



ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ – ΓΕΩΠΟΝΙΚΗ ΣΧΟΛΗ
ΤΜΗΜΑ ΓΕΩΠΟΝΙΑΣ, ΦΥΤΙΚΗΣ ΠΑΡΑΓΩΓΗΣ & ΑΓΡΟΤΙΚΟΥ ΠΕΡΙΒΑΛΛΟΝΤΟΣ

ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
“ΕΠΙΣΤΗΜΕΣ ΚΑΙ ΣΥΣΤΗΜΑΤΑ ΑΕΙΦΟΡΟΥ ΦΥΤΙΚΗΣ ΠΑΡΑΓΩΓΗΣ”

Διπλωματική Εργασία
«ΓΕΩΡΓΙΚΟΣ ΠΕΙΡΑΜΑΤΙΣΜΟΣ ΚΑΙ ΕΦΑΡΜΟΓΕΣ ΤΗΣ R ΣΤΗ
ΓΕΩΡΓΙΑ »

ΓΚΟΥΓΚΟΥΔΗ ΑΝΘΗ

ΒΟΛΟΣ 2023

Copyright © Γκουγκούδη Ανθή 2023

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας διατριβής, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης.

Η έγκριση της Μεταπτυχιακής Διατριβής Ειδίκευσης από το Τμήμα Γεωπονίας Φυτικής Παραγωγής και Αγροτικού Περιβάλλοντος του Πανεπιστημίου Θεσσαλίας δε δηλώνει αποδοχή των γνώμων του συγγραφέα.

«ΓΕΩΡΓΙΚΟΣ ΠΕΙΡΑΜΑΤΙΣΜΟΣ ΚΑΙ ΕΦΑΡΜΟΓΗ ΤΗΣ R ΣΤΗ ΓΕΩΡΓΙΑ»

«AGRICULTURAL EXPERIMENTATION AND APPLICATIONS OF R IN AGRICULTURE»

Γκουγκούδη Ανθή

Εγκρίθηκε από τα Μέλη της Τριμελούς Εξεταστικής Επιτροπής:

Πρώτος Εξεταστής (Επιβλέπων)

Δρ. Νάκας Χρήστος

Καθηγητής, γνωστικό αντικείμενο «Βιομετρία», Τμήμα Γεωπονίας Φυτικής παραγωγής και Αγροτικού Περιβάλλοντος, Πανεπιστήμιο Θεσσαλίας

Δεύτερος Εξεταστής

Δρ. Δαναλάτος Νικόλαος

Καθηγητής, γνωστικό αντικείμενο «Γεωργία και Οικολογία φυτών», Τμήμα Γεωπονίας Φυτικής παραγωγής και Αγροτικού Περιβάλλοντος, Πανεπιστήμιο Θεσσαλίας

Τρίτος Εξεταστής

Δρ. Λεβίζου Ευθυμία

Αναπληρώτρια Καθηγήτρια, γνωστικό αντικείμενο «Φυσιολογία Φυτών», Τμήμα Γεωπονίας Φυτικής παραγωγής και Αγροτικού Περιβάλλοντος, Πανεπιστήμιο Θεσσαλίας

Ευχαριστίες

Στο σημείο αυτό θα ήθελα να ευχαριστήσω θερμά τον επιβλέποντα καθηγητή μου κ. Νάκα Χρήστο για την ευκαιρία που μου έδωσε να με την ανάθεση της συγκεκριμένης διπλωματικής εργασίας, για την πολύτιμη καθοδήγηση του και την άριστη συνεργασία καθ' όλη την διάρκεια εκπόνησής της. Επίσης, ευχαριστώ θερμά τη Μπατάκα Έφη για τη ανεκτίμητη βοήθεια που μου παρείχε για τη διεξαγωγή της παρούσας εργασίας. Και τέλος, όλους τους ανθρώπους, συγγενείς και φίλους, που με στήριξαν κατά την διάρκεια των μεταπτυχιακών μου σπουδών.

Γκουγκούδη Ανθή

Περίληψη

Ο γεωργικός πειραματισμός αφορά στη συστηματική διαδικασία οργάνωσης πειραμάτων και συλλογής αξιόπιστων δεδομένων που αφορούν τη γεωργία. Σε γενικές γραμμές, καθορίζει τον τρόπο με τον οποίο θα διεξαχθεί ένα πείραμα, τις στατιστικές τεχνικές που θα χρησιμοποιηθούν για την οργάνωση, τη συνοπτική παρουσίαση και ανάλυσή των δεδομένων, με στόχο την εξαγωγή χρήσιμων συμπερασμάτων. Τόσο ο γεωργικός πειραματισμός όσο και οι στατιστικές τεχνικές, αποτελούν χρήσιμα εργαλεία σε πολλά και διαφορετικά πλαίσια επιστημονικής έρευνας.

Στη παρούσα εργασία θα παρουσιαστούν θα εξεταστούν και θα αναλυθούν αξιόπιστα δεδομένα ποιοτικών χαρακτηριστικών της ίνας του βαμβακιού από εκκοκκιστήριο στη περιοχή της Θεσσαλίας, προσαρμοσμένα σε κοινά πειραματικά σχέδια μέσω της R γλώσσας προγραμματισμού. Οι μετρήσεις αφορούν τρεις σημαντικές ποικιλίες, από πλευράς καλλιεργούμενης έκτασης τα έτη 2017-2021, από τρεις διαφορετικές περιοχές. Διερευνάται η επίδραση των ποικιλιών, των περιοχών καλλιέργειας και των μετεωρολογικών δεδομένων της εκάστοτε χρονιάς, στις τιμές των χαρακτηριστικών ποιότητας της ίνας. Ειδικότερα, εξετάζονται τα χαρακτηριστικά ποιότητας *micronaire*, αντοχή και μήκος ίνας.

Μετά την επεξεργασία των δεδομένων προέκυψε ότι οι ποικιλίες, οι περιοχές ενδιαφέροντος και οι μετεωρολογικές συνθήκες επιδρούν στις τελικές τιμές των χαρακτηριστικών ποιότητας της ίνας, ωστόσο κρίνεται απαραίτητο να διεξαχθούν περαιτέρω έρευνες που σχετίζονται με τα ποιοτικά χαρακτηριστικά της ίνας του βαμβακιού και τους παράγοντες που τα επηρεάζουν.

Λέξεις Κλειδιά: , ανάλυση δεδομένων, λήψη απόφασης, λογισμικό στατιστικής ανάλυσης, ποιοτικά χαρακτηριστικά ίνας, ποσοτικά χαρακτηριστικά ίνας

Abstract

Agricultural experimentation refers to the systematic process of organizing experiments and collecting reliable data related to agriculture. In general, it defines how an experiment will be conducted, the statistical techniques to be used to organize, summarize and analyze the data, and the aim of drawing useful conclusions. Both agricultural experimentation and statistical techniques are useful tools in many different contexts of scientific research. In this paper, reliable data of quality characteristics of cotton fiber from a gin in Thessaly, adapted to common experimental designs through the R programming language, will be presented, examined, and analyzed. The measurements concern three important varieties, in terms of cultivated area, for the years 2017-2021, from three different areas. The influence of the varieties, the cultivation areas, and the meteorological data of each year on the values of the fiber quality characteristics is investigated. In particular, the quality characteristics of micronaire, strength, and fiber length are examined.

After processing the data, it was found that varieties, areas of interest, and meteorological conditions affect the final values of fiber quality characteristics. However, it is necessary to conduct further research related to cotton fiber quality characteristics and the factors that influence them.

Key Words: data analysis, decision making, qualitative fiber characteristics, quantitative fiber characteristics, statistical analysis software

Πίνακας Περιεχομένων

ΚΕΦΑΛΑΙΟ 1: ΕΙΣΑΓΩΓΗ – ΑΝΤΙΚΕΙΜΕΝΟ ΜΕΛΕΤΗΣ	1
1.1 Αντικείμενο και Στόχοι της Παρούσας Διατριβής	1
1.2 Διάρθρωση και Ανάλυση των Περιεχομένων της Διατριβής	1
ΚΕΦΑΛΑΙΟ 2: ΛΕΠΤΟΜΕΡΙΕΣ ΓΙΑ ΤΗΝ ΙΝΑ ΤΟΥ ΒΑΜΒΑΚΙΟΥ & ΠΑΡΟΥΣΙΑΣΗ ΔΕΔΟΜΕΝΩΝ ΠΕΙΡΑΜΑΤΟΣ	2
2.1 Φύση του βαμβακιού και βάση της ταξινόμησής του	2
2.2 Φυσιολογία της ανάπτυξης των ινών	2
2.3 Εξέλιξη στον έλεγχο της ίνας	3
2.4 Το ενδιαφέρον στα ποιοτικά χαρακτηριστικά της ίνας	3
2.4.1 Τα 3 στάδια παραγωγής και επεξεργασίας βαμβακιού-ίνας	3
2.4.2 Ο παραγωγός- Το εκκοκκιστήριο- Το κλωστήριο	4
2.4.3 Παράγοντες που επηρεάζουν την ανάπτυξη της ίνας	5
2.4.4 Επίδραση του περιβάλλοντος στα ποιοτικά χαρακτηριστικά της ίνας .	5
2.5 Τα δεδομένα του πειράματος	6
2.6 Ποιοτικά χαρακτηριστικά ίνας	7
2.6.1 Αντοχή της ίνας (strength)	7
2.6.2 Λεπτότητα και ωριμότητα της ίνας (micronaire).	7
2.6.3 Μήκος ίνας (length)	7
2.6.4 Τιμές ποιοτικών χαρακτηριστικών ίνας	8
ΚΕΦΑΛΑΙΟ 3: ΑΝΑΛΥΣΗ ΚΟΙΝΩΝ ΠΕΙΡΑΜΑΤΙΚΩΝ ΣΧΕΔΙΩΝ ΜΕ ΠΑΡΑΔΕΙΓΜΑΤΑ	9
3.1 Ανάλυση δεδομένων με Πλήρως Τυχαιοποιημένο Σχέδιο (Completely Randomized Design)	9
3.1.1 Το Πλήρως τυχαιοποιημένο σχέδιο (CRD)	9
3.1.2 Για το ποιοτικό χαρακτηριστικό: Μήκος ίνας	9
3.1.3 Για το ποιοτικό χαρακτηριστικό: MICRONAIRE	14
3.1.4 Για το ποιοτικό χαρακτηριστικό: Αντοχή της ίνας	18
3.2 Ανάλυση δεδομένων με Σχεδιασμός τυχαιοποιημένων ομάδων (Randomized Block Design)	21
3.2.1 Ο Σχεδιασμός τυχαιοποιημένων ομάδων	21
3.2.2 Για το ποιοτικό χαρακτηριστικό: Αντοχή της ίνας	21
3.2.3 Για το ποιοτικό χαρακτηριστικό : Micronaire	28
3.2.4 Για το ποιοτικό χαρακτηριστικό: Μήκος ίνας	34

3.3 Ανάλυση δεδομένων με Ανάλυση συνδιακύμανσης (analysis of covariance).....	39
3.3.1 Η Ανάλυση συνδιακύμανσης (ANCOVA).....	39
3.3.2 Για το ποιοτικό χαρακτηριστικό : MICRONAIRE	40
3.3.3 Για το ποιοτικό χαρακτηριστικό : Αντοχή της ίνας	47
3.3.4 Για το ποιοτικό χαρακτηριστικό : Μήκος ίνας	54
ΚΕΦΑΛΑΙΟ 4: ΣΤΑΤΙΣΤΙΚΗ ΑΝΑΛΥΣΗ ΠΟΙΟΤΙΚΩΝ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ ΙΝΑΣ ΒΑΜΒΑΚΙΟΥ ΜΕ ΒΑΣΗ ΤΟ ΠΡΑΓΜΑΤΙΚΟ ΠΛΑΝΟ ΔΙΕΞΑΓΩΓΗΣ ΕΡΕΥΝΑΣ ...	60
4.1 Για τη μεταβλητή: MICRONAIRE	60
4.2 Για τη μεταβλητή Μήκος ίνας (LENGTH).....	64
4.3 Για τη μεταβλητή : Αντοχή της ίνας (STRENGTH).....	69
ΚΕΦΑΛΑΙΟ 5: ΣΥΜΠΕΡΑΣΜΑΤΑ- ΠΡΟΤΑΣΕΙΣ	74
5.1 ΣΥΖΗΤΗΣΗ	74
5.1.1 Χαρακτηριστικό ποιότητας : Micronaire	74
5.1.2 Χαρακτηριστικό ποιότητας : Αντοχή της ίνας	74
5.1.3 Χαρακτηριστικό ποιότητας : Μήκος της ίνας	75
5.2 Προτάσεις για περαιτέρω έρευνα	76

Πίνακας Πινάκων

Πίνακας 2-1: Όρια τιμών ποιοτικών χαρακτηριστικών ίνας.....	8
Πίνακας 3-1: Δεδομένα πριν την τυχαιοποίηση CRD	10
Πίνακας 3-2: Δεδομένα μετά την τυχαιοποίηση CRD	10
Πίνακας 3-3: Τυχαιοποίηση στην RBD	22
Πίνακας 4-1: Το τελικό μοντέλο με τις αλληλεπιδράσεις	61
Πίνακας 4-2: Επιδράσεις των μεταχειρίσεων.....	61
Πίνακας 4-3: Το τελικό μοντέλο με τις αλληλεπιδράσεις.....	65
Πίνακας 4-4: Επιδράσεις των μεταχειρίσεων.....	65
Πίνακας 4-5: Το τελικό μοντέλο με τις αλληλεπιδράσεις.....	69
Πίνακας 4-6: Επιδράσεις των μεταχειρίσεων.....	70

Πίνακας Διαγραμμάτων

Διάγραμμα 3-1: Ανά ζεύγη συγκρίσεις emmeans	13
Διάγραμμα 3-2: Ανά ζεύγη συγκρίσεις emmeans	16
Διάγραμμα 3-3: Boxplot κατανομής μετρήσεων micronaire	17
Διάγραμμα 3-4: Ανά ζεύγη συγκρίσεις emmeans	20
Διάγραμμα 3-5: Q-Q plot κανονικής κατανομής	23
Διάγραμμα 3-6: Boxplot τιμών αντοχής της ίνας	25
Διάγραμμα 3-7:Ραβδόγραμμα μέσης αντοχής ίνας	27
Διάγραμμα 3-8: Q-Q plot κανονικότητας	29
Διάγραμμα 3-9: Boxplot κατανομής τιμών micronaire	31
Διάγραμμα 3-10:Ραβδόγραμμα μέσων τιμών micronaire	33
Διάγραμμα 3-11: Q-Q κανονικότητας	35
Διάγραμμα 3-12: Boxplot κατανομής τιμών μήκους ίνας	37
Διάγραμμα 3-13: Ραβδόγραμμα μέσου μήκους ίνας	39
Διάγραμμα 3-14: διαγράμματα δεδομένων τιμών micronaire	42
Διάγραμμα 3-15:Διάγραμμα διασποράς	46
Διάγραμμα 3-16: Οπτικοποίηση δεδομένων αντοχής ίνας	48
Διάγραμμα 3-17: διάγραμμα διασποράς αντοχής ίνας	52
Διάγραμμα 3-18: Οπτικοποίηση δεδομένων	55
Διάγραμμα 3-19: διάγραμμα διασποράς	59
Διάγραμμα 4-1: Normal Q-Q plot για τα υπόλοιπα του μοντέλου	63
Διάγραμμα 4-2: Q-Q Plot Διάγραμμα κανονικότητας	64
Διάγραμμα 4-3: Normal Q-Q plot για τα υπόλοιπα του μοντέλου	67
Διάγραμμα 4-4: Q-Q Plot Διάγραμμα κανονικότητας	68
Διάγραμμα 4-5: Normal Q-Q plot για τα υπόλοιπα του μοντέλου	71
Διάγραμμα 4-6: Q-Q Plot Διάγραμμα κανονικότητας	73

ΚΕΦΑΛΑΙΟ 1: ΕΙΣΑΓΩΓΗ – ΑΝΤΙΚΕΙΜΕΝΟ ΜΕΛΕΤΗΣ

1.1 Αντικείμενο και Στόχοι της Παρούσας Διατριβής

Η παρούσα εργασία δημιουργήθηκε για δύο κυρίως λόγους.

- ✓ Πρώτον, για τη παρουσίαση πειραματικών σχεδίων και διαδικασιών και την αντιμετώπισή τους μέσω της R γλώσσας προγραμματισμού
- ✓ Και δεύτερον, για την ανάγκη της αξιοποίησης των δεδομένων της ποιότητας της ίνας του βαμβακιού, τα οποία συγκεντρώθηκαν με στόχο τη ανάλυσή τους και τη διεξαγωγή συμπερασμάτων.

Η εργασία ασχολείται με τα ποιοτικά χαρακτηριστικά της ίνας του βαμβακιού που προορίζεται για κλωστήρια. Τα ποιοτικά χαρακτηριστικά σχετίζονται άμεσα με τη ποιότητα του σύσπορου και κατ'επέκταση εκκοκκισμένου βαμβακιού και εν συνεχεία με την τιμή πώλησης του. Τα ποιοτικά χαρακτηριστικά του εκκοκκισμένου βαμβακιού (micronaire, ομοιομορφία, αντοχή, μήκος ίνας, απόδοση σε εκκοκκισμένο βαμβάκι, ανάκλαση, χρωματισμός) εξαρτώνται κυρίως από:

- ✓ την ποικιλία,
- ✓ τις συνθήκες καλλιέργειας,
- ✓ τις καιρικές συνθήκες,
- ✓ την ποιότητα της διαδικασίας της εκκόκκισης.
- ✓ την επίδραση της κάθε χρονιάς

Στη παρούσα εργασία θα εξεταστεί η επίδραση της ποικιλίας και των καιρικών συνθηκών της περιοχής, στα ποιοτικά χαρακτηριστικά της ίνας του βαμβακιού, με τη χρήση των στατιστικών εργαλείων της γλώσσας προγραμματισμού R 2.14.2.

1.2 Διάρθρωση και Ανάλυση των Περιεχομένων της Διατριβής

Η παρούσα διπλωματική εργασία δομείται σε πέντε κεφάλαια. Στο πρώτο κεφάλαιο περιγράφεται συνοπτικά το αντικείμενο και οι στόχοι της εργασίας και αναλύεται η δομή της. Στο δεύτερο κεφάλαιο παρουσιάζονται τα δεδομένα του πειράματος και γίνεται η αναλυτική περιγραφή τους. Στο τρίτο κεφάλαιο πραγματοποιείται η ανάλυση κοινών πειραματικών σχεδίων (CRD, CBD, ANCOVA) από τα δεδομένα της ποιότητας της ίνας του βαμβακιού. Στο τέταρτο κεφάλαιο γίνεται η ανάλυση όλων των δεδομένων της εργασίας και τα αναλυτικά συμπεράσματα μέσα από γραφήματα. Τέλος στο πέμπτο κεφάλαιο παρουσιάζεται η συγκεντρωτική συζήτηση για τα συμπεράσματα και τα δεδομένα της εργασίας. Στο τελευταίο μέρος της εργασίας παρατίθεται η βιβλιογραφία που χρησιμοποιήθηκε για τη διεκπεραίωση της παρούσας εργασίας. Για τη στατιστική ανάλυση των πειραμάτων χρησιμοποιήθηκε η γλώσσα προγραμματισμού R 2.14.2 (the R foundation for statistical computing V1).

ΚΕΦΑΛΑΙΟ 2: ΛΕΠΤΟΜΕΡΙΕΣ ΓΙΑ ΤΗΝ ΙΝΑ ΤΟΥ ΒΑΜΒΑΚΙΟΥ & ΠΑΡΟΥΣΙΑΣΗ ΔΕΔΟΜΕΝΩΝ ΠΕΙΡΑΜΑΤΟΣ

2.1 Φύση του βαμβακιού και βάση της ταξινόμησής του

Το βαμβάκι είναι μια φυσική ίνα που προέρχεται από τον λοβό σπόρων του φυτού βαμβακιού. Η ίνα του βαμβακιού αποτελείται από έναν μοναδικό συνδυασμό κυτταρίνης, κεριών, πρωτεϊνών και άλλων ουσιών που δίνουν στην ίνα τα μοναδικά χαρακτηριστικά της (Cianci, 2009).

Η βάση της ταξινόμησης του βαμβακιού εξαρτάται από το μήκος, τη λεπτότητα και την αντοχή της ίνας του. Το Long-staple βαμβάκι ταξινομείται με βάση το μήκος του, το οποίο είναι συνήθως πάνω από μιάμιση ίντσα (3.81 cm) (Partridge, 2011).

Η λεπτότητα καθορίζεται από τον αριθμό των γραμμαρίων ανά χίλιες γιάρδες (914.4m) νήματος. Η αντοχή καθορίζεται από την ποσότητα συστροφής ή αντοχής του νήματος. Όσο πιο μακρύ και λεπτό είναι το βαμβάκι, τόσο πιο δυνατό και απαλό είναι. Το βαμβάκι μπορεί επίσης να ταξινομηθεί με βάση το χρώμα του, συμπεριλαμβανομένου του λευκού, του ανοιχτού γκρι, του κίτρινου και του καφέ (Askin, 2012). Ο τύπος του βαμβακιού που χρησιμοποιείται σε ένα ύφασμα θα επηρεάσει την απαλότητα, τη δύναμη και την αντοχή του.

Η ταξινόμηση του βαμβακιού είναι η τέχνη και η επιστήμη της περιγραφής της ποιότητας του βαμβακιού ως προς την ποιότητα, το μήκος των ινών, το micronaire και άλλες ιδιότητες ινών σύμφωνα με τα καθιερωμένα πρότυπα βαμβακιού. (S.Gordon & L.Hsieh, 2006)

2.2 Φυσιολογία της ανάπτυξης των ινών

Η ανάπτυξη των ινών είναι μια πολύπλοκη φυσιολογική διαδικασία που περιλαμβάνει πολλαπλές διεργασίες όπως η κυτταρική διαίρεση, η κυτταρική επέκταση, η συναρμολόγηση και εναπόθεση του κυτταρικού τοιχώματος και ο σχηματισμός δευτερογενούς κυτταρικού τοιχώματος (Ladisch and Lippman, 2007).

Η κυτταρίνη και η λιγνίνη είναι τα κύρια συστατικά μιας ίνας και εναποτίθενται στο κυτταρικό τοίχωμα με συγκεκριμένη σειρά. Η κυτταρίνη παράγεται από εξειδικευμένα κύτταρα που ονομάζονται στοιχεία τραχείας ενώ η εναπόθεση λιγνίνης είναι απαραίτητη για τη μηχανική αντοχή του κυτταρικού τοιχώματος (Bao et al., 2009).

Κατά τη διαδικασία ανάπτυξης των ινών, τα κύτταρα υφίστανται μια διαδικασία σχηματισμού δευτερογενούς τοιχώματος που περιλαμβάνει εναπόθεση λιγνίνης και άλλων συστατικών του δευτερογενούς τοιχώματος (Dong et al.,

2010). Οι φυτικές ορμόνες, όπως η αυξίνη και η κυτοκίνη, παίζουν ζωτικό ρόλο στη ρύθμιση της ανάπτυξης των ινών (Liu et al., 2011). Επιπλέον, το αβιοτικό και βιοτικό στρες μπορεί να επηρεάσει την ανάπτυξη των ινών αλλάζοντας τη δομή του κυτταρικού τοιχώματος, την εναπόθεση λιγνίνης και την έκφραση των γονιδίων που εμπλέκονται στην ανάπτυξη των ινών (Feng et al., 2012).

Λίγες μελέτες έχουν ασχοληθεί με την επίδραση της θερμοκρασίας στα ποιοτικά χαρακτηριστικά των ινών βαμβακιού.

Έχουν γίνει προσπάθειες για την ανάπτυξη μοντέλων για την πρόβλεψη της ποιότητας των ινών σε σχέση με την επίδραση της θερμοκρασίας. Το μήκος της ίνας, το micronaire και η ομοιομορφία των ινών σχετίζονται τετραγωνικά με τη θερμοκρασία, ενώ η αντοχή της ίνας αυξάνεται γραμμικά με την αύξηση της θερμοκρασίας. Η αντοχή της ίνας και το micronaire σχετίζονται κυρίως με το παχυντικό δευτερεύον τοίχωμά της, που επηρεάζεται από την υψηλή θερμοκρασία ανάπτυξης (Ramirez et al., 2009).

Επομένως, οι ίνες που παράγονται σε συνθήκες υψηλής θερμοκρασίας έχουν μεγαλύτερη αντοχή λόγω της αύξησης της πάχυνσης του δευτερεύοντος τοιχώματος.

