ : Το μοντέλο έχει καλή προσαρμογή στα δεδομένα.

H₁ : Αντίθετα.

Στον επόμενο πίνακα δίνονται τα αποτελέσματα του ελέγχου..

Goodness-of-fit	Τιμή χ^2	Βαθμοί Ελευθερίας	Επίπεδο Σημαντικότητας
Deviance	636,577	1098	1

Η τιμή της συνάρτησης ελέγχου είναι $G=636,577$ ακολουθεί την χ^2 κατανομή με 1098 βαθμούς ελευθερίας. Το παρατηρούμενο επίπεδο σημαντικότητας είναι σχεδόν 1, μεγαλύτερο από κάθε σύνηθες επίπεδο σημαντικότητας. Άρα δεν υπάρχουν αρκετές ενδείξεις απόρριψης της μηδενικής υπόθεσης. Έτσι δεν μπορεί να μην γίνει δεκτό ότι το μοντέλο έχει καλή προσαρμογή. Για αυτό θα χρησιμοποιηθεί για να προβλεφθεί η πληθοκάλυψη του φυτού. Οι προβλέψεις και οι πιθανότητες κάθε κατηγορίας έκβασης δίνονται στο επόμενο διάγραμμα.



Στο διάγραμμα παρατηρείται ότι σε υψόμετρα μεγαλύτερα από 800 περίπου μέτρα, το μοντέλο προκρίνει το ενδεχόμενο της απουσίας του φυτού. Στα μικρά υψόμετρα, μέχρι τα 800-900 πιθανολογείται η παρουσία του φυτού. Η κάλυψη που δημιουργεί το φυτό εκεί είναι μεγάλη και θα είναι πάνω από 25%. Από τα παραπάνω θα μπορούσε να βγει το συμπέρασμα ότι το φυτό Brachypodium retusum όταν υπάρχει, θα δημιουργεί και μεγάλες πληθοκαλύψεις.

Έλεγχος υποθέσεων.

Το παραπάνω μοντέλο εκτιμήθηκε με την προϋπόθεση ότι η επίδραση του διανύσματος των ανεξάρτητων μεταβλητών είναι ίδια σε κάθε επίπεδο παρατήρησης της εξαρτημένης μεταβλητής. Αυτή η υπόθεση όμως πρέπει να ελεγχθεί, έτσι οι υποθέσεις του ελέγχου είναι:

H_0 : Όλες οι κατηγορίες έχουν το ίδιο σετ συντελεστών.

H_1 : Αντίθετα.

Τα αποτελέσματα δίνονται στον πίνακα:

Μοντέλο	-2log-likelihood	Τιμή χ^2	Βαθμοί Ελευθερίας	Επίπεδο Σημαντικότητας
Null Hypothesis	685,700			
General	673,526	12,175	6	0,058

Η τιμή της ελεγχοσυνάρτησης είναι $G=12,175$. και ακολουθεί την χ^2 κατανομή με $\text{d.f.}=6$ βαθμούς ελευθερίας. Το παρατηρούμενο επίπεδο σημαντικότητας είναι $p-value=0,058$.

μεγαλύτερο από κάθε σύνηθες επίπεδο σημαντικότητας. Αυτό σημαίνει ότι δεν υπάρχουν αρκετές ενδείξεις ώστε να απορρίπτεται η μηδενική υπόθεση. Έτσι δεν μπορεί να απορριφθεί η υπόθεση ότι όλες οι κατηγορίες έκβασης της εξαρτημένης μεταβλητής έχουν το ίδιο διάνυσμα συντελεστών των ανεξάρτητων μεταβλητών.