2.3 Εξέλιξη στον έλεγχο της ίνας

Η σημασία του ελέγχου των ινών έχει επεκταθεί πάρα πολύ τα τελευταία χρόνια. Κάποτε η τυχαία δοκιμή πολλών δεμάτων βαμβακιού ήταν επαρκής για τις εργασίες του κλωστήριου. Ο νέος εξοπλισμός επεξεργασίας και κλώσης υψηλής ταχύτητας οδήγησε στην ανάγκη για πρόσθετες πληροφορίες για τις ίνες για κάθε μεμονωμένο δέμα βαμβακιού. Για να επιτευχθεί αυτό το έργο, τα παραδοσιακά όργανα όπως ο ινογράφος, το στερεόμετρο έχουν αντικατασταθεί από όργανο υψηλής έντασης (HVI) (Fischer, 2006). Η υψηλή ταχύτητα δοκιμής και η επαναληψιμότητα του HVI επιτρέπει στο κλωστήριο να δοκιμάσει κάθε δέμα βαμβακιού. Ένα κλωστήριο που χρησιμοποιεί πληροφορίες HVI μπορεί τώρα να προσαρμόσει τα προϊόντα του με βάση το 50% έως 70% του συνολικού κόστους του νήματος παραγωγής (Chou, 2010). Η αποτελεσματική διαχείριση του βαμβακιού είναι απαραίτητη για την κάλυψη των σημερινών αναγκών σε μια παγκόσμια αγορά.

2.4 Το ενδιαφέρον στα ποιοτικά χαρακτηριστικά της ίνας

2.4.1 Τα 3 στάδια παραγωγής και επεξεργασίας βαμβακιού-ίνας

Όπως είναι γνωστό, τα ποιοτικά χαρακτηριστικά είναι αλληλεπίδραση παραγόντων όπως η ποικιλία του βαμβακιού, το περιβάλλον που αναπτύσσεται η καλλιέργεια (εισροές μαζί με καλλιεργητικές τεχνικές) καθώς και οι διεργασίες που λαμβάνουν χώρα από τη φάση της συγκομιδής μέχρι και το τελικό στάδιο της νηματοποίησης. Έτσι η αμερόληπτη εκτίμηση της ποιότητας του βαμβακιού δεν είναι μια απλή υπόθεση.

Τα ποιοτικά χαρακτηριστικά του βαμβακιού έχουν διαφορετική σημασία στα 3 διαφορετικά στάδια παραγωγής και επεξεργασίας του. Για το κάθε στάδιο, ο ενδιαφερόμενος (παραγωγός-εκκοκκιστήριο-κλωστήριο), δίνει αλλού το βάρος της έννοιας της ποιότητας.

2.4.2 Ο παραγωγός- Το εκκοκκιστήριο- Το κλωστήριο

Τα ποιοτικά χαρακτηριστικά που ενδιαφέρουν τον παραγωγό και αφορούν τα στάδια πριν τη συγκομιδή (κατά τη παραγωγή του προϊόντος) είναι τα εξής:

- ✓ η καλή φυτρωτική ικανότητα του σπόρου,
- ✓ η αντοχή του φυτού σε εχθρούς (σκουλήκι, λύγκο κλπ)
- ✓ η ανοχή του σε λιγότερο πότισμα
- ✓ η συγκομιδή να γίνεται μία φορά (σε ένα χέρι).
- ✓ το τελικό αποτέλεσμα κατά τη συγκομιδή
- ✓ η απόδοση της παραγωγής (kg/στρέμμα)

Το ενδιαφέρον του εκκοκκιστηρίου χωρίζεται σε δύο επίπεδα.

1. Το πρώτο επίπεδο συνδέεται με τη λειτουργία του εκκοκκιστηρίου στη φάση του εκκοκκισμού. Σε αυτό το επίπεδο σημαντικό ρόλο παίζει:
 - ✓ η απόδοση του συσπόρου στο εκκοκκιστήριο. Η απόδοση αποτελεί ένα συνονθύλευμα της υγρασίας που έχει το σύσπορο βαμβάκι κατά τη συγκομιδή και παράδοσή του και της ποιότητας του γενετικού υλικού.
2. Το δεύτερο επίπεδο αφορά τη πώληση των δεμάτων που θα παραχθούν η οποία ως επί τω πλείστον γίνεται σε κλωστήρια.
 - ✓ Άρα τα ποιοτικά χαρακτηριστικά θα είναι κοινά με αυτά των κλωστηρίων και αναφέρονται παρακάτω.

Για τα κλωστήρια (και όπως είδαμε και για τα εκκοκκιστήρια) το ενδιαφέρον επικεντρώνεται:

- ✓ στην αντοχή της ίνας (strength),
- ✓ στη λεπτότητα της ίνας (length),
- ✓ στο micronaire (λεπτότητα και ωριμότητα ινών),
- ✓ στην ομοιομορφία (uniformity),
- ✓ στα neps (κόμποι).
- ✓ στην υγρασία της ίνας (moisture)
- ✓ και στο χρωματισμό της ίνας

Προκειμένου να αξιολογηθεί η ποιότητα του βαμβακιού με συνεπή και αμερόληπτο τρόπο, έχουν αναπτυχθεί μια σειρά από τυπικές μεθόδους. Τα πιο συνηθισμένα είναι το Όργανο Υψηλού Όγκου (HVI) (που αναφέρεται και

παραπάνω), το Σύστημα Πληροφοριών Προηγμένων Ινών (AFIS) και το Σύστημα Προηγμένης Απόδοσης (AYS) (Auerbach, 2016). Το HVI είναι ένα εργαλείο που χρησιμοποιείται για τη μέτρηση των κύριων χαρακτηριστικών του βαμβακιού, όπως το μήκος, τη δύναμη, το χρώμα και τη λεπτότητα της ίνας. Το AFIS είναι ένα σύστημα που μετρά τις ιδιότητες των ινών και δημιουργεί γραφήματα που δείχνουν τις ιδιότητες των ινών. Τέλος, το AYS είναι ένα σύστημα που μετρά την απόδοση και την ποιότητα του βαμβακόσπορου.

Η χρήση αυτών των τυποποιημένων μεθόδων επέτρεψε την ακριβέστερη αξιολόγηση της ποιότητας του βαμβακιού, δίνοντας έτσι τη δυνατότητα στους παραγωγούς, στα εκκοκκιστήρια και στα νηματουργεία να λαμβάνουν τεκμηριωμένες αποφάσεις σχετικά με την ποιότητα του βαμβακιού που αγοράζουν ή παράγουν.

2.4.3 Παράγοντες που επηρεάζουν την ανάπτυξη της ίνας

Η ποικιλία του βαμβακιού -το γενετικό υλικό με άλλα λόγια- είναι ο πιο σημαντικός παράγοντας μαζί με την αναπτυξιακή λειτουργία της ίνας στο φυτό που προκαθορίζει τα ποιοτικά χαρακτηριστικά της ίνας, πριν το άνοιγμα του καρυδιού (Kumar et al., 2017). Το μήκος, η αντοχή και η λεπτότητα και ωριμότητα της ίνας εξαρτώνται από τα χαρακτηριστικά κάθε ποικιλίας. Ωστόσο, η ωριμότητα και η αντοχή της ίνας είναι μεταξύ των χαρακτηριστικών της ίνας, τα οποία καθορίζονται από την ποικιλία αλλά μπορούν να επηρεαστούν από τις περιβαλλοντικές συνθήκες σε ένα συγκεκριμένο στάδιο ανάπτυξης. Η απόδοση του εκκοκκισμού μαζί με τις παραμέτρους ποιότητας των ινών (μήκος, αντοχή, λεπτότητα, μικρού μεγέθους) επηρεάζονται από τη θερμοκρασία (Khan et al., 2019).

2.4.4 Επίδραση του περιβάλλοντος στα ποιοτικά χαρακτηριστικά της ίνας

Υπάρχουν συγκεκριμένες συνθήκες σε ορισμένα στάδια ανάπτυξης φυτών βαμβακιού που μπορεί να έχουν δυσμενείς επιπτώσεις στην ποιότητα των ινών.

- Οι συνθήκες υψηλής θερμοκρασίας ή η έλλειψη θρεπτικών ουσιών κατά την περίοδο που αναπτύσσεται το μήκος της ίνας, μπορεί να έχει αρνητική επίδραση στα ποιοτικά χαρακτηριστικά της (Brieger, 2018; Gupta et al., 2018).
- Ο βροχερός καιρός κατά τη διάρκεια της περιόδου ανθοφορίας του βαμβακιού μπορεί να οδηγήσει σε ατελή λίπανση του φυτού, με άμεση συνέπεια στη παραγωγή των ινών (Nanda et al., 2018).
- Η πρόωρη αποφύλλωση των φυτών επιδρά στη μείωση της τιμής του micronaire. Αυτή μπορεί να είναι συνέπεια της παύσης της άρδευσης σε περίοδο που ακόμη το φυτό χρειάζεται νερό, ή συνέπεια έντονων εντομολογικών προσβολών.

Στη συνέχεια παρουσιάζονται τα δεδομένα του πειράματος καθώς και τα ποιοτικά χαρακτηριστικά που θα αναλυθούν μέσα από στατιστικές αναλύσεις.

2.5 Τα δεδομένα του πειράματος

Η παρούσα μελέτη αξιοποίησε δεδομένα που συλλέχθηκαν από εκκοκκιστήριο στην περιοχή των Φαρσάλων Λάρισας για μια πενταετία (2017-2021). Για συλλογή δεδομένων επιλέχθηκαν τρεις μεγάλες βαμβακοπαραγωγικές περιοχές των Φαρσάλων Λάρισας, των Τρικάλων και της Λαμίας. Όλα τα χωράφια που χρησιμοποιήθηκαν για την παραγωγή βαμβακιού ποτίστηκαν με στάγδην άρδευση και χρησιμοποιήθηκαν τυπικές τεχνικές καλλιέργειας και προστασίας. Μετά τη συλλογή του βαμβακιού, το βαμβάκι που συγκομίστηκε παραδόθηκε στις αποθήκες εκκοκκιστηρίων και εκκοκκίστηκε ξεχωριστά για να διασφαλιστεί η σωστή επιλογή δείγματος για ανάλυση ινών. Επιλέχθηκαν τρεις πρωτογενείς ποικιλίες βαμβακιού με βάση τις πληροφορίες που παρείχε ο ΟΠΕΚΕΠΕ και συλλέχθηκαν τυχαία 180 πειραματικές μονάδες τριών ποιοτικών χαρακτηριστικών ινών βαμβακιού (micronaire, μήκος, ανθεκτικότητα) από κάθε περιοχή.

Επιπλέον, συλλέχθηκαν μετεωρολογικά στοιχεία για τον μήνα Αύγουστο, που είναι περίοδος ενδιαφέροντος για την παραγωγή βαμβακιού, για καθένα από τα πέντε χρόνια και για κάθε περιοχή.

Τα ποιοτικά χαρακτηριστικά των ινών που διερευνήθηκαν περιελάμβαναν αντοχή, μήκος και λεπτότητα/ωριμότητα (micronaire).

Πραγματοποιήθηκαν αναλύσεις για καθένα από τα τρία χαρακτηριστικά σε εξειδικευμένα εργαστήρια κλώσης στην περιοχή των Φαρσάλων με τη χρήση πιστοποιημένων οργάνων μέτρησης ινών.

2.6 Ποιοτικά χαρακτηριστικά ίνας

Στην ενότητα αυτή παρουσιάζονται τα τρία κυριότερα ποιοτικά χαρακτηριστικά ίνας που θα εξετάσουμε παρακάτω.

2.6.1 Αντοχή της ίνας (strength)

Η αντοχή των ινών είναι πρωτίστως ένα γενετικό χαρακτηριστικό κάθε ποικιλίας, ωστόσο επηρεάζεται από τις εδαφικές συνθήκες, το κλίμα, τη διαχείριση των καλλιεργειών, την επαρκή υγρασία του εδάφους και τη διατροφή των φυτών. Το κάλιο έχει θετική επίδραση στο μήκος των ινών, με αποτέλεσμα πιο δυνατά και λεπτότερα νήματα (Babu et al., 2018). Άλλες φυσικές και χημικές ιδιότητες όπως η λεπτότητα της ίνας, η ωριμότητα, η ευκαμψία της και τα χαρακτηριστικά της επιφάνειας της ίνας επηρεάζουν την αντοχή της ίνας (Kumar et al., 2019).

2.6.2 Λεπτότητα και ωριμότητα της ίνας (micronaire).

Η λεπτότητα και η ωριμότητα, αναφέρονται στη διάμετρο της ίνας, την αντοχή του νήματος, τη στιλπνότητα, την απορρόφηση της βαφής και το πάχος του δευτερεύοντος τοιχώματος, αντίστοιχα (Girish et al., 2018; Lu et al., 2018). Η λεπτότητα είναι κυρίως ένα γενετικό χαρακτηριστικό κάθε ποικιλίας, ενώ η ωριμότητα επηρεάζεται από το περιβάλλον ανάπτυξης, τόσο από τον καιρό όσο και η διαχείριση. Το χαμηλό micronaire συχνά υποδηλώνει είτε λεπτές είτε ανώριμες ίνες, ενώ το υψηλό micronaire σχετίζεται με υψηλές θερμοκρασίες ή υδατικό στρες (Lu et al., 2018).

2.6.3 Μήκος ίνας (length)

Το μήκος καθορίζει την ελαστικότητα της ίνας και είναι εξίσου γενετικό γνώρισμα. Υψηλές τιμές αυτού του χαρακτηριστικού έχουν θετική επίδραση στη διαδικασία της κλώσης.

Το μήκος εξαρτάται σε μεγάλο βαθμό από την ποικιλία, αλλά η θερμοκρασία, το στρες στο νερό, η ανεπάρκεια καλίου και η υπερβολική θραύση των ινών στο εκκοκκιστήριο μπορούν επίσης να μειώσουν το μήκος των ινών.

2.6.4 Τιμές ποιοτικών χαρακτηριστικών ίνας

Οι επιθυμητές τιμές των ποιοτικών χαρακτηριστικών της ίνας παρουσιάζονται στον παρακάτω πίνακα:

Πίνακας 2-1: Όρια τιμών ποιοτικών χαρακτηριστικών ίνας

ΠΟΙΟΤΙΚΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ	ΕΠΙΘΥΜΗΤΕΣ ΤΙΜΕΣ
ΜΗΚΟΣ ΙΝΑΣ	>28.5mm
ΑΝΤΟΧΗ ΙΝΑΣ	>29gr/tex
Micronaire (Λεπτότητα-Ωριμότητα)	3.8-4.5

Αυτό το σύνολο ποιοτικών χαρακτηριστικών είναι επιθυμητό επειδή υποδηλώνει ότι το μήκος της ίνας είναι αρκετά μεγάλο για πολλές εφαρμογές, η αντοχή της ίνας είναι αρκετά ισχυρή για πολλές χρήσεις και το Micronaire βρίσκεται εντός του ιδανικού εύρους. Οι τιμές θα πρέπει να είναι λογικές για αυτά τα ποιοτικά χαρακτηριστικά, καθώς είναι επιθυμητές για πολλές εφαρμογές.

ΚΕΦΑΛΑΙΟ 3: ΑΝΑΛΥΣΗ ΚΟΙΝΩΝ ΠΕΙΡΑΜΑΤΙΚΩΝ ΣΧΕΔΙΩΝ ΜΕ ΠΑΡΑΔΕΙΓΜΑΤΑ

Για τα δεδομένα που συλλέγονται σε αυτήν την αναδρομική μελέτη, θα πρέπει να προσαρμοστεί μια κατάλληλη διαδικασία αξιολόγησης. Αυτό μπορεί να περιλαμβάνει βήματα όπως η επαλήθευση δεδομένων, η στατιστική ανάλυση και άλλες μέθοδοι ανάλυσης δεδομένων, ανάλογα με τα δεδομένα και τους στόχους της μελέτης.

3.1 Ανάλυση δεδομένων με Πλήρως Τυχαιοποιημένο Σχέδιο (Completely Randomized Design)

3.1.1 Το Πλήρως τυχαιοποιημένο σχέδιο (CRD)

Το πλήρως τυχαιοποιημένο σχέδιο (Completely Randomized Design) (Holland, 2020) είναι ένας πειραματικός σχεδιασμός όπου οι μεταχειρίσεις κατανέμονται τυχαία σε πειραματικές μονάδες. Αυτός ο σχεδιασμός χρησιμοποιείται συχνά σε γεωργικά πειράματα (Muskat, 2019) και μπορεί να χρησιμοποιηθεί για τη σύγκριση των επιδράσεων διαφορετικών μεταχειρίσεων στο ίδιο σύνολο πειραματικών μονάδων. Η τυχαιοποίηση των μεταχειρίσεων στο CRD επιτρέπει στον ερευνητή να μετρήσει τα αποτελέσματα των μεταχειρίσεων στις πειραματικές μονάδες και να τις συγκρίνει μεταξύ τους για να διασφαλίσει ότι τα αποτελέσματα είναι ανεξάρτητα από οποιαδήποτε άλλη μεταβλητή (Chambless et al., 2018).

Το επίπεδο σημαντικότητας είναι ένα όριο πιθανότητας που χρησιμοποιείται για τον προσδιορισμό της στατιστικής σημασίας ενός αποτελέσματος. Ένα επίπεδο σημαντικότητας 0.05 (ή 5%) χρησιμοποιείται συνήθως στην επιστημονική έρευνα και την ανάλυση δεδομένων.

Επιλέγεται επίπεδο σημαντικότητας πριν τη συλλογή δεδομένων στο 0.05 (5%).

3.1.2 Για το ποιοτικό χαρακτηριστικό: Μήκος ίνας Τα δεδομένα

Παρακάτω θα γίνει η ανάλυση των δεδομένων με τη μέθοδο του πλήρους τυχαιοποιημένου σχεδίου.

Για την ανάλυση δεδομένων με πλήρως τυχαιοποιημένο σχέδιο επιλέχθηκαν για να εξεταστούν τρεις ποικιλίες βαμβακιού. Χρησιμοποιήθηκαν 60 διαφορετικά δέματα για κάθε ποικιλία για τις χρονιές 2017-2021 και μετρήθηκε το μήκος της ίνας (Length) . Ως πειραματική μονάδα ορίστηκε το δέμα.

Ένας εντελώς τυχαιοποιημένος σχεδιασμός είναι η κατάλληλη μέθοδος για την ανάλυση των δεδομένων σε αυτή την περίπτωση. Χρησιμοποιώντας 60 διαφορετικά δέματα για κάθε ποικιλία κατά τη διάρκεια 5 ετών, το πείραμα μπορεί να παρέχει χρήσιμα δεδομένα σχετικά με τις διαφορές στο μήκος των ινών μεταξύ των διαφορετικών ποικιλιών.

Τυχαιοποίηση

Η τυχαιοποίηση είναι ένα σημαντικό μέρος του πειραματικού σχεδιασμού, καθώς βοηθά να διασφαλιστεί ότι τυχόν παρατηρούμενα αποτελέσματα μπορούν να αποδοθούν στις πειραματικές μεταχειρίσεις και όχι σε άλλους παράγοντες.

Διαδικασία τυχαιοποίησης με κλήρωση

α) αρχικά καθορίζεται ο αριθμός N των πειραματικών μονάδων, στη περίπτωση μας ο αριθμός είναι 180. Οι επεμβάσεις (ποικιλίες) είναι 3.

β) γίνεται απαρίθμηση των πειραματικών μονάδων

γ) διεξάγεται κλήρωση για τη τυχαιοποίηση των επεμβάσεων στις πειραματικές μονάδες. Κάθε πειραματική μονάδα έχει ίση πιθανότητα να επιλεγεί για κάποιο επίπεδο ποικιλίας.

δ) Έτσι, μετά τη διεξαγωγή της κλήρωσης έχουμε την εξής διαμόρφωση του πίνακα:

Πίνακας 3-1: Δεδομένα πριν την τυχαιοποίηση CRD

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54
55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72
73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90
91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108
109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126
127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144
145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162
163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180

Πίνακας 3-2: Δεδομένα μετά την τυχαιοποίηση CRD

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54
55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72
73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90
91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108
109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126
127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144
145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162
163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180

Όπου το μπλε χρώμα αντιπροσωπεύει την ποικιλία Αρμονία, το καφέ τη ποικιλία Elpida και το μωβ τη ποικιλία PRG 9811.

Η ανάλυση (Ο κώδικας)

```
attach(crd_len)
```

```
crd_len
```

```
Variety=factor(Variety);pack=factor(pack)
```

Η ANOVA (ANalysis Of VAriance) είναι μια από τις μεθόδους πειραματικών σχεδιασμών για να προσδιοριστεί εάν δύο ή περισσότεροι μέσοι όροι πληθυσμού είναι διαφορετικοί. Με άλλα λόγια, χρησιμοποιείται για τη σύγκριση δύο ή περισσότερων επιπέδων για να δούμε αν είναι σημαντικά διαφορετικά.

- Μηδενική υπόθεση: οι μέσες τιμές των μετρήσεων δεν διαφέρουν μεταξύ των ποικιλιών

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_p$$

- Εναλλακτικά: Τουλάχιστον, ο μέσος όρος του μήκους ίνας μιας ποικιλίας είναι διαφορετικός από άλλες ποικιλίες μετά τον έλεγχο της επίδρασης, δηλαδή οι προσαρμοσμένες τιμές δεν είναι ίσες

$$H_1 : \text{Τουλάχιστον δύο επίπεδα διαφέρουν στατιστικά σημαντικά.}$$

Με την εντολή `aov` γίνεται η προσαρμογή του μοντέλου ANOVA και με την εντολή `summary` παρουσιάζονται τα αποτελέσματα.

```
fit<-aov(LEN~Variety)
```

```
summary(fit)
```

```
> fit=aov(LEN~Variety)
> summary(fit)
              Df Sum Sq Mean Sq F value Pr(>F)
Variety         2  296.57   148.28   281.8 <2e-16 ***
Residuals     177   93.15     0.53
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Υπάρχει σημαντική διαφορά στα μέσα μήκη των ποικιλιών, με πολύ χαμηλή τιμή p ($< 2e-16$). Αυτό υποδηλώνει ότι τουλάχιστον μία από τις ποικιλίες έχει σημαντικά διαφορετικό μέσο μήκος ίνας από τις άλλες.

Για να το ελέγξουμε αυτό, πρέπει να χρησιμοποιήσουμε άλλου τύπου τεστ, που αναφέρονται ως τεστ post-hoc (στα λατινικά, «εκ των υστέρων δοκιμές», άρα μετά τη λήψη στατιστικά σημαντικών αποτελεσμάτων ANOVA) ή πολλαπλές δοκιμές σύγκρισης κατά ζεύγη, για να προσδιοριστεί ποιες συγκεκριμένες μέσες τιμές είναι στατιστικά διαφορετικές μεταξύ τους.

Σε αυτή τη περίπτωση θα χρησιμοποιήσουμε το Tukey's HSD (Honest Significant Difference).

Το τεστ HSD του Tukey είναι ένα ευρέως χρησιμοποιούμενο τεστ post-hoc που είναι κατάλληλο να χρησιμοποιηθεί κατά τη σύγκριση πολλαπλών μέσων.

TukeyHSD(fit)

```
> TukeyHSD(fit)
Tukey multiple comparisons of means
 95% family-wise confidence level

Fit: aov(formula = LEN ~ Variety)

$Variety
      diff      lwr      upr    p adj
ELPIDA-ARMONIA -0.6965 -1.009555 -0.3834452 1.2e-06
PRG 9811-ARMONIA -3.0035 -3.316555 -2.6904452 0.0e+00
PRG 9811-ELPIDA -2.3070 -2.620055 -1.9939452 0.0e+00
```

Τα αποτελέσματα της δοκιμής TukeyHSD δείχνουν ότι η διαφορά μεταξύ των μέσων των ποικιλιών ELPIDA και ARMONIA είναι σημαντικά διαφορετική (p -value = $1.2e-06$). Οι ποικιλίες PRG 9811 και ARMONIA έχουν επίσης σημαντικά διαφορετικό μέσο όρο (p -value = 0). Τέλος, οι ποικιλίες PRG 9811 και ELPIDA έχουν σημαντικά διαφορετικό μέσο όρο (p -value = 0).

Οι αναλύσεις Tukey's HSD και το γράφημα μπορεί να γίνει και μέσω του R πακέτου emmeans.

```
emm1=emmeans::emmeans(fit, specs = "Variety")
```

emm1

```
> emm1=emmeans::emmeans(fit, specs = "Variety")
> emm1
Variety emmean   SE df lower.CL upper.CL
ARMONIA  30.9 0.0937 177   30.7   31.1
ELPIDA   30.2 0.0937 177   30.0   30.4
PRG 9811  27.9 0.0937 177   27.7   28.1

confidence level used: 0.95
```

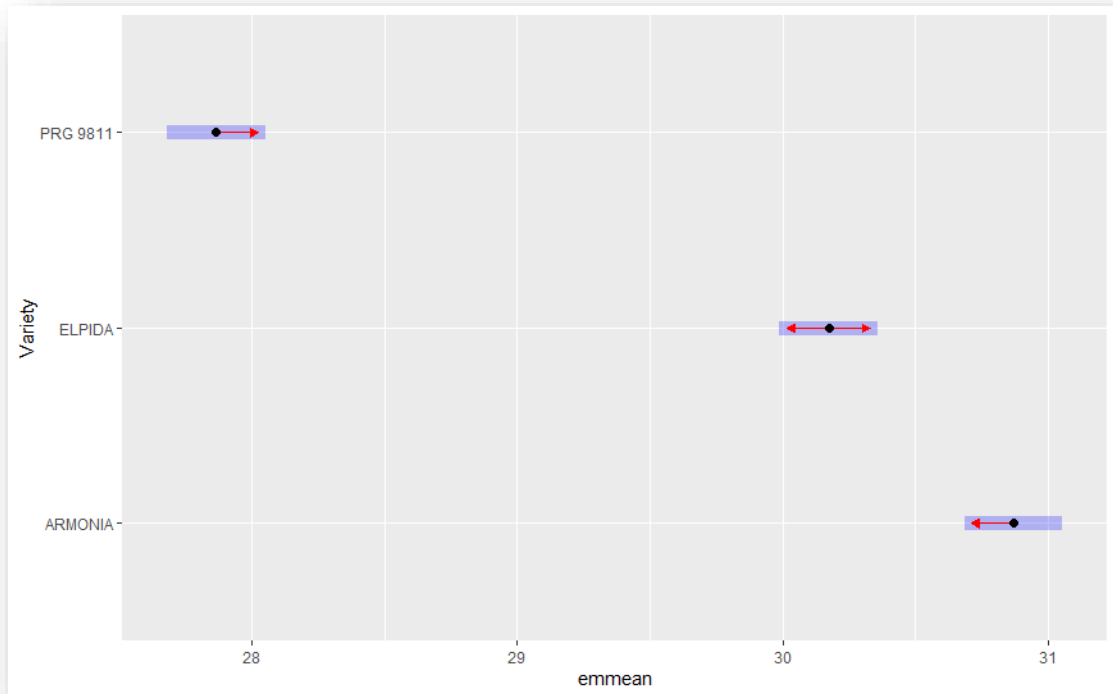
Το μέσο μήκος ίνας της ποικιλίας ARMONIA είναι 30.9 ± 0.0937 , το μέσο μήκος της ποικιλίας ELPIDA είναι 30.2 ± 0.0937 και το μέσο μήκος της ποικιλίας PRG 9811 είναι 27.9 ± 0.0937 . Παρέχονται επίσης τα διαστήματα εμπιστοσύνης 95% για κάθε μέσο όρο.

Αυτό σημαίνει ότι η ποικιλία ARMONIA έχει το μεγαλύτερο μέσο μήκος και ακολουθούν οι ποικιλίες ELPIDA και PRG 9811. Τα διαστήματα εμπιστοσύνης 95% δείχνουν επίσης ότι η διαφορά μεταξύ των μέσων είναι στατιστικά σημαντική.

Το plot που δημιουργείται από το πακέτο emmeans δείχνει τα μέσα μήκη των τριών ποικιλιών, καθώς και τα διαστήματα εμπιστοσύνης 95%. Οι γραμμές που συνδέουν τα σημεία υποδεικνύουν τη σύγκριση των μέσων ανά ζεύγη. Αυτή η

γραφική παράσταση επιβεβαιώνει περαιτέρω ότι η διαφορά μεταξύ των μέσων είναι στατιστικά σημαντική.

`plot(emm1, comparisons = TRUE)`



Διάγραμμα 3-1: Ανά ζεύγη συγκρίσεις *emmeans*

3.1.3 Για το ποιοτικό χαρακτηριστικό: MICRONAIRE

Τα δεδομένα

Για την ανάλυση δεδομένων με πλήρως τυχαιοποιημένο σχέδιο επιλέχθηκαν για να εξεταστούν τρεις ποικιλίες βαμβακιού. Χρησιμοποιήθηκαν 60 διαφορετικά δέματα για κάθε ποικιλία για τις χρονιές 2017-2021 και μετρήθηκε το micronaire της ίνας. Ως πειραματική μονάδα ορίστηκε το δέμα.

Τυχαιοποίηση

Το κομμάτι της τυχαιοποίησης έγινε ακριβώς με τον ίδιο τρόπο όπως και για το προηγούμενο ποιοτικό χαρακτηριστικό σύμφωνα με τους Πίνακες Πίνακας 3-1 & Πίνακας 3-2.

Η ανάλυση (Ο κώδικας)

```
attach(crd_mic)
```

```
crd_mic
```

```
Variety=factor(Variety);pack=factor(pack)
```

Η ANOVA (ANalysis Of VAriance) είναι μια στατιστική δοκιμή για να προσδιοριστεί εάν δύο ή περισσότεροι μέσοι όροι πληθυσμού είναι διαφορετικοί.

- Μηδενική υπόθεση: οι μέσες τιμές των μετρήσεων δεν διαφέρουν μεταξύ των ποικιλιών

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_p$$

- Εναλλακτικά: Τουλάχιστον, ο μέσος όρος του micronaire μιας ποικιλίας είναι διαφορετικός από άλλες ποικιλίες μετά τον έλεγχο της επίδρασης, δηλαδή οι προσαρμοσμένες τιμές δεν είναι ίσες

$$H_1 : \text{Τουλάχιστον δύο επίπεδα διαφέρουν στατιστικά σημαντικά.}$$

```
fit=aov(MIC~Variety)
```

```
summary(fit)
```

```
> summary(fit)
              Df Sum Sq Mean Sq F value Pr(>F)
Variety        2  3.355   1.6774   81.05 <2e-16 ***
Residuals    177  3.663   0.0207
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Υπάρχει στατιστικά σημαντική διαφορά μεταξύ των τριών ποικιλιών βαμβακιού με βάση τις μετρήσεις του micronaire των ινών τους. Η τιμή $F=81,05$ και η τιμή p μικρότερη από $2e-16$ δείχνουν ότι η διαφορά μεταξύ των ποικιλιών είναι σημαντική.

Το επόμενο βήμα θα είναι η περαιτέρω διερεύνηση σχετικά με το ποιες ειδικές διαφορές μεταξύ των ποικιλιών προκαλούν τις σημαντικές διαφορές στις μετρήσεις micronaire. Αυτό θα μπορούσε να γίνει με τη διεξαγωγή δοκιμών post hoc, όπως η δοκιμή Tukey HSD.

Πραγματοποιούμε Posthoc test.

TukeyHSD(fit)

```
> TukeyHSD(fit)
Tukey multiple comparisons of means
 95% family-wise confidence level

Fit: aov(formula = MIC ~ Variety)

$variety
      diff      lwr      upr p adj
ELPIDA-ARMONIA  0.1435000 0.08142083 0.2055792 5e-07
PRG 9811-ARMONIA 0.3333333 0.27125416 0.3954125 0e+00
PRG 9811-ELPIDA  0.1898333 0.12775416 0.2519125 0e+00
```

Οι τιμές p για τις συγκρίσεις ανά ζεύγη δείχνουν ότι οι διαφορές μεταξύ των ποικιλιών είναι σημαντικές. Συγκεκριμένα, οι διαφορές μεταξύ ELPIDA και ARMONIA και μεταξύ PRG 9811 και ARMONIA είναι και οι δύο πολύ σημαντικές, με τιμές p 5e-07 και 0e+00 αντίστοιχα όπως επίσης και η διαφορά μεταξύ PRG 9811 και ELPIDA, με τιμή p 0e+00.

Πραγματοποιούμε την ανάλυση emmean για να προσδιορίσουμε τον εκτιμώμενο οριακό μέσο όρο για κάθε ποικιλία βαμβακιού και να αξιολογήσουμε τη σημασία των διαφορών μεταξύ των μέσων.

```
emm1=emmeans::emmeans(fit, specs = "Variety")
```

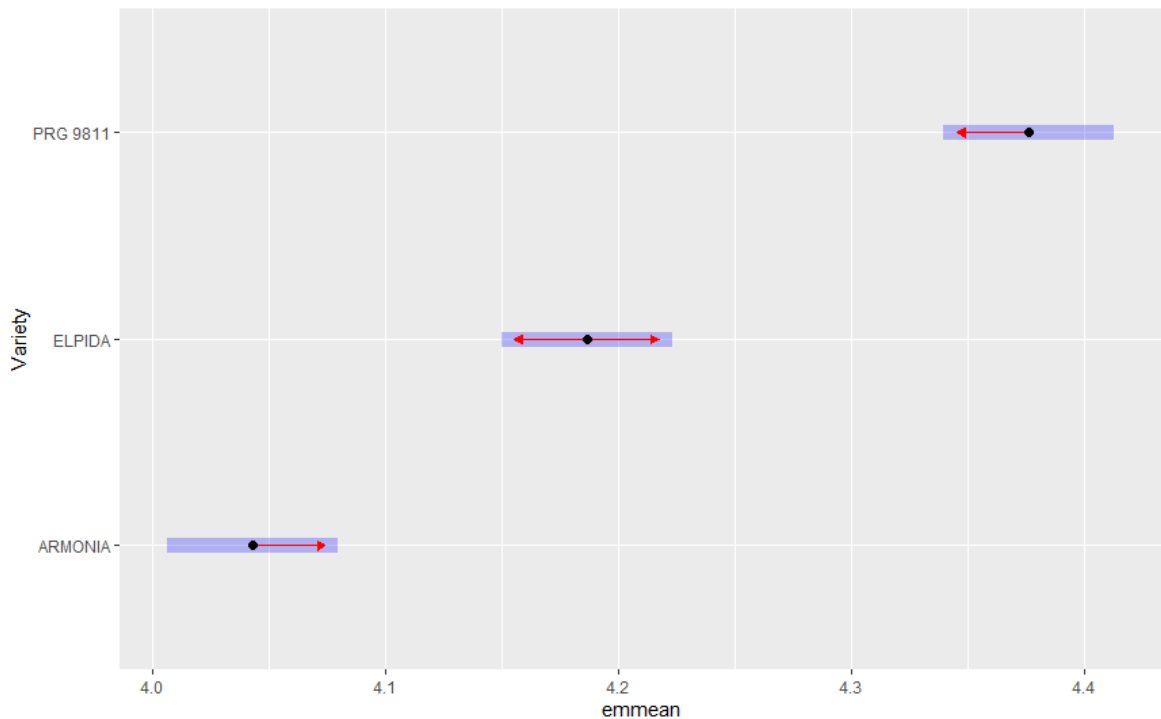
emm1

```
> emm1=emmeans::emmeans(fit, specs = "Variety")
> emm1
Variety emmean   SE df lower.CL upper.CL
ARMONIA  4.04 0.0186 177  4.01  4.08
ELPIDA   4.19 0.0186 177  4.15  4.22
PRG 9811  4.38 0.0186 177  4.34  4.41

confidence level used: 0.95
```

Ο εκτιμώμενος οριακός μέσος όρος για τις τρεις ποικιλίες δείχνει ότι υπάρχει σημαντική διαφορά στις μετρήσεις του micronaire, με την ARMONIA να έχει τον χαμηλότερο μέσο όρο 4.04, την ELPIDA με μέσο όρο 4.19 και το PRG 9811 να έχει τον υψηλότερο μέσο όρο 4.38.

```
plot(emm1, comparisons = TRUE)
```

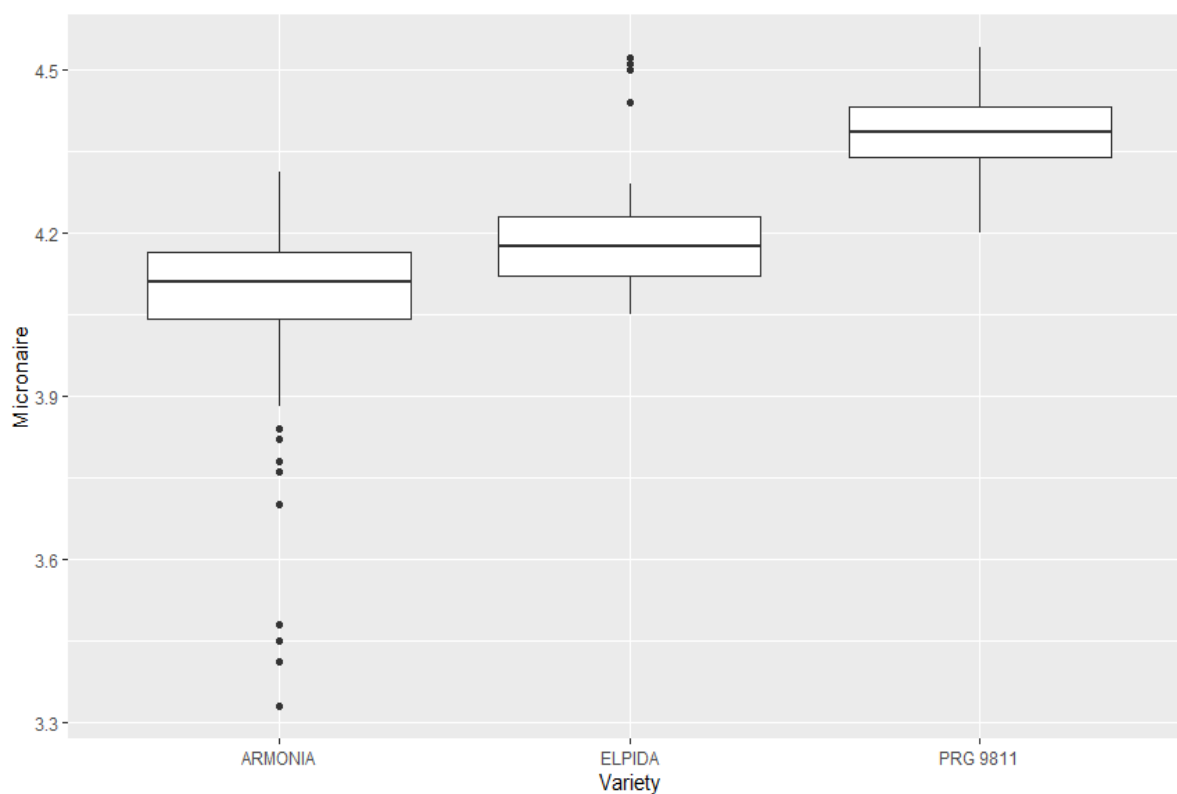


Διάγραμμα 3-2: Ανά ζεύγη συγκρίσεις *emmeans*

Αυτό το διάγραμμα δείχνει τον εκτιμώμενο οριακό μέσο όρο για κάθε ποικιλία βαμβακιού και τα σχετικά διαστήματα εμπιστοσύνης. Συγκεκριμένα, οι διαφορές μεταξύ ELPIDA και ARMONIA και μεταξύ PRG 9811 και ARMONIA είναι και οι δύο πολύ σημαντικές, με τιμές p $5e-07$ και $0e+00$ αντίστοιχα. Σημαντική είναι και η διαφορά μεταξύ PRG 9811 και ELPIDA, με τιμή p $0e+00$.

Μπορούμε να οπτικοποιήσουμε τα δεδομένα δημιουργώντας ένα *boxplot* με τις μετρήσεις *micronaire* για κάθε ποικιλία βαμβακιού.

```
library(ggplot2)
ggplot (fit, aes(x=Variety, y=Micronaire)) +
  geom_boxplot()
```



Διάγραμμα 3-3: Βoxplot κατανομής μετρήσεων micronaire

Αυτό το boxplot δείχνει την κατανομή των μετρήσεων micronaire για κάθε ποικιλία βαμβακιού. Υπάρχει στατιστικά σημαντική διαφορά μεταξύ των ποικιλιών, με την ARMONIA να έχει τη χαμηλότερη μέτρηση micronaire 4.0, την ELPIDA να έχει 4.2 και το PRG 9811 να έχει τον υψηλότερη 4.4.

3.1.4 Για το ποιοτικό χαρακτηριστικό: Αντοχή της ίνας

Τα δεδομένα

Για την ανάλυση δεδομένων με πλήρως τυχαιοποιημένο σχέδιο επιλέχθηκαν για να εξεταστούν τρεις ποικιλίες βαμβακιού. Χρησιμοποιήθηκαν 60 διαφορετικά δέματα για κάθε ποικιλία για τις χρονιές 2017-2021 και μετρήθηκε η αντοχή της ίνας. Ως πειραματική μονάδα ορίστηκε το δέμα.

Τυχαιοποίηση

Το κομμάτι της τυχαιοποίησης έγινε ακριβώς με τον ίδιο τρόπο όπως και για τα προηγούμενα ποιοτικά χαρακτηριστικά.

Η ανάλυση

```
attach(crd_str)
```

```
crd_str
```

```
Variety=factor(Variety);pack=factor(pack)
```

Η ANOVA (ANalysis Of VAriance) είναι μια στατιστική δοκιμή για να προσδιοριστεί εάν δύο ή περισσότεροι μέσοι όροι πληθυσμού είναι διαφορετικοί. Με άλλα λόγια, χρησιμοποιείται για τη σύγκριση δύο ή περισσότερων επιπέδων για να δούμε αν είναι σημαντικά διαφορετικά.

- Μηδενική υπόθεση : οι μέσες τιμές των μετρήσεων δεν διαφέρουν μεταξύ των ποικιλιών

$$H_0: \mu_1 = \mu_2 = \dots = \mu_p$$

- Εναλλακτικά: Τουλάχιστον, ο μέσος όρος της αντοχής της ίνας μιας ποικιλίας είναι διαφορετικός από άλλες ποικιλίες μετά τον έλεγχο της επίδρασης, δηλαδή οι προσαρμοσμένες τιμές δεν είναι ίσες

$$H_1: \text{Τουλάχιστον δύο επίπεδα διαφέρουν στατιστικά σημαντικά.}$$

```
fit=aov(STRENGTH~Variety)
```

```
summary(fit)
```

```
> fit=aov(STRENGTH~Variety)
> summary(fit)
              Df Sum Sq Mean Sq F value    Pr(>F)
Variety         2   21.16   10.579    10.8 3.75e-05 ***
Residuals    177  173.33    0.979
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Υπάρχει στατιστικά σημαντική διαφορά στην αντοχή των ποικιλιών βαμβακιού, με τιμή F 10.8 και τιμή p 3.75e-05. Αυτό υποδηλώνει ότι οι ποικιλίες βαμβακιού έχουν επίδραση στην αντοχή της ίνας.

Κάνουμε μια δοκιμή TukeyHSD για να συγκρίνουμε τις μέσες τιμές κάθε ποικιλίας βαμβακιού προκειμένου να προσδιορίσουμε ποιες ποικιλίες έχουν την υψηλότερη ή τη χαμηλότερη αντοχή.

Πραγματοποιούμε Posthoc test.

TukeyHSD(fit)

```
> TukeyHSD(fit)
Tukey multiple comparisons of means
 95% family-wise confidence level

Fit: aov(formula = STRENGTH ~ Variety)

$variety
      diff      lwr      upr    p adj
ELPIDA-ARMONIA  0.2560000 -0.1710377  0.6830377  0.3343681
PRG 9811-ARMONIA 0.8206667  0.3936290  1.2477043  0.0000305
PRG 9811-ELPIDA  0.5646667  0.1376290  0.9917043  0.0058583
```

Αυτή η ανάλυση TukeyHSD δείχνει ότι υπάρχει μια στατιστικά σημαντική διαφορά μεταξύ της μέσης αντοχής ινών των ποικιλιών βαμβακιού ELPIDA και PRG 9811, με τιμή p 0,0000305. Υπάρχει επίσης στατιστικά σημαντική διαφορά των ποικιλιών PRG 9811 και ARMONIA, με τιμή p 0,0058583.

Είναι σκόπιμο να διεξαχθεί μια δοκιμή emmeans μετά τη δοκιμή TukeyHSD προκειμένου να διερευνηθούν περαιτέρω οι διαφορές μεταξύ των μέσων τιμών των ποικιλιών. Αυτό μπορεί να γίνει με τη διεξαγωγή συγκρίσεων ανά ζεύγη και τον υπολογισμό των διαστημάτων εμπιστοσύνης για τις διαφορές μεταξύ των μέσων.

Διενεργούμε ένα emmeans test.

```
emm1=emmeans::emmeans(fit, specs = "Variety")
```

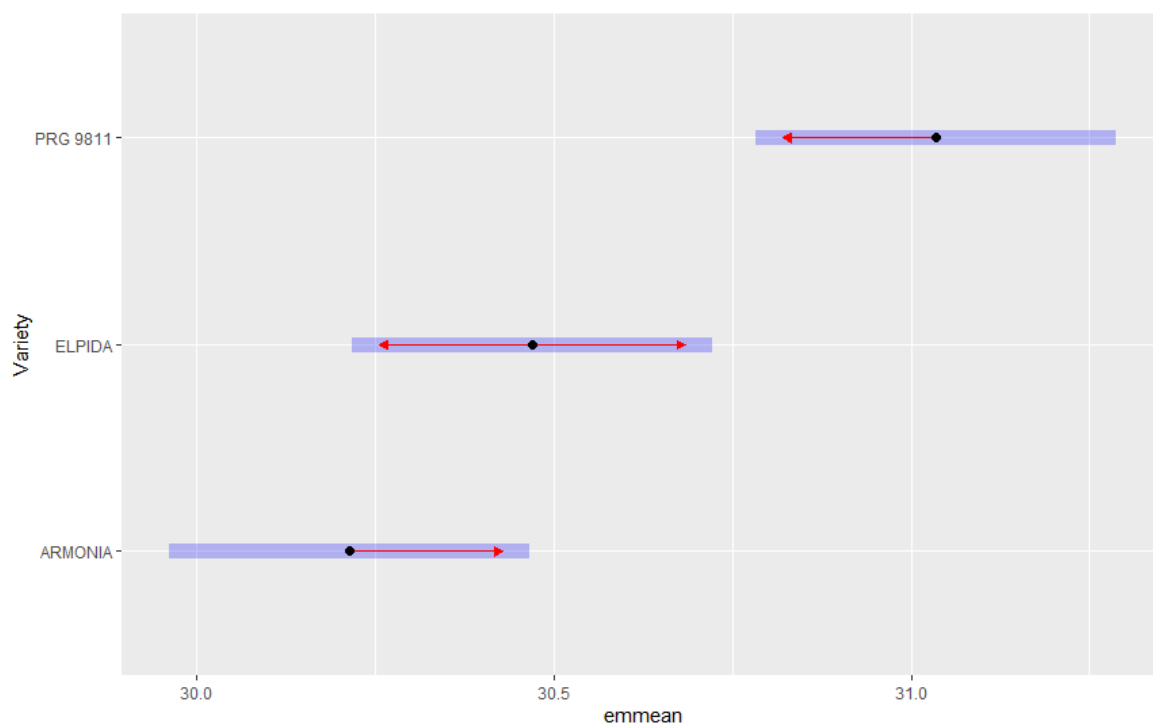
emm1

```
> emm1=emmeans::emmeans(fit, specs = "Variety")
> emm1
Variety emmean  SE  df lower.CL upper.CL
ARMONIA  30.2 0.128 177   30.0   30.5
ELPIDA   30.5 0.128 177   30.2   30.7
PRG 9811  31.0 0.128 177   30.8   31.3

Confidence level used: 0.95
```

Η μέση αντοχή της ίνας για την ARMONIA είναι 30.2, με διάστημα εμπιστοσύνης 95% από 30.0 έως 30.5. Η μέση αντοχή ίνας της ELPIDA είναι 30.5, με διάστημα εμπιστοσύνης 95% από 30.2 έως 30.7. Τέλος, η μέση αντοχή της ίνας του PRG 9811 είναι 31.0, με διάστημα εμπιστοσύνης 95% από 30.8 έως 31.3. Αυτό υποδηλώνει ότι η PRG 9811 έχει την υψηλότερη μέση αντοχή ίνας, ακολουθούμενη από την ELPIDA και μετά την ARMONIA.