ΚΕΦΑΛΑΙΟ 5. ΣΥΜΠΕΡΑΣΜΑΤΑ

Με την ολοκλήρωση της εργασίας επιχειρείται μια προσπάθεια ανακεφαλαίωσης των παρατηρήσεων και των συμπερασμάτων που εξήχθησαν κατά την εκπόνηση της. Τα πρώτα συμπεράσματα προέκυψαν από την στατιστική ανάλυση των περιβαλλοντικών – οικολογικών παραγόντων που καταγράφηκαν. Με τον τρόπο αυτό, οι σχέσεις των μεταβλητών αντανακλούν πολύ εύγλωττα την πραγματική μορφολογική σύσταση του συγκεκριμένου Όρους, αλλά και την δειγματοληπτική τεχνική που εφαρμόσθηκε. Αρχικά διαφαίνεται ότι η κατανομή τόσο του υψομέτρου όσο και της κλίσης του εδάφους εμφανίζουν συμμετρικότητα. Καθώς και ότι υπάρχει θετική συσχέτιση μεταξύ των μεταβλητών υψομέτρου και κλίσης του εδάφους των δειγματοληψιών. Αυτό εξηγείται από το ότι πηγαίνοντας σε μεγαλύτερα υψηλότερα στο συγκεκριμένο όρος, συναντώνται και μεγαλύτερες κλίσεις. Ακόμη παρατηρείται ότι το είδος του πετρώματος του εδάφους εξαρτάται από το υψόμετρο. Η εξάρτηση που υπάρχει είναι η εξής: Στα υψηλά υψόμετρα υπάρχουν (περισσότερο του αναμενόμενου) ασβεστολιθικά πετρώματα, στα ενδιάμεσα πετρώματα του ραδιολαρίτη και στα χαμηλά πετρώματα από φλύσχη. Αποδεικνύεται στατιστικά ότι υπάρχει επίσης εξάρτηση μεταξύ του πετρώματος του εδάφους και της κλίσης του. Συγκεκριμένα ο φλύσχης προτιμά τις μικρές κλίσεις ο ραδιολαρήτης είναι φιλικότερος στις ενδιάμεσες κλίσεις και ο ασβεστόλιθος τις μεγάλες. Δηλαδή στα ασβεστολιθικά πετρώματα δημιουργούνται κατακρήμνισης και σχηματίζονται γκρεμοί. Η μορφολογία αυτή, για το συγκεκριμένο όρος παρατηρείται στα υψηλά υψόμετρα, ένα γεγονός που διαπιστώνεται και με οπτική παρατήρηση του όρους Ερυμάνθου. Ενδιαφέρον έχει ακόμη και η εξάρτηση του πετρώματος και του προσανατολισμού του εδάφους. Παρατηρείται ότι, ο ασβεστόλιθος προτιμά την νότια έκθεση του όρους Ερυμάνθου, ο φλύσχης την δυτική και ο ραδιολαρήτης την βόρεια. Αξίζει επίσης να επισημανθεί ότι υπάρχει σχέση εξάρτησης μεταξύ των μεταβλητών προσανατολισμός δειγματοληψιών και υψόμετρου του εδάφους. Αυτό που διακρίνεται είναι ότι στα ανατολικά του όρους δεν υπάρχουν μικρά υψόμετρα, γεγονός που οφείλεται στη μορφολογία της ευρύτερης ορεινής περιοχής. Ακόμη στην δυτική πλευρά ελήφθησαν πολλές δειγματοληψίες στα χαμηλά υψόμετρα. Αυτό έγινε ώστε να επιτευχθεί η αντιπροσωπευτικότατα της χλωρίδας. Για τον ίδιο λόγο στην νότια έκθεση, πραγματοποιήθηκαν περισσότερες δειγματοληψίες στα μεγάλα υψόμετρα. Στατιστική τεκμηρίωση επίσης υπάρχει για την σχέση εξάρτησης μεταξύ της κλίσης του εδάφους και του προσανατολισμού των δειγματοληψιών. Πιο συγκεκριμένα στα δυτικά παρατηρούνται περισσότερες δειγματοληψίες στις μικρές

κλίσεις. Αυτό δικαιολογείται επειδή στην έκθεση αυτή υπάρχει περισσότερος φλύσχης, πέτρωμα που δεν δημιουργεί μεγάλες κλίσεις.

Στο επόμενο κεφάλαιο επιχειρείται μια παρουσίαση των προβλέψεων της δυναμικής της βλάστησης του Όρους Ερυμάνθου. Τα αποτελέσματα έτσι όπως προκύπτουν με την μέθοδο του proportional – odds μοντέλου, για κάποιες από τις παρατηρούμενες μονάδες βλάστησης είναι:

- **Anemone Blanda.** Προβλέπει την απουσία του από την χλωρίδα του Όρους. Η πιθανότητα του να μην παρατηρηθεί το συγκεκριμένο είδος, είναι αρχικά μεγάλη και σταδιακά μειώνεται με την αύξηση του υψόμετρου. Η σύνδεση που χρησιμοποιείται είναι complementary log-log. Ως ανεξάρτητη μεταβλητή χρησιμοποιείται το υψόμετρο του εδάφους, και δεν μπορεί να μην θεωρηθεί στατιστικά σημαντική. Τα Pseudo R-square του μοντέλου είναι έως 4,7%.
- **Quercus coccifera** (πουρνάρι). Εμφανίζεται φιλικότερο στα μικρά υψόμετρα και περισσότερο διαδεδομένο εκεί. Ο τρόπος σύνδεσης του μοντέλου είναι Negative log-log. Επεξηγηματική μεταβλητή είναι το υψόμετρο του εδάφους και δεν μπορεί να μην θεωρηθεί στατιστικά σημαντική για το μοντέλο. Οι τιμές των Pseudo R-square του μοντέλου φτάνουν έως 15,3%.
- **Rosa canina.** Προβλέπει ότι το φυτό σπάνια μπορεί να παρατηρηθεί στο όρος Ερύμανθο. Το πιθανότερο είναι να εμφανιστεί στα υψηλά υψόμετρα με μικρή όμως κάλυψη. Η συνάρτηση σύνδεσης είναι η logit. Ανεξάρτητη μεταβλητή είναι επίσης το υψόμετρο του εδάφους και δεν μπορεί να μην θεωρηθεί στατιστικά σημαντική για το μοντέλο. Οι τιμές για τα Pseudo R-square του μοντέλου είναι έως 3,2%.
- **Campanula spatulata.** Προβλέπει ότι το φυτό εμφανίζεται σε υψόμετρα άνω των 1000 μέτρων. Η αναμενόμενη κάλυψη είναι έως 5%. Η συνάρτηση σύνδεσης του μοντέλου είναι η Cauchit. Η ανεξάρτητη μεταβλητή δεν μπορεί να μην θεωρηθεί στατιστικά σημαντική και είναι το υψόμετρο του εδάφους. Τα Pseudo Pseudo R-square για το μοντέλο είναι μέχρι 10,6%.
- **Acer mouspesulanum.** Προβλέπει ότι το φυτό είναι σπάνιο στην χλωρίδα του όρους Ερυμάνθου. Τα πιθανότερα υψόμετρα που μπορεί να παρατηρηθεί, είναι τα μεσαία με μικρή ωστόσο πληθοκάλυψη. Το link του μοντέλου είναι Probit. Ως επεξηγηματικές μεταβλητές λαμβάνεται το υψόμετρο και το υψόμετρο στη δευτέρα. Οι δύο αυτές μεταβλητές δεν μπορεί να μην θεωρηθούν στατιστικά σημαντικές για το μοντέλο. Το Pseudo R-square έχει τιμές έως 13,4%.

- **Festuca jeepertii.** Για την συγκεκριμένη μονάδα βλάστησης χρησιμοποιήθηκαν δυο διαφορετικά μοντέλα περιγραφής και πρόβλεψης. Ως συνάρτηση σύνδεσης επιλέγεται η Complementary log-log. Το 1^o μοντέλο έχει μια επεξηγηματική μεταβλητή μόνο το υψόμετρο του εδάφους η οποία είναι στατιστικά σημαντική. Η τιμή των Pseudo R-square είναι έως 7,1%. Το μοντέλο προβλέπει ότι το φυτό δεν θα παρατηρηθεί στο όρος Ερύμανθο. Με πιθανότητα που μειώνεται με την αύξηση του υψομέτρου. Το 2^o μοντέλο έχει ως ανεξάρτητες μεταβλητές το υψόμετρο και την κλίση του εδάφους. Για τις μεταβλητές αυτές δεν υπάρχουν αρκετές ενδείξεις ώστε να μην θεωρηθούν στατιστικά σημαντικές. Το μοντέλο με τις δυο επεξηγηματικές μεταβλητές έχει υψηλότερες τιμές των Pseudo R-square που φτάνουν έως 11,1%. Οι γενικές προβλέψεις όμως ως προς την παρουσία και την πληθοκάλυψη του φυτού, δεν παρουσιάζουν σημαντικές διαφοροποιήσεις από το 1^o μοντέλο.
- **Philyrea latifolia.** Το μοντέλο εδώ παρουσιάζει εξαιρετικά υψηλές τιμές στα Pseudo R-square που φτάνουν έως 87,2%. Η συνάρτηση σύνδεσης είναι η Complementary log-log. Η ανεξάρτητη μεταβλητή είναι το υψόμετρο και θεωρείται στατιστικά σημαντική για το μοντέλο. Προβλέπει ότι το φυτό παρατηρείται σε υψόμετρα έως 500 μέτρα, σε αρκετά ανεπτυγμένη μορφή (δέντρο). Σε μεγαλύτερα υψόμετρα πιθανολογείται ότι απουσιάζει το συγκεκριμένο φυτό, αλλά εάν παρατηρηθεί αναμένεται μην δημιουργεί μεγάλες καλύψεις. Για το μοντέλο αυτό όμως δεν είναι δυνατός ο έλεγχος παραλληλίας των γραμμών. Έτσι δεν είναι αξιόπιστες οι προβλέψεις του.
- **Brachypodium retusum.** Προβλέπει την παρουσία του σε υψόμετρα έως 800 μέτρα. Εκεί θα δημιουργεί πολύ μεγάλες πληθοκαλύψεις. Ως link του μοντέλου λαμβάνεται το logit. Ανεξάρτητες μεταβλητές είναι δυο, το υψόμετρο και η κλίση του εδάφους οι οποίες από τα υπάρχοντα στοιχειά δεν μπορούν να μην θεωρηθούν στατιστικά σημαντικές. Οι τιμές των Pseudo R-square φτάνουν έως 28,2%.