```
plot(emm1, comparisons =TRUE)
```



Διάγραμμα 3-4: Ανά ζεύγη συγκρίσεις emmeans

Αυτό το διάγραμμα δείχνει τον εκτιμώμενο μέσο όρο των τριών ποικιλιών βαμβακιού, καθώς και τα διαστήματα εμπιστοσύνης 95%. Η PRG 9811 έχει την υψηλότερη μέση αντοχή, ακολουθούμενη από την ELPIDA και μετά την ARMONIA.

3.2 Ανάλυση δεδομένων με Σχεδιασμός τυχαιοποιημένων ομάδων (Randomized Block Design)

3.2.1 Ο Σχεδιασμός τυχαιοποιημένων ομάδων

Υπάρχουν περιπτώσεις που όταν κάνουμε πειράματα στον αγρό ξέρουμε ότι υπάρχουν διαφορές σε ότι αφορά στη γονιμότητα του αγρού κατά μήκος ή κατά πλάτος, και θέλουμε να απομακρύνουμε τη πηγή αυτής παραλλακτικότητας από την επίδραση των επεμβάσεων. Για το λόγο αυτό στρεφόμαστε στη χρήση του σχεδίου Τυχαιοποιημένων Ομάδων.

Σύμφωνα με αυτό, η περιοχή που πρόκειται να χρησιμοποιηθεί για το πείραμα, χωρίζεται σε λωρίδες ίσες σε αριθμό με τον αριθμό επαναλήψεων του πειράματος. Η κάθε λωρίδα, ή επανάληψη, ή ομάδα, έχει συνήθως τόσα πειραματικά τεμάχια όσος είναι και ο αριθμός των επεμβάσεων. Οι επεμβάσεις (λιπάσματα, ποικιλίες, ψεκασμοί κλπ) μετά τυχαιοποιούνται, ανεξάρτητα σε κάθε επανάληψη.

Σκοπός της τοποθέτησης των επεμβάσεων κατά ομάδες είναι οι συνθήκες των πειραματικών μονάδων στο εσωτερικό κάθε ομάδας να είναι ομοιόμορφες.

Στο πειραματικό σχέδιο Σχεδιασμός πλήρων τυχαιοποιημένων ομάδων, RCBD (Randomized complete block design) έχουμε μια επανάληψη εντός των ομάδων όπως ορίζει η θεωρία. Στο συγκεκριμένο πείραμα όμως τα δεδομένα είναι κάπως διαφορετικά. Καταλήγουμε λοιπόν να εφαρμόσουμε ένα RBD μοντέλο εφόσον έχουμε 3 επαναλήψεις εντός των ομάδων μας.

Επιλέγεται επίπεδο σημαντικότητας πριν τη συλλογή δεδομένων στο 0,05 (5%).

3.2.2 Για το ποιοτικό χαρακτηριστικό: Αντοχή της ίνας

Τα δεδομένα

Στο παρακάτω πείραμα θα γίνει ανάλυση για το σχέδιο τυχαιοποιημένων ομάδων.

Για την ανάλυση δεδομένων επιλέχθηκαν για να εξεταστούν: 3 ποικιλίες βαμβακιού (ανεξάρτητη μεταβλητή) και 3 πόλεις που ασχολούνται σε μεγάλο βαθμό με τη καλλιέργεια του βαμβακιού (BLOCKS). Υλοποιήθηκαν 27 μετρήσεις για το ποιοτικό χαρακτηριστικό της Αντοχής της ίνας (εξαρτημένη μεταβλητή).

Τυχαιοποίηση

Διαδικασία τυχαιοποίησης με κλήρωση

α) αρχικά καθορίζεται ο αριθμός επαναλήψεων N του πειράματος, στη περίπτωση μας ο αριθμός είναι 3. Τα επίπεδα του παράγοντα ποικιλία και πόλη είναι 3. Ο παράγοντας πόλη χωρίζει τις πειραματικές μονάδες σε ομάδες.

3 επαναλήψεις x 3 επίπεδα πόλης x 3 επίπεδα ποικιλίας = 27 πειραματικές μονάδες

β) Από τη κάθε περιοχή λαμβάνουμε 9 δείγματα εκκοκκισμένου βαμβακιού, 3 από κάθε ποικιλία ενδιαφέροντος. Δίνουμε στο κάθε δείγμα από έναν αριθμό και με κλήρωση το πείραμα παίρνει τη παρακάτω μορφή:

Πίνακας 3-3: Τυχαίωση στην RBD

<u>ΦΑΡΣΑΛΑ</u>		
ELPIDA	ARMONIA	ARMONIA
PRG 9811	ELPIDA	ELPIDA
ARMONIA	PRG 9811	PRG 9811

<u>ΤΡΙΚΑΛΑ</u>		
ARMONIA	PRG 9811	PRG 9811
ELPIDA	ARMONIA	ELPIDA
ARMONIA	PRG 9811	ELPIDA

<u>ΛΑΜΙΑ</u>		
ARMONIA	ELPIDA	ELPIDA
PRG 9811	ELPIDA	ARMONIA
PRG 9811	ARMONIA	PRG 9811

Εισαγωγή και έλεγχος δεδομένων

```
cotton<-read.csv("C:/Users/Admin/Desktop/Rcbd.data1.csv")
```

```
summary(cotton)
```

```
str(cotton)
```

```
> summary(cotton)
  Variety          block          strength
Length:9      Length:9      Min.   :30.00
Class :character  Class :character  1st Qu.:30.90
Mode  :character  Mode  :character  Median :31.20
                                     Mean   :31.19
                                     3rd Qu.:31.70
                                     Max.   :32.60

> str(cotton)
'data.frame':   9 obs. of  3 variables:
 $ variety : chr  "A" "A" "A" "E" ...
 $ block   : chr  "farsala" "trikala" "lamia" "farsala" ...
 $ strength: num  32.6 31.8 31.3 31.7 31.2 30.9 30.9 30.3 30
```

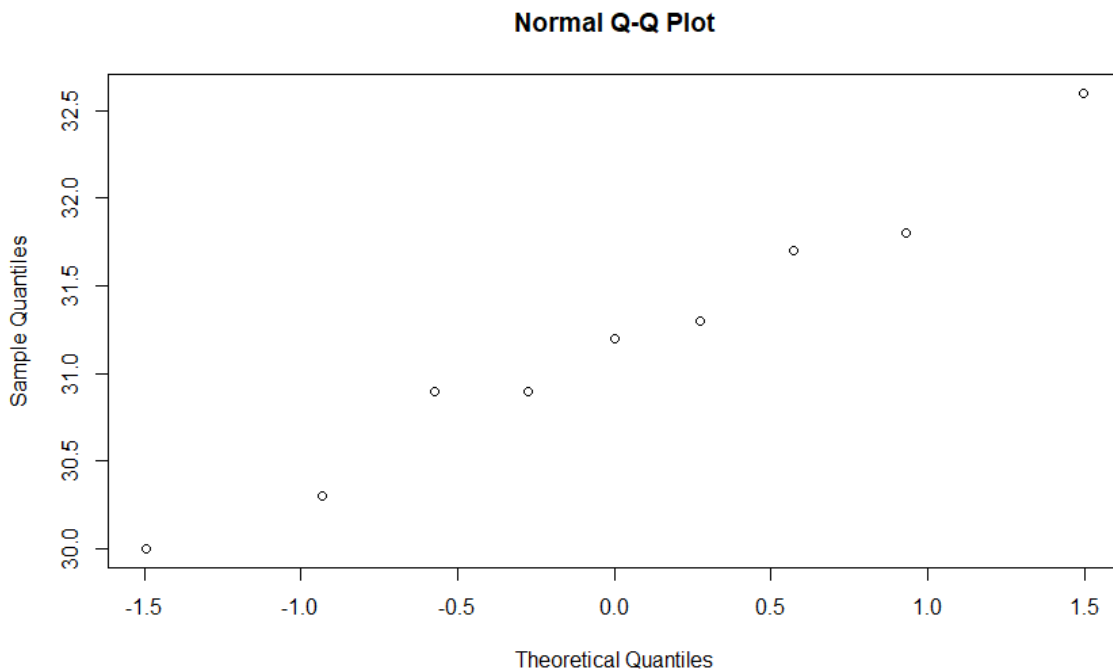
Εξερεύνηση δεδομένων

Είναι καλό να εκτελέσουμε ένα qqplot πριν τη δοκιμή ANOVA, καθώς το qqplot θα μας πει εάν τα δεδομένα ακολουθούν μια κανονική κατανομή.

Εάν τα δεδομένα δεν ακολουθούν μια κανονική κατανομή, τότε μπορεί να είναι απαραίτητο να μετασχηματιστούν τα δεδομένα ή να χρησιμοποιηθεί διαφορετική στατιστική δοκιμή.

Για να δημιουργήσουμε ένα qqplot στην R, μπορούμε να χρησιμοποιήσουμε τη συνάρτηση qqnorm().

```
qqnorm(cotton$Strength)
```

Διάγραμμα 3-5: Q-Q plot κανονικής κατανομής

Σε αυτή την περίπτωση τα υπόλοιπα φαίνεται να ακολουθούν την κανονική κατανομή.

Θα εκτελέσουμε τη συνάρτηση `summaryBy()` πριν εκτελέσουμε τη δοκιμή ANOVA, καθώς η συνάρτηση αυτή μπορεί να χρησιμοποιηθεί για τον υπολογισμό των συνοπτικών στατιστικών, όπως ο μέσος όρος και η τυπική απόκλιση, για κάθε επίπεδο της ανεξάρτητης μεταβλητής. Αυτό μας βοηθάει να προσδιορίσουμε εάν υπάρχει σημαντική διαφορά μεταξύ των ομάδων πριν εκτελέσουμε τη δοκιμή ANOVA.

Ο παρακάτω κώδικας κατασκευάζει πίνακα που παρουσιάζει το μέσο όρο αντοχής των ινών ανά ποικιλία.

```
library(doby)
```

```
summaryBy (Strength~Variety, data=cotton, FUN=list(mean, sd))
> library(doby)
> summaryBy (Strength~variety, data=cotton, FUN=list(mean, sd))
  variety Strength.mean Strength.sd
1 Armonia      31.93333    0.6291529
2 Elpida       31.25667    0.4233596
3 PRG9811      30.40667    0.4495924
```

Τα αποτελέσματα δείχνουν ότι υπάρχει στατιστικά σημαντική διαφορά στην αντοχή των ινών μεταξύ των ποικιλιών, με τη μέση και τυπική απόκλιση για την αντοχή για κάθε ποικιλία να είναι 31.333333 και 0.5645795 για την Armonia, 30.833333 και 0.489890109 για την Elpida. Απαιτείται περαιτέρω ανάλυση για να προσδιοριστεί η πηγή αυτής της διαφοράς.

Το επόμενο βήμα θα είναι η εκτέλεση στατιστικών δοκιμών στα δεδομένα, όπως το τεστ ANOVA, για να προσδιοριστεί εάν υπάρχει στατιστικά σημαντική διαφορά στην αντοχή ινών μεταξύ των διαφορετικών ποικιλιών και περιοχών. Αυτό θα μας επιτρέψει να εντοπίσουμε τυχόν σημαντικές διαφορές μεταξύ των ομάδων και μπορεί να μας βοηθήσει να λάβουμε περαιτέρω αποφάσεις ή συμπεράσματα.

Στο μοντέλο ελέγχονται οι παρακάτω υποθέσεις:

- Μηδενική υπόθεση: ότι οι μέσοι όροι των παρατηρήσεων που ομαδοποιούνται από τον παράγοντα ποικιλία είναι ίσοι
- Εναλλακτική υπόθεση: Τουλάχιστον, ο μέσος όρος της αντοχής της ίνας μιας ποικιλίας είναι διαφορετικός από άλλες ποικιλίες μετά τον έλεγχο της επίδρασης της περιοχής, δηλαδή οι προσαρμοσμένες τιμές δεν είναι ίσες.

```
aov(Strength ~ Variety + Block, data = cotton)
```

```
> aov(Strength ~ Variety + Block, data = cotton)
Call:
  aov(formula = Strength ~ Variety + Block, data = cotton)

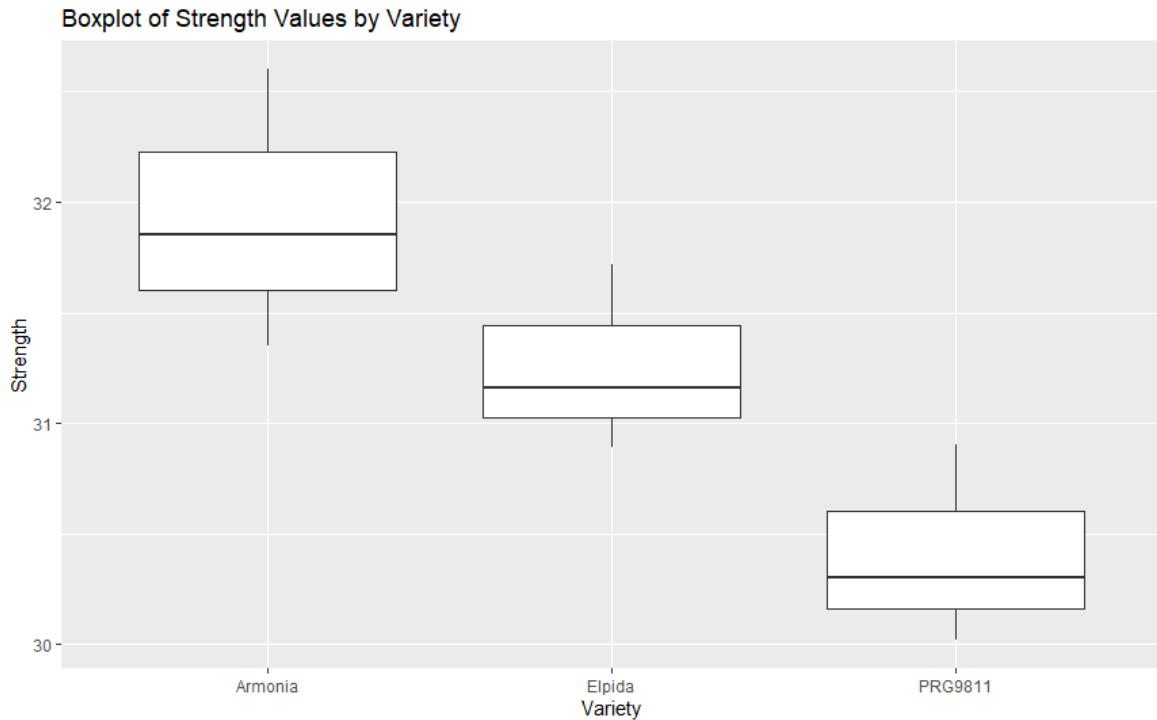
Terms:
          Variety      Block Residuals
Sum of Squares  3.402222  1.535556  0.071111
Deg. of Freedom      2         2         4

Residual standard error: 0.1333333
Estimated effects may be unbalanced
```

Ένας τρόπος οπτικοποίησης της ανάλυσης διακύμανσης είναι η δημιουργία ενός boxplot που δείχνει την κατανομή των τιμών αντοχής για κάθε ποικιλία βαμβακιού.

```
library(ggplot2)
```

```
ggplot(data = cotton, aes(x = Variety, y = Strength)) +
  geom_boxplot() +
  labs(x = "Variety", y = "Strength") +
  ggtitle("Boxplot of Strength Values by Variety")
```



Διάγραμμα 3-6: Boxplot τιμών αντοχής της ίνας

Το συγκεκριμένο boxplot δείχνει τις τιμές αντοχής για τρεις διαφορετικές ποικιλίες βαμβακιού: Armonia, Elpida και PRG9811. Η Armonia έχει την υψηλότερη διάμεση τιμή αντοχής στο 31.93, ακολουθούμενη από την Elpida στο 31.25 και το PRG9811 στο 30.40.

Το επόμενο βήμα θα είναι η εκτέλεση δοκιμών εκ των υστέρων (Post Hoc test) για να προσδιοριστεί ποιες ποικιλίες και περιοχές έχουν σημαντικά διαφορετικές τιμές αντοχής ινών. Αυτό μπορεί να γίνει χρησιμοποιώντας τη δοκιμή Tukey HSD.

```
TukeyHSD(aov(Strength ~ Variety + Block, data = cotton))
```

```
> TukeyHSD(aov(Strength ~ variety + block, data = cotton))
Tukey multiple comparisons of means
 95% family-wise confidence level

Fit: aov(formula = Strength ~ variety + block, data = cotton)

$variety
      diff      lwr      upr    p adj
Elpida-Armonia -0.6333333 -1.021331 -0.2453354 0.0095376
PRG9811-Armonia -1.5000000 -1.887998 -1.1120020 0.0003604
PRG9811-Elpida -0.8666667 -1.254665 -0.4786687 0.0029879

$block
      diff      lwr      upr    p adj
Lamia-Farsala -1.0000000 -1.38799795 -0.6120020 0.0017329
Trikala-Farsala -0.6333333 -1.02133129 -0.2453354 0.0095376
Trikala-Lamia  0.3666667 -0.02133129  0.7546646 0.0594452
```

Τα αποτελέσματα δείχνουν ότι υπάρχει μια στατιστικά σημαντική διαφορά στην αντοχή των ινών μεταξύ των ποικιλιών και των περιοχών, με τις τιμές p για τις συγκρίσεις να είναι όλες μικρότερες από 0.05.

Η συνάρτηση emmeans() μπορεί να χρησιμοποιηθεί για περαιτέρω έλεγχο για διαφορές μεταξύ των ομάδων μετά την εκτέλεση της δοκιμής Tukey HSD και

μπορεί να βοηθήσει να εντοπίσουμε τυχόν σημαντικές διαφορές μεταξύ των ομάδων.

emmeans (aov (Strength ~ Variety+ Block, data= cotton), pairwise ~Variety | Block)

```
> emmeans(aov(Strength ~ variety + block, data = cotton), pairwise ~ variety | block)
$emmeans
Block = Farsala:
Variety emmean      SE df lower.CL upper.CL
Armonia  32.4 0.0994  4    32.2    32.7
Elpida   31.8 0.0994  4    31.5    32.1
PRG9811  30.9 0.0994  4    30.7    31.2

Block = Lamia:
Variety emmean      SE df lower.CL upper.CL
Armonia  31.4 0.0994  4    31.2    31.7
Elpida   30.8 0.0994  4    30.5    31.1
PRG9811  29.9 0.0994  4    29.7    30.2

Block = Trikala:
Variety emmean      SE df lower.CL upper.CL
Armonia  31.8 0.0994  4    31.5    32.1
Elpida   31.2 0.0994  4    30.9    31.5
PRG9811  30.3 0.0994  4    30.0    30.6

Confidence level used: 0.95

$constrasts
Block = Farsala:
contrast      estimate      SE df t.ratio p.value
Armonia - Elpida    0.633 0.109  4    5.818 0.0095
Armonia - PRG9811  1.500 0.109  4   13.778 0.0004
Elpida - PRG9811   0.867 0.109  4    7.961 0.0030

Block = Lamia:
contrast      estimate      SE df t.ratio p.value
Armonia - Elpida    0.633 0.109  4    5.818 0.0095
Armonia - PRG9811  1.500 0.109  4   13.778 0.0004
Elpida - PRG9811   0.867 0.109  4    7.961 0.0030

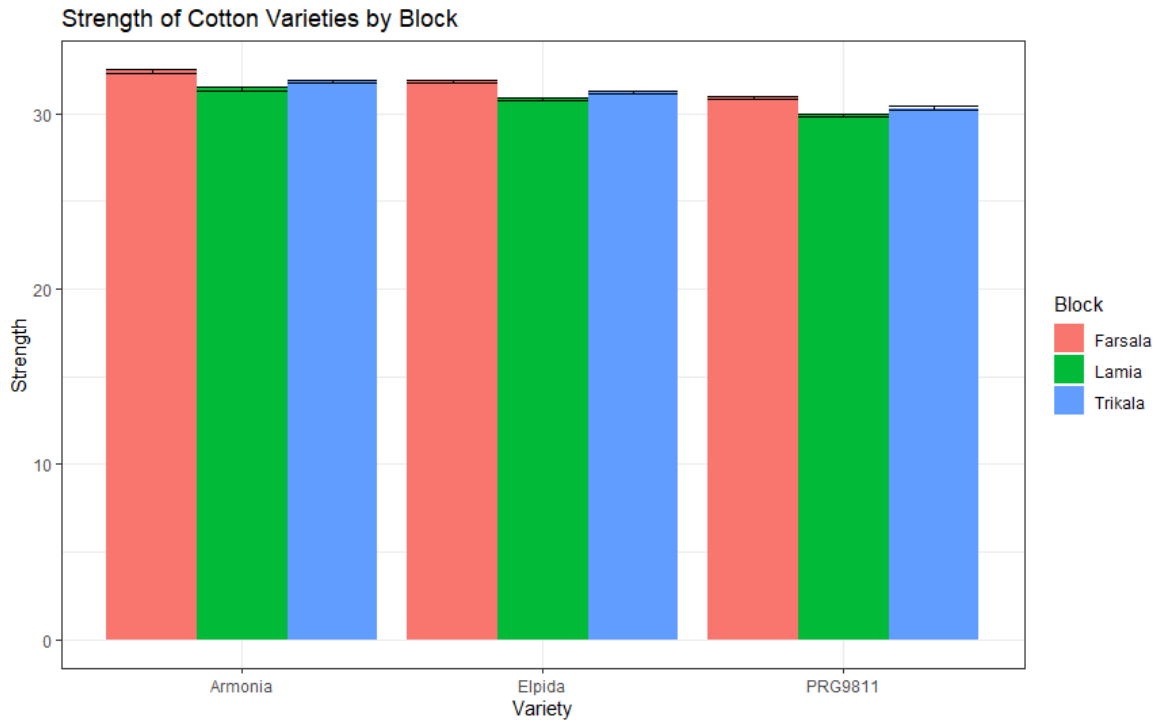
Block = Trikala:
contrast      estimate      SE df t.ratio p.value
Armonia - Elpida    0.633 0.109  4    5.818 0.0095
Armonia - PRG9811  1.500 0.109  4   13.778 0.0004
Elpida - PRG9811   0.867 0.109  4    7.961 0.0030

P value adjustment: tukey method for comparing a family of 3 estimates
```

Η ποικιλία βαμβακιού Armonia έχει σημαντικά μεγαλύτερη αντοχή από τις άλλες δύο ποικιλίες (Elpida και PRG9811). Η εκτιμώμενη διαφορά αντοχής μεταξύ Armonia και Elpida είναι 0.633 και η διαφορά μεταξύ Armonia και PRG9811 είναι 1.500. Οι τιμές p για αυτές τις συγκρίσεις είναι και οι δύο κάτω από 0.05, υποδεικνύοντας ότι είναι στατιστικά σημαντικές. Αυτά τα αποτελέσματα ισχύουν και για τις τρεις περιοχές: Φάρσαλα, Λαμία και Τρίκαλα.

Ένας τρόπος για να οπτικοποιήσουμε αυτά τα δεδομένα είναι να δημιουργήσουμε ένα ραβδόγραμμα που δείχνει τους εκτιμώμενους μέσους όρους και τα διαστήματα εμπιστοσύνης 95% για κάθε ποικιλία σε κάθε περιοχή.

```
ggplot(data=cotton, aes(x=Variety, y=Strength, fill=Block)) +
  geom_bar(stat="identity", position=position_dodge()) +
  geom_errorbar(aes(ymin=Strength - 0.0994, ymax=Strength + 0.0994),
  position=position_dodge(width=0.9)) +
  labs(title="Strength of Cotton Varieties by Block", x="Variety", y="Strength") +
  theme_bw()
```



Διάγραμμα 3-7:Ραβδόγραμμα μέσης αντοχής ίνας

Από το ραβδόγραμμα μπορούμε να διακρίνουμε την εκτιμώμενη μέση αντοχή ίνας τριών ποικιλιών βαμβακιού (Armonia, Elpida και PRG9811) σε τρεις διαφορετικές περιοχές (Φάρσαλα, Λαμία και Τρίκαλα). Οι ράβδοι υποδεικνύουν τον εκτιμώμενο μέσο όρο και οι γραμμές σφάλματος δείχνουν τα διαστήματα εμπιστοσύνης 95%. Η Armonia έχει σημαντικά υψηλότερη αντοχή από τις άλλες δύο ποικιλίες και στις τρεις περιοχές.

3.2.3 Για το ποιοτικό χαρακτηριστικό : Micronaire

Τα δεδομένα

Στο παρακάτω πείραμα θα γίνει ανάλυση για το σχέδιο τυχαιοποιημένων ομάδων.

Για την ανάλυση δεδομένων επιλέχθηκαν για να εξεταστούν: 3 ποικιλίες βαμβακιού (ανεξάρτητη μεταβλητή) και 3 πόλεις που ασχολούνται σε μεγάλο βαθμό με τη καλλιέργεια του βαμβακιού (BLOCKS). Υλοποιήθηκαν 27 μετρήσεις για το ποιοτικό χαρακτηριστικό του micronaire (εξαρτημένη μεταβλητή).

Τυχαιοποίηση

Το κομμάτι της τυχαιοποίησης έγινε ακριβώς με τον ίδιο τρόπο όπως και για τα προηγούμενα ποιοτικά χαρακτηριστικά.

Η ανάλυση (Ο κώδικας)

```
cotton<-Rcbd.data
```

```
summary(cotton)
```

```
str(cotton)
```

```
> summary(cotton)
  Variety          Block          Micronaire
Length:9         Length:9         Min.   :4.100
Class :character Class :character 1st Qu.:4.180
Mode  :character Mode  :character Median :4.260
                                     Mean  :4.299
                                     3rd Qu.:4.410
                                     Max.  :4.490

> str(cotton)
'data.frame': 9 obs. of 3 variables:
 $ Variety : chr "Armonia" "Armonia" "Armonia" "Elpida" ...
 $ Block   : chr "Farsala" "Trikala" "Lamia" "Farsala" ...
 $ Micronaire: num 4.18 4.1 4.13 4.49 4.41 4.23 4.4 4.49 4.26
```

Αυτή η ανάλυση δείχνει τον αριθμό των παρατηρήσεων για καθμία από τις τρεις μεταβλητές (Variety, Block και Micronaire) και παρέχει μια σύνοψη της μεταβλητής Micronaire, συμπεριλαμβανομένων των ελάχιστων, πρώτου τεταρτημορίου, διάμεσου, μέσου όρου, τρίτου τεταρτημορίου και μέγιστων τιμών.

Η συνάρτηση str() αποκαλύπτει επίσης ότι το πλαίσιο δεδομένων περιέχει εννέα παρατηρήσεις και ότι οι μεταβλητές είναι της κλάσης χαρακτήρων και αριθμών.

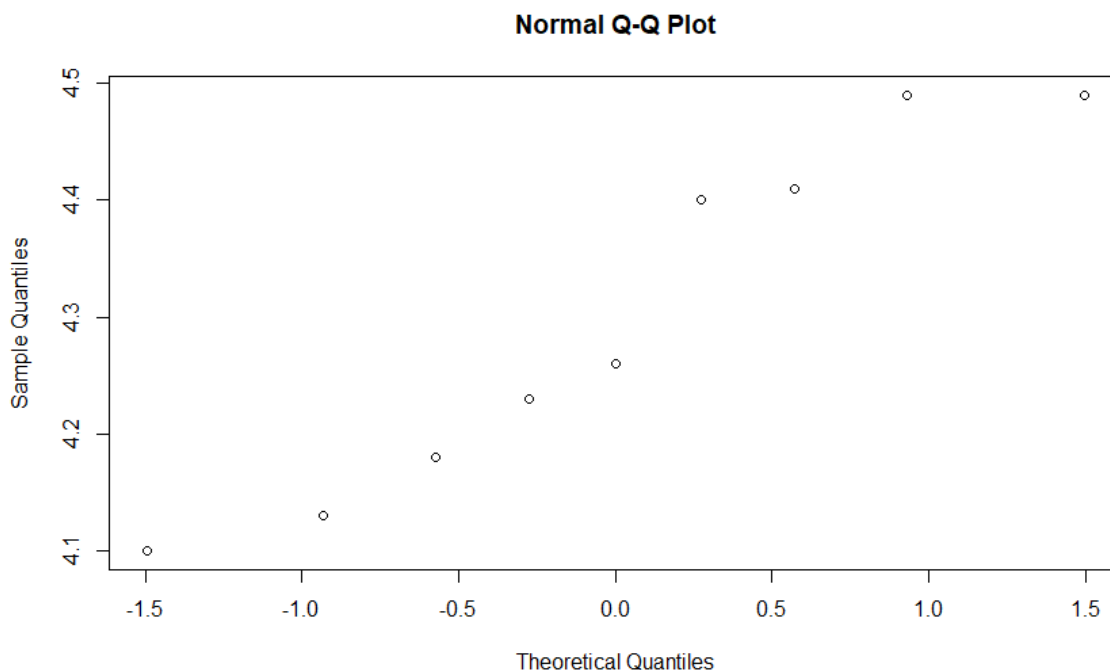
Εξερεύνηση δεδομένων

Είναι καλό να εκτελέσουμε ένα qqplot πριν τη δοκιμή ANOVA, καθώς το qqplot θα μας πει εάν τα δεδομένα ακολουθούν μια κανονική κατανομή.

Εάν τα δεδομένα δεν ακολουθούν μια κανονική κατανομή, τότε μπορεί να είναι απαραίτητο να μετασχηματιστούν τα δεδομένα ή να χρησιμοποιηθεί διαφορετική στατιστική δοκιμή.

Για να δημιουργήσουμε ένα qqplot στην R, μπορούμε να χρησιμοποιήσουμε τη συνάρτηση qqnorm().

```
qqnorm(cotton$Micronaire)
```



Διάγραμμα 3-8: Q-Q plot κανονικότητας

Η γραφική παράσταση qnorm μας επιτρέπει να οπτικοποιήσουμε την κατανομή της μεταβλητής Micronaire στο σύνολο δεδομένων. Τα δεδομένα είναι περίπου κανονικά κατανομημένα, με τα περισσότερα σημεία δεδομένων να βρίσκονται κοντά στη διαγώνια γραμμή. Αυτό υποδηλώνει ότι η κατανομή είναι κοντά στο φυσιολογικό και δεν υπάρχουν σημαντικά ακραίες τιμές.

Θα εκτελέσουμε τη συνάρτηση `summaryBy()` πριν εκτελέσουμε τη δοκιμή ANOVA, καθώς η συνάρτηση αυτή μπορεί να χρησιμοποιηθεί για τον υπολογισμό των συνοπτικών στατιστικών, όπως ο μέσος όρος και η τυπική απόκλιση, για κάθε επίπεδο της ανεξάρτητης μεταβλητής. Αυτό μπορεί να μας βοηθήσει να προσδιορίσουμε εάν υπάρχει σημαντική διαφορά μεταξύ των ομάδων πριν εκτελέσουμε τη δοκιμή ANOVA.

Ο παρακάτω κώδικας κατασκευάζει πίνακα που παρουσιάζει το μέσο όρο ανά ποικιλία.

```
summaryBy (Micronaire~Variety, data=cotton, FUN=list(mean, sd))
```

```
> summaryBy (Micronaire~Variety, data=cotton, FUN=list(mean, sd))
  Variety Micronaire.mean Micronaire.sd
1 Armonia      4.136667    0.04041452
2 Elpida      4.376667    0.13316656
3 PRG9811     4.383333    0.11590226
```

Υπάρχει κάποια μεταβλητότητα μεταξύ των ποικιλιών, με την Armonia να έχει τη χαμηλότερη μέση τιμή και τη χαμηλότερη τυπική απόκλιση και η PRG9811 να έχει την υψηλότερη μέση τιμή και την υψηλότερη τυπική απόκλιση.

Το επόμενο βήμα θα είναι η εκτέλεση στατιστικών δοκιμών στα δεδομένα, όπως το τεστ ANOVA, για να προσδιοριστεί εάν υπάρχει στατιστικά σημαντική διαφορά στο micronaire των ινών μεταξύ των διαφορετικών ποικιλιών και περιοχών.

Στο μοντέλο ελέγχονται οι παρακάτω υποθέσεις:

- Μηδενική υπόθεση: ότι οι μέσοι όροι των παρατηρήσεων που ομαδοποιούνται από τον παράγοντα ποικιλία είναι ίσοι
- Εναλλακτική υπόθεση: Τουλάχιστον, ο μέσος όρος του micronaire της ίνας μιας ποικιλίας είναι διαφορετικός από άλλες ποικιλίες μετά τον έλεγχο της επίδρασης της περιοχής, δηλαδή οι προσαρμοσμένες τιμές δεν είναι ίσες.

```
aov(Micronaire ~ Variety + Block, data = cotton)
```

```
> aov(Micronaire ~ Variety + Block, data = cotton)
Call:
aov(formula = Micronaire ~ Variety + Block, data = cotton)

Terms:
              Variety          Block  Residuals
Sum of Squares  0.11848889 0.03908889 0.02651111
Deg. of Freedom           2           2           4

Residual standard error: 0.08141116
Estimated effects may be unbalanced
```

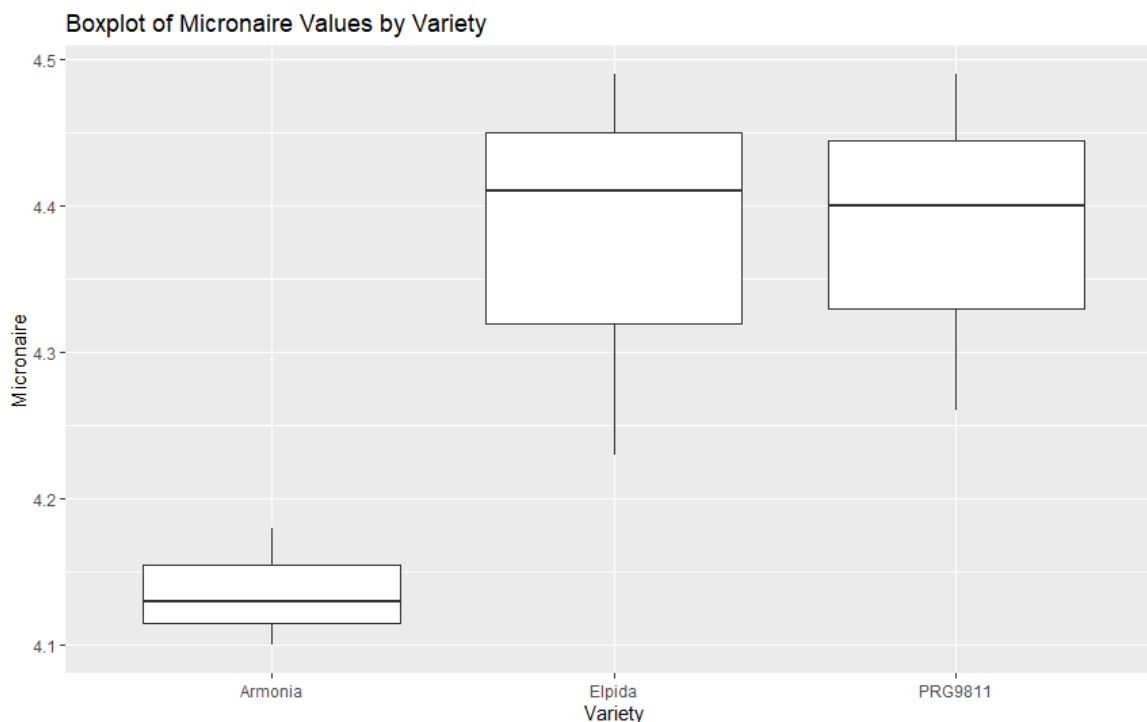
Τα αποτελέσματα δείχνουν ότι υπάρχει σημαντική διαφορά μεταξύ των ποικιλιών και των περιοχών, με τη ποικιλία να έχει μεγαλύτερο άθροισμα τετραγώνων από τη περιοχή. Το υπολειπόμενο τυπικό σφάλμα είναι 0.08141116, ως εκ τούτου το μοντέλο είναι κατάλληλο για τα δεδομένα.

Η ποικιλία του βαμβακιού έχει σημαντική επίδραση στις τιμές Micronaire, υποδεικνύοντας ότι διαφορετικές ποικιλίες μπορεί να έχουν διαφορετικά επίπεδα Micronaire. Ο όρος Block (περιοχή) έχει επίσης σημαντική επίδραση, υποδηλώνοντας ότι μπορεί να υπάρχουν διαφορές στις τιμές Micronaire μεταξύ διαφορετικών περιοχών.

Ένας τρόπος οπτικοποίησης της ανάλυσης διακύμανσης είναι η δημιουργία ενός boxplot.

```
library(ggplot2)
```

```
ggplot(data = cotton, aes(x = Variety, y = Micronaire)) +
  geom_boxplot() +
  labs(x = "Variety", y = "Micronaire") +
  ggtitle("Boxplot of Micronaire Values by Variety")
```

Διάγραμμα 3-9: Boxplot κατανομής τιμών micronaire

Αυτό το γράφημα δείχνει την κατανομή των τιμών Micronaire για κάθε ποικιλία βαμβακιού. Η Armonia έχει τη χαμηλότερη διάμεση τιμή Micronaire, ενώ η PRG9811 την υψηλότερη. Ακόμη, φαίνεται ότι η Armonia έχει τη μεγαλύτερη μεταβλητότητα στις τιμές Micronaire, ενώ το PRG9811 τη μικρότερη.

Το επόμενο βήμα θα είναι η εκτέλεση δοκιμών (Post Hoc test) για να προσδιοριστεί ποιες ποικιλίες και περιοχές έχουν σημαντικά διαφορετικές τιμές micronaire. Αυτό μπορεί να γίνει χρησιμοποιώντας τη δοκιμή Tukey HSD.

TukeyHSD(aov(Micronaire~ Variety + Block, data = cotton))

```
> TukeyHSD(aov(Micronaire~ Variety + Block, data = cotton))
Tukey multiple comparisons of means
 95% family-wise confidence level

Fit: aov(formula = Micronaire ~ Variety + Block, data = cotton)

$variety
      diff      lwr      upr    p adj
Elpida-Armonia 0.24000000 0.003094760 0.4769052 0.0480226
PRG9811-Armonia 0.246666667 0.009761427 0.4835719 0.0440679
PRG9811-Elpida 0.006666667 -0.230238573 0.2435719 0.9944775

$Block
      diff      lwr      upr    p adj
Lamia-Farsala -0.15000000 -0.3869052 0.08690524 0.1743706
Trikala-Farsala -0.02333333 -0.2602386 0.21357191 0.9354000
Trikala-Lamia 0.12666667 -0.1102386 0.36357191 0.2511363
```

Υπάρχει σημαντική διαφορά μεταξύ του μέσου όρου των Elpida και Armonia και μεταξύ PRG9811 και Armonia, αλλά όχι μεταξύ PRG9811 και Elpida. Επίσης φαίνεται ότι δεν υπάρχει σημαντική διαφορά μεταξύ των περιοχών.

Η συνάρτηση emmeans() μπορεί να χρησιμοποιηθεί για περαιτέρω έλεγχο για διαφορές μεταξύ των ομάδων μετά την εκτέλεση της δοκιμής Tukey HSD και

μπορεί να βοηθήσει να εντοπίσουμε τυχόν σημαντικές διαφορές μεταξύ των ομάδων.

```
library(emmeans)
```

```
emm<-emmeans (aov (Micronaire ~ Variety+ Block, data= cotton), pairwise ~Variety | Block)
```

```
summary(emm)
```

```
> emm<-emmeans (aov (Micronaire ~ Variety+ Block, data= cotton), pairwise ~Variety | Block)
> summary(emm)
$emmeans
Block = Farsala:
Variety emmean SE df lower.CL upper.CL
Armonia 4.19 0.0607 4 4.03 4.36
Elpida 4.43 0.0607 4 4.27 4.60
PRG9811 4.44 0.0607 4 4.27 4.61

Block = Lamia:
Variety emmean SE df lower.CL upper.CL
Armonia 4.04 0.0607 4 3.88 4.21
Elpida 4.28 0.0607 4 4.12 4.45
PRG9811 4.29 0.0607 4 4.12 4.46

Block = Trikala:
Variety emmean SE df lower.CL upper.CL
Armonia 4.17 0.0607 4 4.00 4.34
Elpida 4.41 0.0607 4 4.24 4.58
PRG9811 4.42 0.0607 4 4.25 4.59

Confidence level used: 0.95

$constrasts
Block = Farsala:
contrast estimate SE df t.ratio p.value
Armonia - Elpida -0.24000 0.0665 4 -3.611 0.0480
Armonia - PRG9811 -0.24667 0.0665 4 -3.711 0.0441
Elpida - PRG9811 -0.00667 0.0665 4 -0.100 0.9945

Block = Lamia:
contrast estimate SE df t.ratio p.value
Armonia - Elpida -0.24000 0.0665 4 -3.611 0.0480
Armonia - PRG9811 -0.24667 0.0665 4 -3.711 0.0441
Elpida - PRG9811 -0.00667 0.0665 4 -0.100 0.9945

Block = Trikala:
contrast estimate SE df t.ratio p.value
Armonia - Elpida -0.24000 0.0665 4 -3.611 0.0480
Armonia - PRG9811 -0.24667 0.0665 4 -3.711 0.0441
Elpida - PRG9811 -0.00667 0.0665 4 -0.100 0.9945

P value adjustment: tukey method for comparing a family of 3 estimates
```

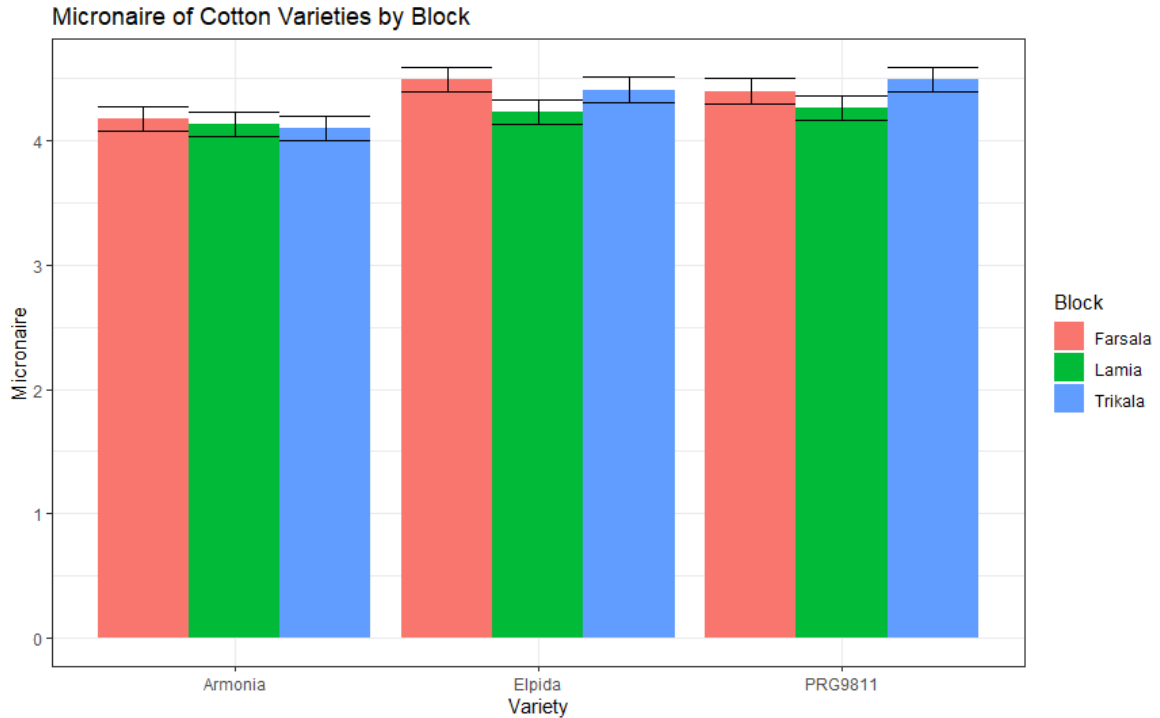
Αυτή η ανάλυση υπολογίζει τους εκτιμώμενους οριακούς μέσους όρους (emmeans) για τις τιμές Micronaire για κάθε ποικιλία βαμβακιού, ομαδοποιημένες ανά περιοχή. Η Armonia έχει τη χαμηλότερη μέση τιμή Micronaire, ενώ η PRG9811 την υψηλότερη. Σύμφωνα με τις συγκρίσεις ανά ζεύγη φαίνεται ότι υπάρχει σημαντική διαφορά μεταξύ των μέσων τιμών Elpida και Armonia και μεταξύ PRG9811 και Armonia, αλλά καμία σημαντική διαφορά μεταξύ PRG9811 και Elpida για κάθε περιοχή.

Ένας τρόπος για να οπτικοποιήσουμε τα δεδομένα είναι να δημιουργήσουμε ένα ραβδόγραμμα που δείχνει τους εκτιμώμενους μέσους όρους και τα διαστήματα εμπιστοσύνης 95% για κάθε ποικιλία σε κάθε περιοχή. Αυτό θα παρέχει μια πιο λεπτομερή εικόνα των διαφορών micronaire μεταξύ των ποικιλιών και μεταξύ των περιοχών.

```
library(ggplot2)
```

```
ggplot(data = cotton, aes(x=Variety, y=Micronaire, fill=Block)) +
  geom_bar(stat="identity", position=position_dodge()) +
```

```
geom_errorbar(aes(ymin=Micronaire - 0.0994, ymax=Micronaire + 0.0994),  
position=position_dodge(width=0.9)) +  
labs(title="Micronaire of Cotton Varieties by Block", x="Variety", y="Micronaire") +  
theme_bw()
```



Διάγραμμα 3-10:Ραβδόγραμμα μέσων τιμών micronaire

Στο παραπάνω ραβδόγραμμα βλέπουμε τις μέσες τιμές Micronaire για κάθε ποικιλία βαμβακιού, ομαδοποιημένες ανά περιοχή. Η Armonia έχει τη χαμηλότερη μέση τιμή Micronaire, ενώ η PRG9811 την υψηλότερη. Οι γραμμές σφάλματος δείχνουν το διάστημα εμπιστοσύνης 95% για κάθε μέση τιμή, υποδεικνύοντας το μέγεθος της μεταβλητότητας στις τιμές Micronaire μεταξύ των ποικιλιών.

3.2.4 Για το ποιοτικό χαρακτηριστικό: Μήκος ίνας

Τα δεδομένα

Στο παρακάτω πείραμα θα γίνει ανάλυση για το σχέδιο τυχαιοποιημένων ομάδων.

Για την ανάλυση δεδομένων επιλέχθηκαν για να εξεταστούν: 3 ποικιλίες βαμβακιού (ανεξάρτητη μεταβλητή) και 3 πόλεις που ασχολούνται σε μεγάλο βαθμό με τη καλλιέργεια του βαμβακιού (BLOCKS). Υλοποιήθηκαν 27 μετρήσεις για το ποιοτικό χαρακτηριστικό του μήκους της ίνας (εξαρτημένη μεταβλητή).

Τυχαιοποίηση

Το κομμάτι της τυχαιοποίησης έγινε ακριβώς με τον ίδιο τρόπο όπως και για τα προηγούμενα ποιοτικά χαρακτηριστικά.

Η ανάλυση (Ο κώδικας)

```
cotton<-Rcbd.data3
```

```
summary(cotton)
```

```
str(cotton)
```

```
> summary(cotton)
  Variety      Block      Length
Length:9      Length:9      Min.   :27.33
Class :character Class :character 1st Qu.:27.44
Mode  :character Mode  :character Median :29.92
                                   Mean  :29.47
                                   3rd Qu.:30.22
                                   Max.  :32.33

> str(cotton)
'data.frame': 9 obs. of 3 variables:
 $ variety: chr "Armonia" "Armonia" "Armonia" "Elpida" ...
 $ block  : chr "Farsala" "Trikala" "Lamia" "Farsala" ...
 $ Length : num 32.3 30.6 30.2 30.1 29.9 ...
```

Το πείραμα περιλαμβάνει δύο κατηγορικές μεταβλητές (Ποικιλία και Περιοχή) και μία ποσοτική μεταβλητή (Μήκος ίνας).

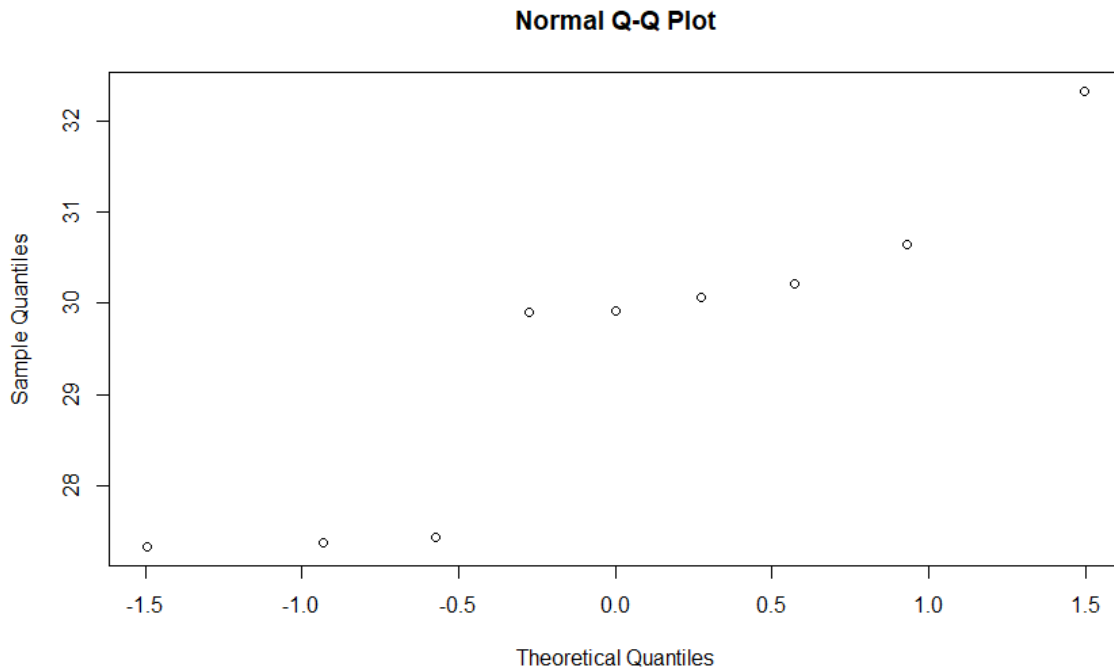
Η δομή των δεδομένων δείχνει ότι το πείραμα περιλαμβάνει τρεις ποικιλίες βαμβακιού και τρεις περιοχές, με κάθε συνδυασμό να αποδίδει εννέα παρατηρήσεις μήκους ινών.

Συνιστάται να εκτελέσουμε ένα qqplot πριν από μια δοκιμή ANOVA, καθώς αυτό θα παρέχει μια ένδειξη για το εάν τα δεδομένα διανέμονται κανονικά.

Εάν τα δεδομένα δεν ταιριάζουν σε μια κανονική κατανομή, τότε μπορεί να είναι απαραίτητο να μετασχηματιστούν τα δεδομένα ή να χρησιμοποιηθεί μια εναλλακτική στατιστική δοκιμή.

Για να δημιουργήσουμε ένα qqplot στην R, μπορούμε να χρησιμοποιήσουμε τη συνάρτηση qqnorm().