Από την εφαρμογή του μοντέλου **proportional – odds**, στις παραπάνω μονάδες βλάστησης προκύπτει ότι είναι δυνατή η χρήση του για περιγραφή και πρόβλεψη της δυναμικής της βλάστησης του Όρους Ερυμάνθου. Μέσω των συγκεκριμένων περιβαλλοντικών παραγόντων. Έτσι θα μπορούσε να ειπωθεί ότι το μοντέλο προσαρμόζεται, γενικά, στα δεδομένα. Αφού οι επεξηγηματικές μεταβλητές διαπιστώθηκε ότι ήταν πάντα στατιστικά σημαντικές. Η ένταση της όμως δεν μπορεί να χαρακτηριστεί ιδιαίτερα ισχυρή. Αυτό προκύπτει από τις μικρές σχετικά τιμές των Pseudo R-square. Η αύξηση των τιμών, αυτών επιτυγχάνεται με την προσθήκη ανεξάρτητων μεταβλητών, έχοντας κόστος βέβαια την απλότητα του μοντέλου. Στις περιπτώσεις που δεν ήταν ικανή

η μοντελοποίηση, εξαιτίας του ότι προέκυπταν μοντέλα που κατά τον έλεγχο της παραλληλίας των γραμμών δεν ήταν αποδεκτά. Διαπιστώθηκε, ότι αυτά είχαν ιδιαίτερα υψηλές τιμές των Pseudo R-square.

ΚΕΦΑΛΑΙΟ 6. ΒΙΒΛΙΟΓΡΑΦΙΑ

- Agresti A. (1996). An Introduction to Categorical Data Analysis. John Wiley & Sons, Inc.
- Braun – Blanquet J. 1964. Pflanzensociologie – Grundzüge der Vegetationskunde. 3. Aufl. Springer.
- Draper N., Smith H. (1997). Applied regression analysis. 2nd edition. John Wiley & Sons, Inc.
- Guisan A., Harrell F. (2000). Ordinal response models in ecology. Journal of Vegetation Science, 11: 617-626.
- Hosmer D.W., Lemeshow S. (1989). Applied Logistic Regression. John Wiley & Sons, Inc.
- Jongman R. H. G., ter Braak C.J.F. & van Tongeren O.F.R. (1995) Data analysis in community and landscape ecology. Cambridge University Press.
- Kent M. & Coker P. 1992. Vegetation Description and Analysis. John Wiley & Sons., Chichester.
- Kleinbaum D., Kupper L., Muller K. & Nizam A. (1998). Applied regression analysis and other multivariate methods. Duxbury Press.
- Lemmel A. Moyer. Multiple Analyses in Clinical Trials.
- Μαρούλης Γ. (2003). Χλωρίδα και Βλάστηση των Οικοσυστημάτων του όρους Ερυμάνθου (ΒΔ Πελοπόννησος). Διδακτορική Διατριβή. Πανεπιστήμιο Πατρών.
- Mc Cullagh P. (1980). Regression models for Ordinal Data. J. R. Statist. Soc, 42, No 2, 109 – 142.
- Πανάρετος Ι., Ξεκαλάκη Ε. (2000). Εισαγωγή στη στατιστική σκέψη. Ergo μ&p Γραφικές τέχνες Ε.Π.Ε.
- Πανάρετος Ι. (2001). Γραμμικά μοντέλα με έμφαση στις εφαρμογές. Ergo μ&p Γραφικές τέχνες Ε.Π.Ε.
- Scott S., Goldberg M., Mayo N. (1997). Statistical Assessment of Ordinal Outcomes in Comparative Studies. J. Clin. Epidemiol. Vol. 50, No 1, pp. 45 – 55.
- Venables W. N. & Ripley B.D. (1994) Modern applied statistics with s-plus. Springer-Verlag.