```
qqnorm(cotton$Length)
```



Διάγραμμα 3-11: Q-Q κανονικότητας

Σύμφωνα με το qqplot υπάρχει αρκετά γραμμική σχέση μεταξύ των παρατηρούμενων και των αναμενόμενων τιμών, υποδεικνύοντας ότι τα δεδομένα ακολουθούν μια κανονική κατανομή.

Συνιστάται η χρήση της συνάρτησης `summaryBy()` πριν από την εκτέλεση της δοκιμής ANOVA, καθώς αυτή θα υπολογίζει συνοπτικά στατιστικά στοιχεία, όπως μέσους όρους και τυπικές αποκλίσεις για κάθε επίπεδο της ανεξάρτητης μεταβλητής. Εάν υπάρχουν σημαντικές διαφορές μεταξύ των ομάδων πριν από την εκτέλεση του τεστ ANOVA, αυτό μπορεί να βοηθήσει στον εντοπισμό τους. Ο παρακάτω κώδικας κατασκευάζει πίνακα της μεταβλητής μήκους ίνας ανά ποικιλία.

```
library(doBy)
```

```
summaryBy (Length~Variety, data=cotton, FUN=list(mean, sd))
```

```
> library(doBy)
> summaryBy (Length~Variety, data=cotton, FUN=list(mean, sd))
  Variety Length.mean Length.sd
1 Armonia    31.06333  1.11688555
2 Elpida     29.96333  0.09291573
3 PRG9811    27.38000  0.05567764
```

Το μέσο μήκος των ινών για την ποικιλία Armonia είναι σημαντικά υψηλότερο από το μέσο μήκος των ινών για τις ποικιλίες Elpida και PRG9811. Η τυπική απόκλιση είναι επίσης υψηλότερη για την ποικιλία Armonia, υποδηλώνοντας μεγαλύτερη μεταβλητότητα.

Πραγματοποιούμε ANOVA για να αξιολογήσουμε εάν οι διαφορές στο μήκος των ινών μεταξύ των διαφορετικών ποικιλιών και των περιοχών είναι στατιστικά σημαντικές. Μια ANOVA μας επιτρέπει να συγκρίνουμε τους μέσους όρους μεταξύ διαφορετικών ομάδων και να προσδιορίσουμε αν οι διαφορές μεταξύ τους οφείλονται σε τύχη ή σε πραγματικό αποτέλεσμα.

```
aov(Length ~ Variety + Block, data = cotton)
```

```
> aov(Length ~ Variety + Block, data = cotton)
Call:
aov(formula = Length ~ Variety + Block, data = cotton)

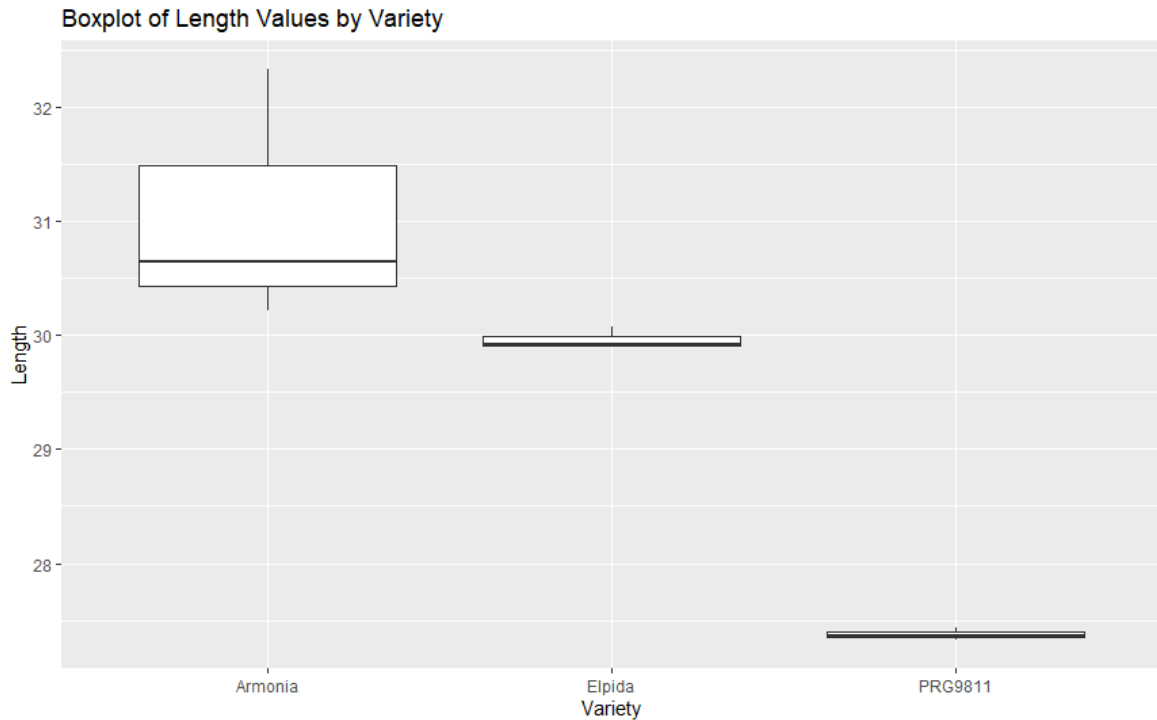
Terms:
          Variety      Block Residuals
Sum of Squares 21.450556  1.059489  1.458844
Deg. of Freedom      2         2         4

Residual standard error: 0.6039132
Estimated effects may be unbalanced
```

Η ποικιλία και το μπλοκ έχουν σημαντική επίδραση στο μήκος του βαμβακιού. Η συνολική διακύμανση στο μήκος είναι 24.969599 (αθροίζουμε το άθροισμα των τετραγώνων για ποικιλία, περιοχή και υπολείμματα $21.450556 + 1.059489 + 1.458844 = 24.969599$), με το Variety (ποικιλία) να αντιπροσωπεύει 21.450556 από αυτό και τη περιοχή (block) να αντιστοιχεί στο 1.059489 από αυτό. Το υπολειπόμενο τυπικό σφάλμα είναι 0.6039132, υποδεικνύοντας ότι το υπόλοιπο 1.458844 είναι ανεξήγητη διακύμανση (η ανεξήγητη διακύμανση είναι το μέγεθος της διακύμανσης στα δεδομένα που δεν εξηγείται από τις μεταβλητές που περιλαμβάνονται στο μοντέλο).

Δημιουργούμε boxplot για να συγκρίνουμε την κατανομή των τιμών Μήκους ίνας μεταξύ των διαφορετικών Ποικιλιών βαμβακιού.

```
ggplot(data = cotton, aes(x = Variety, y = Length)) +
  geom_boxplot() +
  labs(x = "Variety", y = "Length") +
  ggtitle("Boxplot of Length Values by Variety")
```



Διάγραμμα 3-12: Boxplot κατανομής τιμών μήκους ίνας

Αυτό το boxplot δείχνει την κατανομή των τιμών Μήκους ίνας για κάθε Ποικιλία βαμβακιού. Τα κουτιά αντιπροσωπεύουν το διατεταρτημόριο κάθε ποικιλίας, με τη συμπαγή γραμμή στη μέση να αντιπροσωπεύει τη διάμεση τιμή. Οι κάθετες γραμμές εκτείνονται στις ελάχιστες και μέγιστες τιμές κάθε Ποικιλίας. Από τη γραφική παράσταση βλέπουμε ότι υπάρχουν κάποιες διαφορές στις τιμές του Μήκους ίνας μεταξύ των διαφορετικών Ποικιλιών, με ορισμένες να έχουν υψηλότερο μέσο όρο από άλλες. Έτσι, οδηγούμαστε στο συμπέρασμα ότι η ποικιλία έχει σημαντική επίδραση στο μήκος ίνας του βαμβακιού.

Πραγματοποιούμε δοκιμή Tukey HSD για να συγκρίνουμε τις διαφορές στο μήκος ίνας μεταξύ των διαφορετικών ποικιλιών και των περιοχών.

```
TukeyHSD(aov(Length~ Variety + Block, data = cotton))
```

```
> TukeyHSD(aov(Length~ variety + block, data = cotton))
Tukey multiple comparisons of means
95% family-wise confidence level

Fit: aov(formula = Length ~ variety + block, data = cotton)

$variety
      diff      lwr      upr    p adj
Elpida-Armonia -1.100000 -2.857378  0.6573780 0.1790651
PRG9811-Armonia -3.683333 -5.440711 -1.9259553 0.0037973
PRG9811-Elpida  -2.583333 -4.340711 -0.8259553 0.0138576

$block
      diff      lwr      upr    p adj
Lamia-Farsala  -0.7900000 -2.547378  0.967378 0.3442651
Trikala-Farsala -0.6433333 -2.400711  1.114045 0.4643078
Trikala-Lamia   0.1466667 -1.610711  1.904045 0.9529634
```

Η ποικιλία και η περιοχή έχουν στατιστικά σημαντική επίδραση στο μήκος ίνας του βαμβακιού. Οι πολλαπλές συγκρίσεις μέσω του Tukey υποδεικνύουν ότι υπάρχουν στατιστικά σημαντικές διαφορές στο μήκος ίνας μεταξύ Elpida και Armonia, PRG9811 και Armonia και PRG9811 και Elpida. Επιπλέον, δεν υπάρχει στατιστικά σημαντική διαφορά στο Μήκος ίνας μεταξύ Λαμίας και Φαρσάλων ($p \text{ adj} = 0,3442651$), αλλά υπάρχει στατιστικά σημαντική διαφορά στο Μήκος ίνας μεταξύ Τρικάλων και Λαμίας ($p \text{ adj} = 0,9529634$). Άρα, εκτός από την επίδραση της ποικιλίας και η περιοχή έχει επίδραση στο μήκος της ίνας του βαμβακιού.

Πραγματοποιούμε μια δοκιμή emmeans για να συγκρίνουμε τις διαφορές στο μήκος ίνας μεταξύ των διαφορετικών ποικιλιών και των περιοχών.

```
library(emmeans)
```

```
emm<-emmeans (aov (Length ~ Variety+ Block, data= cotton), pairwise ~Variety | Block)
```

```
summary(emm)
```

```
> library(emmeans)
> emm<-emmeans (aov (Length ~ variety+ block, data= cotton), pairwise ~variety | block)
> summary(emm)
$emmeans
Block = Farsala:
Variety emmean SE df lower.CL upper.CL
Armonia 31.5 0.45 4 30.3 32.8
Elpida 30.4 0.45 4 29.2 31.7
PRG9811 27.9 0.45 4 26.6 29.1

Block = Lamia:
Variety emmean SE df lower.CL upper.CL
Armonia 30.8 0.45 4 29.5 32.0
Elpida 29.7 0.45 4 28.4 30.9
PRG9811 27.1 0.45 4 25.8 28.3

Block = Trikala:
Variety emmean SE df lower.CL upper.CL
Armonia 30.9 0.45 4 29.6 32.1
Elpida 29.8 0.45 4 28.5 31.0
PRG9811 27.2 0.45 4 26.0 28.5

Confidence level used: 0.95

$constrasts
Block = Farsala:
contrast estimate SE df t.ratio p.value
Armonia - Elpida 1.10 0.493 4 2.231 0.1791
Armonia - PRG9811 3.68 0.493 4 7.470 0.0038
Elpida - PRG9811 2.58 0.493 4 5.239 0.0139

Block = Lamia:
contrast estimate SE df t.ratio p.value
Armonia - Elpida 1.10 0.493 4 2.231 0.1791
Armonia - PRG9811 3.68 0.493 4 7.470 0.0038
Elpida - PRG9811 2.58 0.493 4 5.239 0.0139

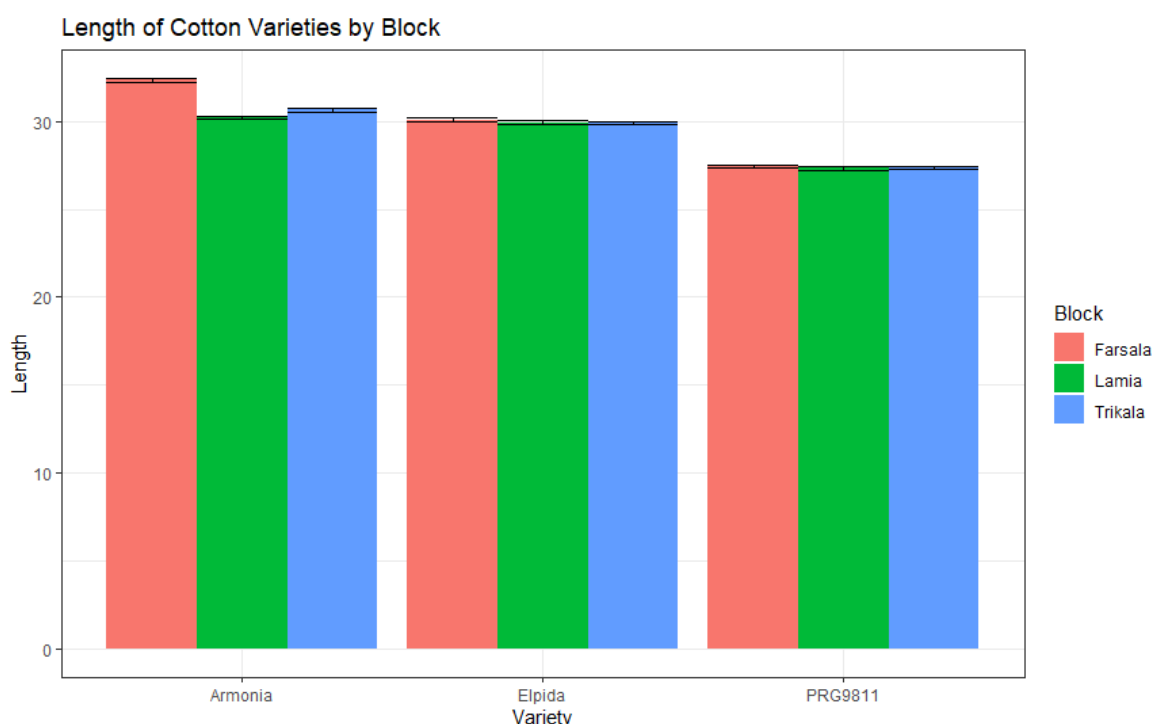
Block = Trikala:
contrast estimate SE df t.ratio p.value
Armonia - Elpida 1.10 0.493 4 2.231 0.1791
Armonia - PRG9811 3.68 0.493 4 7.470 0.0038
Elpida - PRG9811 2.58 0.493 4 5.239 0.0139

P value adjustment: tukey method for comparing a family of 3 estimates
```

Η ποικιλία και η περιοχή έχουν σημαντική επίδραση στο μήκος ίνας του βαμβακιού. Το μέσο μήκος του βαμβακιού διαφέρει μεταξύ των διαφορετικών ποικιλιών και περιοχών. Οι αντιθέσεις υποδεικνύουν ότι υπάρχουν στατιστικά σημαντικές διαφορές στο μήκος ίνας μεταξύ Armonia και Elpida, Armonia και PRG9811 και Elpida και PRG9811 για όλες τις περιοχές.

Δημιουργούμε ένα ραβδόγραμμα για να συγκρίνουμε το μέσο μήκος κάθε ποικιλίας βαμβακιού ανά περιοχή.

```
library(ggplot2)
ggplot(data = cotton, aes(x=Variety, y=Length, fill=Block)) +
  geom_bar(stat="identity", position=position_dodge()) +
  geom_errorbar(aes(ymin=Length - 0.0994, ymax=Length + 0.0994),
  position=position_dodge(width=0.9)) +
  labs(title="Length of Cotton Varieties by Block", x="Variety", y="Length") +
  theme_bw()
```



Διάγραμμα 3-13: Ραβδόγραμμα μέσου μήκους ίνας

Η γραφική παράσταση δείχνει ότι υπάρχουν διαφορές στο μέσο μήκος μεταξύ των διαφορετικών ποικιλιών και περιοχών.

3.3 Ανάλυση δεδομένων με Ανάλυση συνδιακύμανσης (analysis of covariance)

3.3.1 Η Ανάλυση συνδιακύμανσης (ANCOVA)

Ένα γενικό γραμμικό μοντέλο με τουλάχιστον μία συνεχή και μία κατηγορική ανεξάρτητη μεταβλητή είναι γνωστό ως ANCOVA. Όταν υπάρχει επίδραση των μεταχειρίσεων υπάρχει υποψία επίδρασης μιας επιπλέον συνεχούς μεταβλητής στη μελέτη, η ανάλυση των αποτελεσμάτων θα υλοποιηθεί με τη χρήση της ανάλυσης ANCOVA (ανάλυση συνδιακύμανσης). Μια συμμεταβλητή είναι μια πρόσθετη συνεχή ανεξάρτητη μεταβλητή (πιθανόν συγχυτική). Λαμβάνοντας

υπόψιν την επιρροή της συμμεταβλητής, η ANCOVA αξιολογεί τις διαφορές μεταξύ των ομάδων σε μια κατηγορική ανεξάρτητη μεταβλητή (καταργώντας τη διακύμανση που σχετίζεται με τη συμμεταβλητή).

Επιλέγεται επίπεδο σημαντικότητας πριν τη συλλογή δεδομένων στο 0.05 (5%).

3.3.2 Για το ποιοτικό χαρακτηριστικό : MICRONAIRE

Τα δεδομένα

Τα ακόλουθα δεδομένα του παραδείγματος αποτελούνται από τον παράγοντα ποικιλία (X) (3 επίπεδα) και τις μετρήσεις λεπτότητας και ωριμότητας των ινών (micronaire). Οι 5 διαφορετικές χρονιές μετρήσεων αποτελούν τη συμμεταβλητή (COV). Το μοντέλο ANCOVA αναλύει την επίδραση των ποικιλιών βαμβακιού στην λεπτότητα και ωριμότητα της ίνας, ενώ ελέγχει την επίδραση της συμμεταβλητής. Με την ANCOVA, η επίδραση διαφορετικών ποικιλιών στην λεπτότητα και ωριμότητα της ίνας, μπορεί να αναλυθεί με ακρίβεια, ενώ ελέγχεται και η επίδραση ανά χρονιά.

Υποθέσεις της ANCOVA

- Μηδενική υπόθεση : Οι μέσοι όροι όλων των ποικιλιών είναι ίσοι
 $H_0 : \mu_1 = \mu_2 = \dots = \mu_p$
- Εναλλακτικά : Τουλάχιστον δύο διαφορετικές ποικιλίες έχουν διαφορετικό μέσο όρο
 $H_1 : \text{Τουλάχιστον δύο } \mu \text{ δεν είναι ίσα}$

Η ανάλυση

Παρουσιάζονται συνοπτικά στατιστικά στοιχεία με βάση την εξαρτημένη μεταβλητή και τη συμμεταβλητή. Αρχικά φορτώνουμε τα δεδομένα και τα αποθηκεύουμε στη μεταβλητή cotton.

```
summary(cotton)
```

```
str(cotton)
```

```
> summary(cotton)
  variety      Year      MIC
Length:15      Min.   :2017   Min.   :3.800
Class :character 1st Qu.:2018   1st Qu.:4.145
Mode  :character Median :2019   Median :4.230
              Mean  :2019   Mean  :4.237
              3rd Qu.:2020   3rd Qu.:4.385
              Max.   :2021   Max.   :4.490

> str(cotton)
'data.frame':  15 obs. of  3 variables:
 $ variety: chr  "ARMONIA" "ARMONIA" "ARMONIA" "ARMONIA" ...
 $ Year   : int  2017 2018 2019 2020 2021 2017 2018 2019 2020 2021 ...
 $ MIC    : num  4.09 3.8 4.23 4.23 4.18 4.15 4.13 4.14 4.26 4.49 ...
```

Τα δεδομένα αποτελούνται από 15 παρατηρήσεις 3 μεταβλητών. Η μεταβλητή 'Variety' (ποικιλία) είναι χαρακτήρας, ενώ οι μεταβλητές 'Year'(έτος) και 'MIC' (micronaire) είναι αριθμητικές.

Πραγματοποιούμε τη παρακάτω ανάλυση προκειμένου να κατανοηθούν καλύτερα οι τιμές micronaire των ποικιλιών βαμβακιού Armonia, Elpida και PRG9811 σε μια περίοδο πέντε ετών. Αναλύοντας τα δεδομένα, μπορούμε να αποκτήσουμε μια εικόνα για το πώς αυτές οι ποικιλίες έχουν αλλάξει με την πάροδο του χρόνου και εάν μπορούν να παρατηρηθούν οποιεσδήποτε τάσεις.

```
library(rstatix)
```

Το πακέτο gtsummary παρέχει έναν κομψό και ευέλικτο τρόπο δημιουργίας αναλυτικών και συνοπτικών πινάκων έτοιμοι για δημοσίευση χρησιμοποιώντας τη γλώσσα προγραμματισμού R. Το πακέτο gtsummary συνοψίζει σύνολα δεδομένων, μοντέλα παλινδρόμησης και πολλά άλλα, χρησιμοποιώντας λογικές προεπιλογές με εξαιρετικά προσαρμόσιμες δυνατότητες.

Το πακέτο groupby χρησιμοποιείται για τη συλλογή πανομοιότυπων δεδομένων σε ομάδες στο DataFrame και την εκτέλεση συγκεντρωτικών λειτουργιών στα ομαδοποιημένα δεδομένα.

```
cotton %>% group_by(Variety) %>% get_summary_stats(MIC, type="common")
```

```
cotton %>% group_by(Variety) %>% get_summary_stats(Year, type="common")
```

```
> library(rstatix)
> cotton %>% group_by(Variey) %>% get_summary_stats(MIC, type="common")
# A tibble: 3 x 11
  Variety variable   n  min  max median  iqr mean  sd  se  ci
<chr> <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1 ARMONIA MIC       5  3.8  4.23  4.18  0.14  4.11 0.18 0.081 0.224
2 ELPIDA MIC       5  4.13  4.49  4.15  0.12  4.23 0.152 0.068 0.189
3 PRG9811 MIC      5  4.28  4.4  4.39  0.02  4.37 0.051 0.023 0.063
> cotton %>% group_by(Variety) %>% get_summary_stats(Year, type="common")
# A tibble: 3 x 11
  Variety variable   n  min  max median  iqr mean  sd  se  ci
<chr> <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1 ARMONIA Year       5  2017 2021 2019    2 2019 1.58 0.707 1.96
2 ELPIDA Year       5  2017 2021 2019    2 2019 1.58 0.707 1.96
3 PRG9811 Year      5  2017 2021 2019    2 2019 1.58 0.707 1.96
```

οι τιμές micronaire για κάθε ποικιλία έχουν μέσο όρο 4.11, 4.23 και 4.37 αντίστοιχα, με τυπική απόκλιση 0.18, 0.15 και 0.05 αντίστοιχα. Από την ανάλυση βλέπουμε ότι οι τιμές micronaire για τις ποικιλίες βαμβακιού Armonia, Elpida και PRG9811 είναι σχετικά σταθερές κατά την πενταετία. Τέλος, οι τιμές micronaire για κάθε ποικιλία είναι εντός του ίδιου εύρους, με ελάχιστη απόκλιση μεταξύ τους.

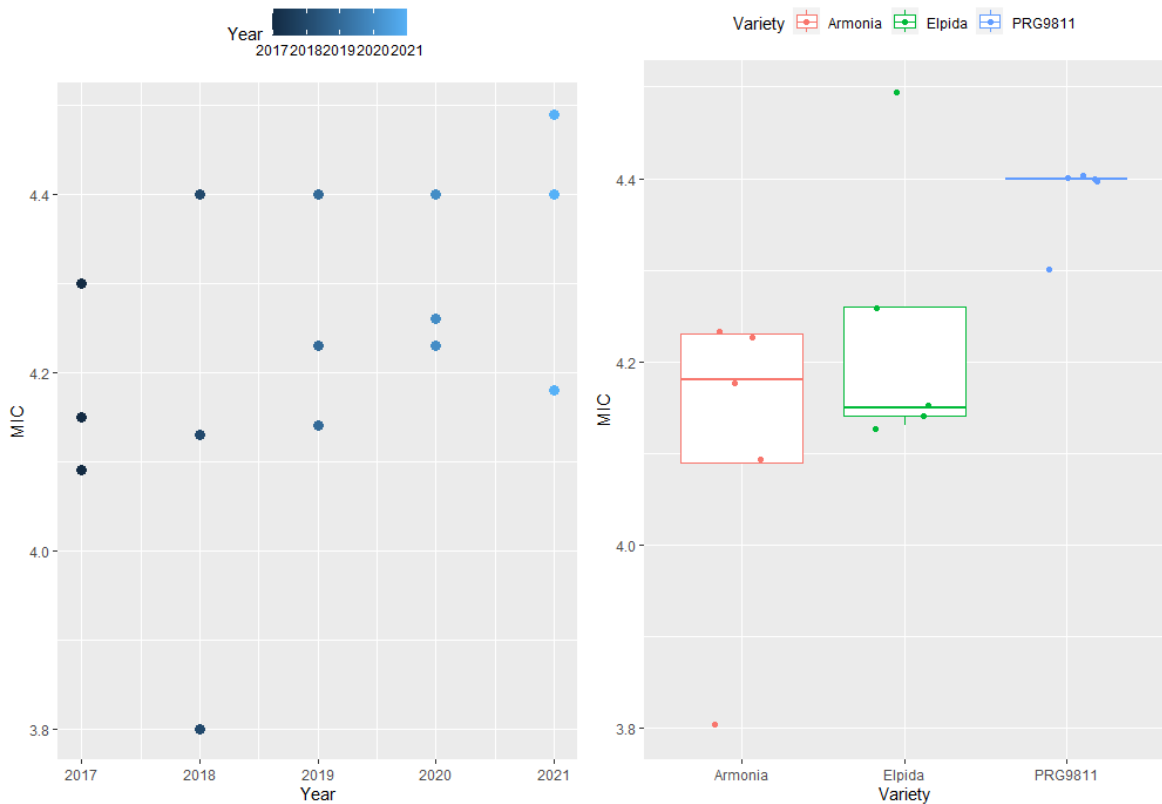
Πραγματοποιούμε ggplot για να απεικονιστούν οι τιμές micronaire των ποικιλιών βαμβακιού Armonia, Elpida και PRG9811 σε μια περίοδο πέντε ετών.

```
library(gridExtra)
```

```
library(ggplot2)
```

```
p1 <- ggplot(cotton, aes(Year, MIC, colour = Year)) + geom_point(size = 3) +
theme(legend.position="top")
```

```
p2 <- ggplot(cotton, aes(x = Variety, y = MIC, col = Variety)) + geom_boxplot(outlier.shape = NA) + geom_jitter(width = 0.2) + theme(legend.position="top")
grid.arrange(p1, p2, ncol=2)
```



Διάγραμμα 3-14: διαγράμματα δεδομένων τιμών micronaire

Είναι σαφές από το scatterplot ότι οι τιμές micronaire για την Armonia είναι συνολικά υψηλότερες από ό,τι για την Elpida και το PRG9811, ενώ το boxplot δείχνει ότι οι τιμές micronaire για την Elpida και το PRG9811 είναι πιο διασκορπισμένες από εκείνες για την Armonia.

Η παρακάτω ανάλυση πραγματοποιείται για να εκτιμηθούν οι διαφορές στις τιμές micronaire μεταξύ των ποικιλιών βαμβακιού Armonia, Elpida και PRG9811. Διεξάγοντας τη δοκιμή ANOVA, μπορούμε να προσδιορίσουμε το μέγεθος της διαφοράς στις τιμές micronaire μεταξύ των ποικιλιών και εάν κάποιος από τους παράγοντες (όπως το έτος) έχει επίδραση στις τιμές micronaire.

```
anova_test(data = cotton, formula = MIC ~ Year + Variety, type = 3, detailed = TRUE)
```

```
> anova_test(data = cotton, formula = MIC ~ Year + Variety, type = 3, detailed = TRUE)
Coefficient covariances computed by hccm()
ANOVA Table (type III tests)
```

Effect	SSn	SSd	DFn	DFd	F	p	p<.05	ges
1 (Intercept)	0.086	0.14	1	11	6.741	0.025	*	0.380
2 Year	0.093	0.14	1	11	7.279	0.021	*	0.398
3 Variety	0.174	0.14	2	11	6.824	0.012	*	0.554

Υπάρχει στατιστικά σημαντική διαφορά στις τιμές micronaire μεταξύ των ποικιλιών βαμβακιού Armonia, Elpida και PRG9811. Από τον πίνακα ANOVA παρατηρούμε ότι το μέγεθος του (ges) είναι 0.554, γεγονός που υποδηλώνει ότι η διακύμανση στις τιμές micronaire προκαλείται κυρίως από την ποικιλία του βαμβακιού. Επιπροσθέτως, το έτος είχε στατιστικά σημαντική επίδραση, με μέγεθος επίδρασης 0.398.

Πραγματοποιούμε μια δοκιμή EMMeans προκειμένου να διερευνήσουμε περαιτέρω τις διαφορές στις τιμές micronaire μεταξύ των ποικιλιών βαμβακιού Armonia, Elpida και PRG9811.

```
adj_means <- emmeans_test(data = cotton, formula = MIC ~ Variety, covariate = Year)
```

```
get_emmeans(adj_means)
```

```
> adj_means <- emmeans_test(data = cotton, formula = MIC ~ Variety, covariate = Year)
> get_emmeans(adj_means)
# A tibble: 3 x 8
  Year Variety emmean   se    df conf.low conf.high method
<dbl> <fct>   <dbl> <dbl> <dbl> <dbl>   <dbl> <chr>
1  2019 ARMONIA  4.11 0.0505  11    3.99    4.22 Emmeans test
2  2019 ELPIDA  4.23 0.0505  11    4.12    4.35 Emmeans test
3  2019 PRG9811 4.37 0.0505  11    4.26    4.48 Emmeans test
```

Οι προσαρμοσμένες μέσες τιμές για τις τιμές micronaire των ποικιλιών είναι 4.11, 4.23 και 4.37 αντίστοιχα, με τυπικό σφάλμα 0.0505. Σύμφωνα με τα αποτελέσματα οι τιμές micronaire για κάθε ποικιλία είναι εντός του ίδιου εύρους, με ελάχιστη απόκλιση μεταξύ τους. Επιπλέον, δεν υπάρχει στατιστικά σημαντική διαφορά στους προσαρμοσμένους μέσους όρους για τις τιμές micronaire μεταξύ των ποικιλιών ή μεταξύ των ετών.

Η ποικιλία PRG 9811 έχει την υψηλότερη απόδοση για τη χρονιά 2019 (4.37).

Για να γνωρίζουμε τις ποικιλίες με στατιστικά σημαντικές διαφορές, θα πραγματοποιήσουμε το post-hoc τεστ με τη μέθοδο για τον έλεγχο πολλαπλών υποθέσεων.

Η κλασική διόρθωση Sidak επιστρέφει προσαρμοσμένες τιμές p, διασφαλίζοντας ισχυρό έλεγχο ρυθμού σφάλματος family wise (FWER) υπό την υπόθεση της ανεξαρτησίας των τιμών p εισόδου.

Η μέθοδος Sidak χρησιμοποιεί την ελάχιστη τιμή p, αλλά τη διορθώνει για τον αριθμό των τιμών p που συγκεντρώνονται. Η διαδικασία είναι γενικότερα έγκυρη για θετικές εξαρτώμενες στατιστικές δοκιμών. Συνιστάται η χρήση της αναβαθμισμένης έκδοσης της διόρθωσης SidakSD, η οποία ισχύει με τις ίδιες ακριβώς υποθέσεις, αλλά είναι πιο ισχυρή.

```
emmeans_test(data = cotton, formula = MIC ~ Variety, covariate = Year, p.adjust.method = "sidak")
```

```
> emmeans_test(data = cotton, formula = MIC ~ Variety, covariate = Year, p.adjust.method = "sidak")
# A tibble: 3 x 9
  term      .y. group1 group2    df statistic      p p.adj p.adj.signif
* <chr>    <chr> <chr>  <chr>  <dbl> <dbl> <dbl> <dbl> <chr>
1 Year*Variety MIC  ARMONIA ELPIDA    11   -1.79 0.101 0.273 ns
2 Year*Variety MIC  ARMONIA PRG9811  11   -3.69 0.00354 0.0106 *
3 Year*Variety MIC  ELPIDA  PRG9811  11   -1.90 0.0835 0.230 ns
```

Υπάρχει μια στατιστικά σημαντική διαφορά στις τιμές micronaire μεταξύ της Armonia και του PRG9811 ($p = 0.0106$, $p.adj = 0.0106$). Επιπλέον, δεν υπάρχει στατιστικά σημαντική διαφορά μεταξύ Armonia και Elpida ή μεταξύ Elpida και PRG9811 ($p > 0.05$). Η Armonia έχει σημαντικά χαμηλότερη τιμή micronaire από την PRG9811, ενώ οι άλλες δύο ποικιλίες έχουν παρόμοιες τιμές.

Δοκιμή υποθέσεων ANCOVA

Υποθέσεις κανονικότητας

Τα υπολείμματα θα πρέπει να είναι περίπου κανονικά κατανομημένα. Η δοκιμή Shapiro-Wilk μπορεί να χρησιμοποιηθεί για τον έλεγχο της κανονικής κατανομής των υπολειμμάτων. *Μηδενική υπόθεση* : τα δεδομένα προέρχονται από μια κανονική κατανομή.

```
shapiro.test(resid(aov(MIC ~ Variety + Year, data = cotton)))
```

```
> shapiro.test(resid(aov(MIC ~ Variety + Year, data = cotton)))

      Shapiro-Wilk normality test

data:  resid(aov(MIC ~ Variety + Year, data = cotton))
W = 0.94432, p-value = 0.4398
```

Σύμφωνα με τη δοκιμή, τα υπολείμματα κατανέμονται κανονικά, με στατιστική W 0.94432 και τιμή p 0.4398. Ως εκ τούτου, πληρούνται οι παραδοχές του μοντέλου ANOVA και ότι το μοντέλο είναι ένα κατάλληλο μοντέλο για τα δεδομένα.

Υποθέσεις ομοιογένειας διακυμάνσεων

Η διακύμανση πρέπει να είναι παρόμοια για όλες τις ποικιλίες. Το τεστ Bartlett μπορεί να χρησιμοποιηθεί για τον έλεγχο της ομοιογένειας των διακυμάνσεων.

Μηδενική υπόθεση: δείγματα από πληθυσμούς έχουν ίσες διακυμάνσεις.

```
bartlett.test(MIC ~ Variety, data = cotton)
```

```
> bartlett.test(MIC ~ Variety, data = cotton)

Bartlett test of homogeneity of variances

data: MIC by Variety
Bartlett's K-squared = 4.7554, df = 2, p-value = 0.09277
```

Αφού η τιμή p είναι μεγαλύτερη από 0.05 οι διακυμάνσεις των δύο μεταβλητών είναι ομοιογενείς και οι παραδοχές του μοντέλου ANOVA είναι πιθανό να πληρούνται.

Υποθέσεις ομοιογένειας (συντελεστές συμμεταβλητών)

Αυτή είναι μια σημαντική υπόθεση στην ANCOVA. Δεν πρέπει να υπάρχει αλληλεπίδραση μεταξύ της κατηγορικής ανεξάρτητης μεταβλητής και της συμμεταβλητής. Αυτό μπορεί να ελεγχθεί χρησιμοποιώντας όρους αλληλεπίδρασης μεταξύ ποικιλίας και χρονιάς σε anova. Εάν παραβιαστεί αυτή η υπόθεση, το αποτέλεσμα της μεταχείρισης δεν θα είναι το ίδιο σε διάφορα επίπεδα της συμμεταβλητής.

Anova(aov(MIC ~ Variety * Year, data = cotton), type = 3)

```
> Anova(aov(MIC ~ Variety * Year, data = cotton), type = 3)
Anova Table (Type III tests)

Response: MIC
      Sum Sq Df F value Pr(>F)
(Intercept)  0.034770  1  2.5161 0.1471
Variety      0.016155  2  0.5845 0.5772
Year        0.037210  1  2.6927 0.1352
Variety:Year 0.016107  2  0.5828 0.5781
Residuals   0.124370  9
```

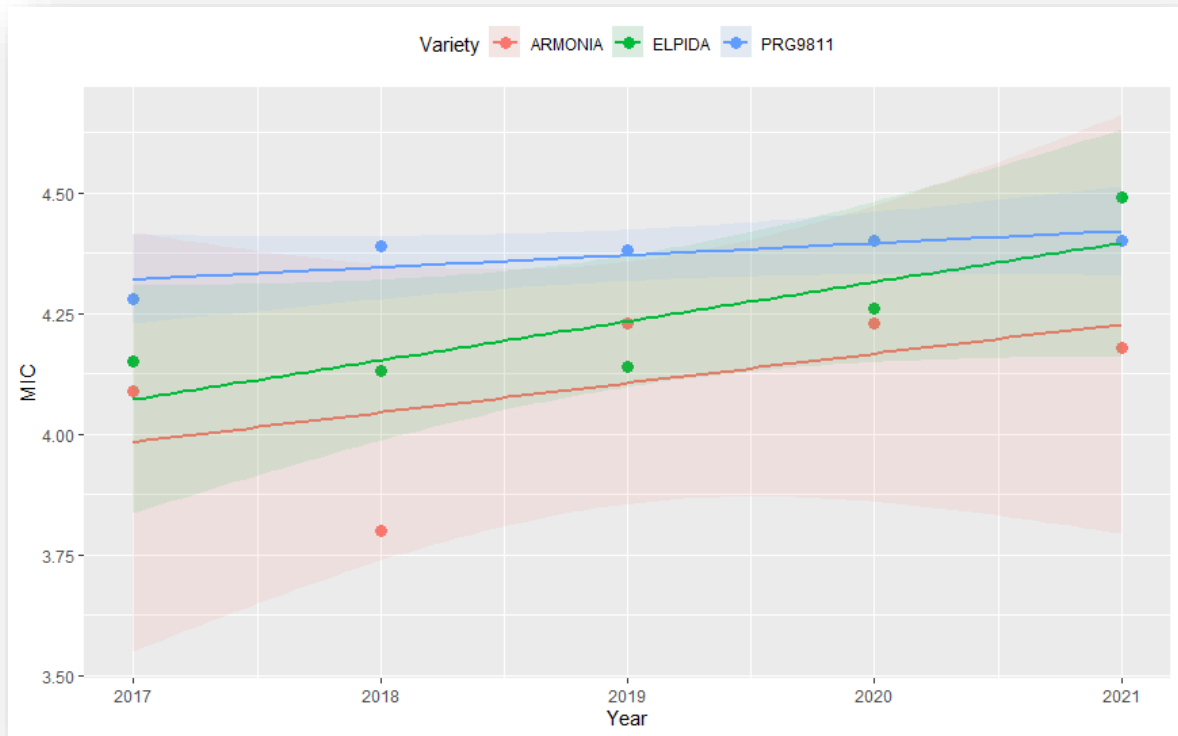
Καθώς η τιμή p είναι για την αλληλεπίδραση (χρονιά και ποικιλία) δεν είναι σημαντική ($p > 0.05$), δεν υπάρχει αλληλεπίδραση μεταξύ χρονιάς και ποικιλίας.

Υποθέσεις γραμμικότητας

Η σχέση μεταξύ της συμμεταβλητής σε κάθε ομάδα της ανεξάρτητης μεταβλητής πρέπει να είναι γραμμική. Το διάγραμμα διασποράς της συμμεταβλητής και της εξαρτημένης μεταβλητής σε κάθε ομάδα της ανεξάρτητης μεταβλητής μπορεί να χρησιμοποιηθεί για την αξιολόγηση αυτής της υπόθεσης. Τα σημεία δεδομένων πρέπει να βρίσκονται στην ευθεία γραμμή για να πληρούν την υπόθεση γραμμικότητας.

Παρακάτω παρουσιάζεται μια απεικόνιση του συνόλου δεδομένων βαμβακιού που δείχνει τη σχέση μεταξύ της μεταβλητής απόκρισης *micronaire* και των δύο παραγόντων, Ποικιλία και Έτος. Τα χρώματα διευκολύνουν τη διάκριση των διαφορετικών Ποικιλιών και οι γραμμές παλινδρόμησης δείχνουν πώς η μεταβλητή απόκρισης επηρεάζεται από τους δύο παράγοντες.

```
ggplot(cotton, aes(Year, MIC, colour = Variety)) + geom_point(size = 3) +  
geom_smooth(method = "lm", aes(fill = Variety), alpha = 0.1) +  
theme(legend.position="top")
```



Διάγραμμα 3-15: Διάγραμμα διασποράς

Αυτή η γραφική παράσταση δείχνει ένα διάγραμμα διασποράς του Έτους σε σχέση με το *micronaire*, με κάθε ποικιλία να αντιπροσωπεύεται από διαφορετικό χρώμα και μια γραμμή γραμμικής παλινδρόμησης για κάθε ποικιλία. Από τις γραμμές παλινδρόμησης οι ποικιλίες *Elpida* και *Armonia* εμφανίζουν θετική γραμμική συσχέτιση κ η ποικιλία *PRG9811* σχεδόν μηδενική γραμμική συσχέτιση.

Σε γενικές γραμμές η ποικιλία *PRG9811* εμφανίζει τις υψηλότερες τιμές *micronaire* εκτός του έτους 2021 που την υψηλότερη τιμή εμφανίζει η ποικιλία *Elpida*. Η *Elpida* εμφανίζει υψηλότερες τιμές σε σχέση με την *Armonia* εκτός του έτους 2019. Τις μεγαλύτερες διαφοροποιήσεις τιμών *micronaire* διακρίνουμε το έτος 2018 ανάμεσα στις 3 ποικιλίες. Τέλος παρατηρούμε ότι η ποικιλία *Elpida* εμφανίζει ανοδικές τιμές ανά τα έτη, ενώ για τη ποικιλία *Armonia* δε μπορούμε να βγάλουμε ανάλογα συμπεράσματα.

3.3.3 Για το ποιοτικό χαρακτηριστικό : Αντοχή της ίνας Τα δεδομένα

Τα ακόλουθα δεδομένα του παραδείγματος αποτελούνται από τον παράγοντα ποικιλία (X) (3 επίπεδα) και τις μετρήσεις αντοχής των ινών (Strength). Οι 5 διαφορετικές χρονιές μετρήσεων αποτελούν τη συμμεταβλητή (COV). Το μοντέλο ANCOVA αναλύει την επίδραση των ποικιλιών βαμβακιού στην αντοχή της ίνας, ενώ ελέγχει την επίδραση της συμμεταβλητής. Με την ANCOVA, η επίδραση διαφορετικών ποικιλιών στην αντοχή της ίνας, μπορεί να αναλυθεί με ακρίβεια, ενώ ελέγχεται και η επίδραση ανά χρονιά.

Υποθέσεις της ANCOVA

- Μηδενική υπόθεση : Οι μέσοι όροι όλων των ποικιλιών είναι ίσοι
 $H_0 : \mu_1 = \mu_2 = \dots = \mu_p$
- Εναλλακτικά : Τουλάχιστον δύο διαφορετικές ποικιλίες έχουν διαφορετικό μέσο όρο
 $H_1 : \text{Τουλάχιστον δύο } \mu \text{ δεν είναι ίσα}$

Η ανάλυση

Παρουσιάζονται συνοπτικά στατιστικά στοιχεία με βάση την εξαρτημένη μεταβλητή και τη συμμεταβλητή. Αρχικά φορτώνουμε τα δεδομένα και τα αποθηκεύουμε στη μεταβλητή cotton.

```
summary(cotton)
```

```
str(cotton)
```

```
> cotton<-anc_dataa
> summary(cotton)
  Variety      Year      Strength
Length:15      Min.   :2017      Min.   :28.80
Class :character 1st Qu.:2018      1st Qu.:30.00
Mode  :character Median :2019      Median :30.80
              Mean  :2019      Mean   :30.67
              3rd Qu.:2020      3rd Qu.:31.40
              Max.  :2021      Max.   :32.60

> str(cotton)
'data.frame': 15 obs. of 3 variables:
 $ Variety : chr  "Armonia" "Armonia" "Armonia" "Armonia" ...
 $ Year    : int  2017 2018 2019 2020 2021 2017 2018 2019 2020 2021 ...
 $ Strength: num  29.5 30.1 29.1 30.8 32.6 29.9 30.2 28.8 31.8 31.7 ...
> |
```

Σύμφωνα με τον παραπάνω πίνακα υπάρχουν 15 παρατηρήσεις 3 μεταβλητών στο σύνολο δεδομένων. Η μεταβλητή "Variety" (ποικιλία) είναι μια μεταβλητή χαρακτήρων, ενώ η "Year" (έτος) και η "Strength" (αντοχή της ίνας) είναι αριθμητικές μεταβλητές.

```
library(rstatix)
```

```
cotton %>% group_by(Variety) %>% get_summary_stats(STRENGTH, type="common")
```

```
cotton %>% group_by(Variety) %>% get_summary_stats(Year, type="common")
```

```
> library(rstatix)
> cotton %>% group_by(Variety) %>% get_summary_stats(Strength, type="common")
# A tibble: 3 x 11
  Variety variable     n  min  max median  iqr  mean  sd  se  ci
<chr>   <chr>   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1 Armonia Strength     5  29.1  32.6  30.1  1.3  30.4  1.38  0.616  1.71
2 Elpida  Strength     5  28.8  31.8  30.2  1.8  30.5  1.27  0.569  1.58
3 PRG9811 Strength     5  30.5  32.3  30.9  0.3  31.1  0.694  0.31  0.862
> cotton %>% group_by(Variety) %>% get_summary_stats(Year, type="common")
# A tibble: 3 x 11
  Variety variable     n  min  max median  iqr  mean  sd  se  ci
<chr>   <chr>   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1 Armonia Year         5  2017  2021  2019    2  2019  1.58  0.707  1.96
2 Elpida  Year         5  2017  2021  2019    2  2019  1.58  0.707  1.96
3 PRG9811 Year         5  2017  2021  2019    2  2019  1.58  0.707  1.96
```

Τα συνοπτικά στατιστικά στοιχεία για την αντοχή δείχνουν ότι η μέση αντοχή για τις τρεις ποικιλίες είναι μεταξύ 30.4 και 31.1, με την ποικιλία Armonia να έχει την υψηλότερη μέση αντοχή.

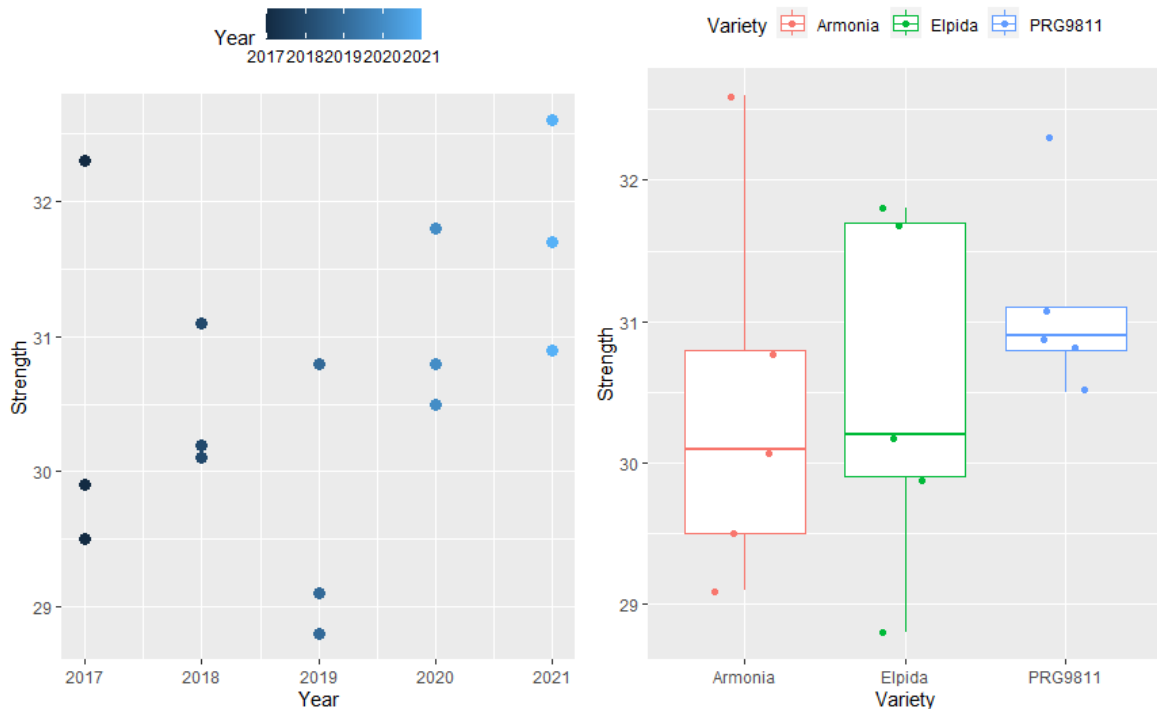
```
library(gridExtra)
```

```
library(ggplot2)
```

```
p1 <- ggplot(cotton, aes(Year, Strength, colour = Year)) + geom_point(size = 3) +
  theme(legend.position="top")
```

```
p2 <- ggplot(cotton, aes(x = Variety, y = Strength, col = Variety)) +
  geom_boxplot(outlier.shape = NA) + geom_jitter(width = 0.2) +
  theme(legend.position="top")
```

```
grid.arrange(p1, p2, ncol=2)
```



Διάγραμμα 3-16: Οπτικοποίηση δεδομένων αντοχής ίνας

Οπτικοποίηση δεδομένων

Αυτή η γραφική παράσταση παρέχει μια οπτική αναπαράσταση των δεδομένων.

Το γράφημα στα αριστερά δείχνει ένα διάγραμμα διασποράς του Έτους με την αντοχή της ίνας, με κάθε έτος να αντιπροσωπεύεται από διαφορετικό χρώμα. Η αντοχή ποικίλλει μεταξύ ετών, με υψηλότερες τιμές αντοχής το 2021 από ό,τι το 2017.

Το boxplot της αντοχής για κάθε Ποικιλία δείχνει ότι η αντοχή ποικίλλει μεταξύ των ποικιλιών, με την Αρμονία να έχει την υψηλότερη μέση αντοχή και τη μεγαλύτερη διακύμανση στην αντοχή.

```
anova_test(data = cotton, formula = Strength ~ Year + Variety, type = 3, detailed = TRUE)
```

```
> anova_test(data = cotton, formula = Strength ~ Year + Variety, type = 3, detailed = TRUE)
Coefficient covariances computed by hccm()
ANOVA Table (type III tests)
```

	Effect	SSn	SSd	DFn	DFd	F	p	p<.05	ges
1	(Intercept)	2.266	13.461	1	11	1.851	0.201		0.144
2	Year	2.523	13.461	1	11	2.062	0.179		0.158
3	Variety	1.505	13.461	2	11	0.615	0.558		0.101

Τα αποτελέσματα δείχνουν ότι η επίδραση του Έτους στη μεταβλητή Strength είναι στατιστικά σημαντική ($p = 0.179$), ενώ η επίδραση του Variety όχι ($p = 0.558$). Συμπεραίνουμε ότι η αντοχή της ίνας επηρεάζεται από το Έτος, αλλά όχι από την Ποικιλία.

```
adj_means <- emmeans_test(data = cotton, formula = Strength ~ Variety, covariate = Year)
```

```
get_emmeans(adj_means)
```

```
> adj_means <- emmeans_test(data = cotton, formula = Strength ~ Variety, covariate = Year)
> get_emmeans(adj_means)
# A tibble: 3 x 8
  Year variety emmean se df conf.low conf.high method
<dbl> <fct> <dbl> <dbl> <dbl> <dbl> <dbl> <chr>
1 2019 Armonia 30.4 0.495 11 29.3 31.5 Emmeans test
2 2019 Elpida 30.5 0.495 11 29.4 31.6 Emmeans test
3 2019 PRG9811 31.1 0.495 11 30.0 32.2 Emmeans test
```

Σύμφωνα με τον πίνακα το 2019, η μέση αντοχή της ίνας για την Αρμονία είναι 30.4, για την Ελπίδα είναι 30.5 και για την PRG9811 είναι 31.1. Αρα, η αντοχή επηρεάζεται από το Έτος, αλλά όχι από την Ποικιλία.

Η ποικιλία PRG 9811 έχει την υψηλότερη αντοχή της ίνας για τη χρονιά 2019 (31.1).

Το Emmeans δίνει τον εκτιμώμενο οριακό μέσο όρο (EMMs) που είναι επίσης γνωστός ως μέσος όρος ελαχίστων τετραγώνων. Τα EMM είναι προσαρμοσμένα για κάθε ποικιλία.

Για να γνωρίζουμε τις ποικιλίες με στατιστικά σημαντικές διαφορές, πραγματοποιούμε το post-hoc τεστ με τη μέθοδο για τον έλεγχο πολλαπλών υποθέσεων σε διάστημα 5%.

```
emmeans_test(data = cotton, formula = Strength ~ Variety, covariate = Year,
p.adjust.method = "sidak")
```

```
> emmeans_test(data = cotton, formula = Strength ~ Variety, covariate = Year, p.adjust.method = "sidak")
# A tibble: 3 x 9
  term          .y.      group1 group2      df statistic      p p.adj p.adj.signif
* <chr>         <chr>    <chr>  <chr>    <dbl> <dbl> <dbl> <dbl> <chr>
1 Year*Variety Strength Armonia Elpida     11  -0.0858  0.933  1.00    ns
2 Year*Variety Strength Armonia PRG9811    11  -1.00    0.339  0.711    ns
3 Year*Variety Strength Elpida  PRG9811    11  -0.915   0.380  0.762    ns
```

Παρατηρούμε ότι δεν υπάρχει στατιστικά σημαντική διαφορά Αντοχής μεταξύ των ποικιλιών ($p = 0.762$). Ως εκ τούτου, η αντοχή δεν επηρεάζεται από την Ποικιλία.

Δοκιμή υποθέσεων ANCOVA

Υποθέσεις κανονικότητας

Τα υπολείμματα θα πρέπει να είναι περίπου κανονικά κατανομημένα. Η δοκιμή Shapiro-Wilk μπορεί να χρησιμοποιηθεί για τον έλεγχο της κανονικής κατανομής των υπολειμμάτων. *Μηδενική υπόθεση* : τα δεδομένα προέρχονται από μια κανονική κατανομή.

```
shapiro.test(resid(aov(Strength ~ Variety + Year, data = cotton)))
```

```
> shapiro.test(resid(aov(Strength ~ Variety + Year, data = cotton)))  
  
      shapiro-wilk normality test  
  
data:  resid(aov(Strength ~ Variety + Year, data = cotton))  
W = 0.96912, p-value = 0.8447
```

Τα αποτελέσματα δείχνουν ότι τα υπολείμματα δεν κατανέμονται κανονικά ($p\text{-value}=0.8447$). Αυτό υποδηλώνει ότι το μοντέλο ANOVA μπορεί να μην είναι κατάλληλο για τα δεδομένα.

Υποθέσεις ομοιογένειας διακυμάνσεων

Η διακύμανση πρέπει να είναι παρόμοια για όλες τις ποικιλίες. Το τεστ Bartlett μπορεί να χρησιμοποιηθεί για τον έλεγχο της ομοιογένειας των διακυμάνσεων.

Μηδενική υπόθεση : δείγματα από πληθυσμούς έχουν ίσες διακυμάνσεις.

```
bartlett.test(Strength ~ Variety, data = cotton)
```

```
> bartlett.test(Strength ~ Variety, data = cotton)  
  
      Bartlett test of homogeneity of variances  
  
data:  Strength by Variety  
Bartlett's K-squared = 1.6884, df = 2, p-value = 0.4299
```

Οι διακυμάνσεις της αντοχής για τις τρεις ποικιλίες δεν διαφέρουν σημαντικά μεταξύ τους ($p\text{-value} = 0,4299$). Άρα το μοντέλο ANOVA μπορεί να είναι κατάλληλο για τα δεδομένα.

Υποθέσεις ομοιογένειας (συντελεστές συμμεταβλητών)

Αυτή είναι μια σημαντική υπόθεση στην ANCOVA. Δεν πρέπει να υπάρχει αλληλεπίδραση μεταξύ της κατηγορικής ανεξάρτητης μεταβλητής και της συμμεταβλητής. Αυτό μπορεί να ελεγχθεί χρησιμοποιώντας όρους αλληλεπίδρασης μεταξύ ποικιλίας και χρονιάς σε `aov`. Εάν παραβιαστεί αυτή η υπόθεση, το αποτέλεσμα της μεταχείρισης δεν θα είναι το ίδιο σε διάφορα επίπεδα της συμμεταβλητής.

```
Anova(aov(Strength ~ Variety * Year, data = cotton), type = 3)
```

```
> Anova(aov(Strength ~ Variety * Year, data = cotton), type = 3)
Anova Table (Type III tests)

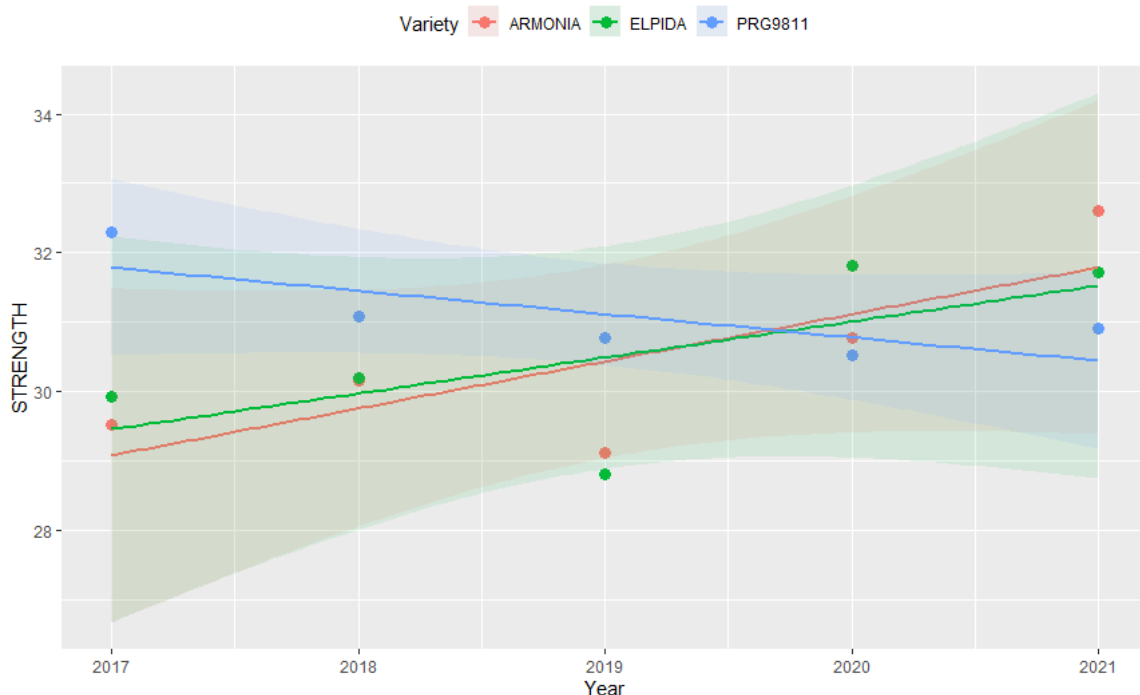
Response: Strength
      Sum Sq Df F value    Pr(>F)
(Intercept)  4.5553  1  5.5681 0.04262 *
Variety      6.1022  2  3.7295 0.06611 .
Year        4.7610  1  5.8195 0.03910 *
Variety:Year  6.0980  2  3.7269 0.06621 .
Residuals    7.3630  9
---
signif. codes:  0 '****' 0.001 '***' 0.01 '**' 0.05 '.' 0.1 ' ' 1
```

Η επίδραση του Έτους στη μεταβλητή Strength είναι στατιστικά σημαντική ($p = 0.03910$), ενώ τα αποτελέσματα της ποικιλίας και της ποικιλίας σε συνάρτηση με το έτος όχι ($p = 0.06611$ και 0.06621 , αντίστοιχα). Αρα, η αντοχή επηρεάζεται από το Έτος, αλλά όχι από την Ποικιλία ή την αλληλεπίδραση μεταξύ Ποικιλίας και Έτους.

Υποθέσεις γραμμικότητας

Η σχέση μεταξύ της συµμεταβλητής σε κάθε ομάδα της ανεξάρτητης μεταβλητής πρέπει να είναι γραμμική. Το διάγραμμα διασποράς της συµμεταβλητής και της εξαρτηµένης μεταβλητής σε κάθε ομάδα της ανεξάρτητης μεταβλητής μπορεί να χρησιμοποιηθεί για την αξιολόγηση αυτής της υπόθεσης. Τα σηµεία δεδοµένων πρέπει να βρίσκονται στην ευθεία γραµµή για να πληρούν την υπόθεση γραμμικότητας.

```
ggplot(cotton, aes(Year, Strength, colour = Variety)) + geom_point(size = 3) +
geom_smooth(method = "lm", aes(fill = Variety), alpha = 0.1) +
theme(legend.position="top")
```



Διάγραμμα 3-17: διάγραμμα διασποράς αντοχής ίνας

Αυτή η γραφική παράσταση δείχνει ένα διάγραμμα διασποράς του Έτους σε σχέση με την αντοχή της ίνας, με κάθε ποικιλία να αντιπροσωπεύεται από διαφορετικό χρώμα και μια γραμμή γραμμικής παλινδρόμησης για κάθε ποικιλία. Παρατηρούμε ότι υπάρχει θετική σχέση μεταξύ του έτους και της αντοχής και ότι η αντοχή των δύο από τις τρεις ποικιλίες (Elpida, Armonia) αυξάνεται με παρόμοιο ρυθμό με την πάροδο του χρόνου, ενώ για τη ποικιλία PRG9811 έχουμε αρνητική συσχέτιση.

Το έτος 2019 παρατηρούμε χαμηλές τιμές και για τις 3 ποικιλίες χωρίς ωστόσο αυτό να επηρεάζει τη γενική εικόνα που έχουμε για το σύνολο των τιμών όλων των ετών.

3.3.4 Για το ποιοτικό χαρακτηριστικό : Μήκος ίνας

Τα δεδομένα

Τα ακόλουθα δεδομένα του παραδείγματος αποτελούνται από τον παράγοντα ποικιλία (X) (3 επίπεδα) και τις μετρήσεις μήκους των ινών (Length). Οι 5 διαφορετικές χρονιές μετρήσεων αποτελούν τη συμμεταβλητή (COV). Το μοντέλο ANCOVA αναλύει την επίδραση των ποικιλιών βαμβακιού στην αντοχή της ίνας, ενώ ελέγχει την επίδραση της συμμεταβλητής. Με την ANCOVA, η επίδραση διαφορετικών ποικιλιών στην αντοχή της ίνας, μπορεί να αναλυθεί με ακρίβεια, ενώ ελέγχεται και η επίδραση ανά χρονιά.

Υποθέσεις της ANCOVA

- Μηδενική υπόθεση : Οι μέσοι όροι όλων των ποικιλιών είναι ίσοι
 $H_0 : \mu_1 = \mu_2 = \dots = \mu_p$
- Εναλλακτικά : Τουλάχιστον δύο διαφορετικές ποικιλίες έχουν διαφορετικό μέσο όρο
 $H_1 : \text{Τουλάχιστον δύο } \mu \text{ δεν είναι ίσα}$

Η ανάλυση

Παρουσιάζονται συνοπτικά στατιστικά στοιχεία με βάση την εξαρτημένη μεταβλητή και τη συμμεταβλητή. Αρχικά φορτώνουμε τα δεδομένα και τα αποθηκεύουμε στη μεταβλητή cotton.

```
summary(cotton)
```

```
str(cotton)
```

```
> summary(cotton)
  Variety      Year      LENGTH
Length:15      Min.   :2017   Min.   :27.14
Class :character 1st Qu.:2018   1st Qu.:28.53
Mode  :character Median :2019   Median :30.17
              Mean  :2019   Mean  :29.65
              3rd Qu.:2020   3rd Qu.:30.60
              Max.  :2021   Max.  :32.33

> str(cotton)
'data.frame':   15 obs. of  3 variables:
 $ Variety: chr  "ARMONIA" "ARMONIA" "ARMONIA" "ARMONIA" ...
 $ Year   : int  2017 2018 2019 2020 2021 2017 2018 2019 2020 2021 ...
 $ LENGTH: num  30.9 30.4 30.5 30.7 32.3 ...
```

Η Ποικιλία είναι μεταβλητή χαρακτήρων, ενώ οι μεταβλητές Έτος και Μήκος ίνας είναι και οι δύο αριθμητικές μεταβλητές. Επιπλέον, φαίνεται ότι οι ελάχιστες και μέγιστες τιμές της μεταβλητής «LENGTH» είναι 27.14 και 32.33 αντίστοιχα, ενώ οι μέσες και διάμεσες τιμές είναι 29.65 και 30.17 αντίστοιχα.

```
library(rstatix)
```

```
cotton %>% group_by(Variety) %>% get_summary_stats(LENGTH,
type="common")
cotton %>% group_by(Variety) %>% get_summary_stats(Year,
type="common")
```



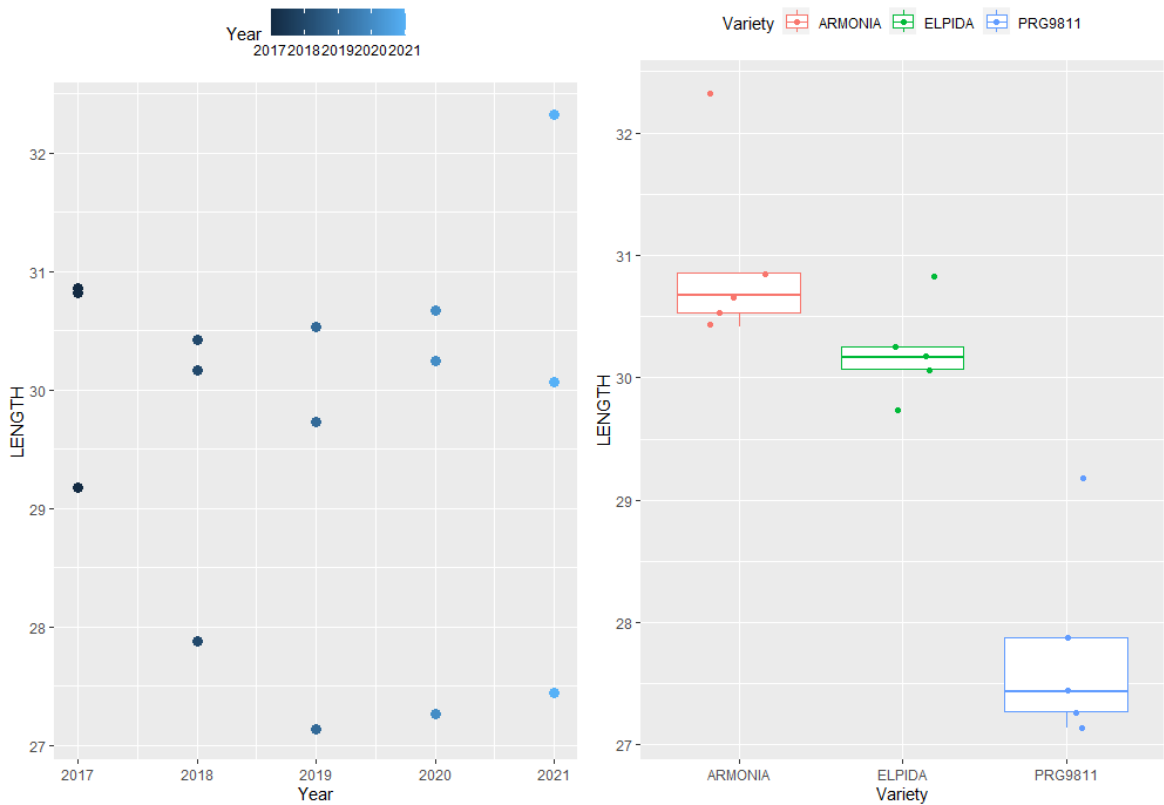
```
> library(rstatix)
> cotton %>% group_by(Variety) %>% get_summary_stats(LENGTH, type="common")
# A tibble: 3 x 11
  Variety variable     n  min  max median  iqr mean  sd  se  ci
  <chr>   <chr>   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1 ARMONIA LENGTH     5  30.4  32.3  30.7  0.33  31.0  0.782  0.35  0.971
2 ELPIDA  LENGTH     5  29.7  30.8  30.2  0.18  30.2  0.395  0.177  0.491
3 PRG9811 LENGTH     5  27.1  29.2  27.4  0.61  27.8  0.83  0.371  1.03
> cotton %>% group_by(Variety) %>% get_summary_stats(Year, type="common")
# A tibble: 3 x 11
  Variety variable     n  min  max median  iqr mean  sd  se  ci
  <chr>   <chr>   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1 ARMONIA Year     5  2017  2021  2019    2  2019  1.58  0.707  1.96
2 ELPIDA  Year     5  2017  2021  2019    2  2019  1.58  0.707  1.96
3 PRG9811 Year     5  2017  2021  2019    2  2019  1.58  0.707  1.96
```

```
library(gridExtra)
library(ggplot2)
```

```
p1 <- ggplot(cotton, aes(Year, LENGTH, colour = Year)) + geom_point(size = 3) +
  theme(legend.position="top")
```

```
p2 <- ggplot(cotton, aes(x = Variety, y = LENGTH, col = Variety)) +
  geom_boxplot(outlier.shape = NA) + geom_jitter(width = 0.2) +
  theme(legend.position="top")
```

```
grid.arrange(p1, p2, ncol=2)
```



Διάγραμμα 3-18: Οπτικοποίηση δεδομένων

Οπτικοποίηση δεδομένων

Αυτή η γραφική παράσταση εμφανίζει δύο διαφορετικά γραφήματα: μια γραφική παράσταση διασποράς του Έτους σε σχέση με το Μήκος ίνας και μια γραφική παράσταση boxplot με τη Ποικιλία σε σχέση με το Μήκος ίνας.

Από το scatterplot, φαίνεται ότι το μήκος των ινών βαμβακιού ποικίλλει ελαφρώς από έτος σε έτος, με τις υψηλότερες τιμές να εμφανίζονται το 2020 και το 2021.

Από το boxplot, παρατηρούμε ότι η διακύμανση στο μήκος μεταξύ των ποικιλιών είναι πολύ μεγαλύτερη από την διακύμανση μεταξύ ετών. Επιπλέον, η ποικιλία Αρμονία εμφανίζεται να έχει το μεγαλύτερο μέσο μήκος, με διάμεσο 31.30.

```
anova_test(data = cotton, formula = LENGTH ~ Year + Variety, type = 3, detailed = TRUE)
```

ANOVA Table (type III tests)								
	Effect	SSn	SSd	DFn	DFd	F	p <.05	ges
1	(Intercept)	0.254	5.649	1	11	0.495	4.96e-01	0.043
2	Year	0.179	5.649	1	11	0.349	5.66e-01	0.031
3	Variety	27.611	5.649	2	11	26.884	5.82e-05	* 0.830

Η επίδραση της ποικιλίας είναι η πιο σημαντική, με τιμή F 26.884 και τιμή p 5.82e-05. Αυτό μας οδηγεί στο συμπέρασμα ότι υπάρχει σημαντική διαφορά στο Μήκος της ίνας μεταξύ των διαφορετικών ποικιλιών βαμβακιού. Η επίδραση του Έτους δεν είναι σημαντική, με τιμή F 0.349 και τιμή p 0.566.

```
adj_means <- emmeans_test(data = cotton, formula = LENGTH ~ Variety, covariate = Year)
```

```
get_emmeans(adj_means)
```

```
> adj_means <- emmeans_test(data = cotton, formula = LENGTH ~ Variety, covariate = Year)
> get_emmeans(adj_means)
# A tibble: 3 x 8
  Year Variety emmean se df conf.low conf.high method
<dbl> <fct> <dbl> <dbl> <dbl> <dbl> <dbl> <chr>
1 2019 ARMONIA 31.0 0.320 11 30.3 31.7 Emmeans test
2 2019 ELPIDA 30.2 0.320 11 29.5 30.9 Emmeans test
3 2019 PRG9811 27.8 0.320 11 27.1 28.5 Emmeans test
```

Η ποικιλία ARMONIA έχει τον υψηλότερο προσαρμοσμένο μέσο όρο, στο 31.0, ενώ οι ποικιλίες ELPIDA και PRG9811 έχουν προσαρμοσμένο μέσο όρο 30.2 και 27.8 αντίστοιχα. Αυτό υποδηλώνει ότι η ποικιλία ARMONIA έχει το μεγαλύτερο μέσο μήκος ίνας, με τις ποικιλίες ELPIDA και PRG9811 να είναι ελαφρώς μικρότερες κατά μέσο όρο.

Το Emmeans δίνει τον εκτιμώμενο οριακό μέσο όρο (EMMs) που είναι επίσης γνωστός ως μέσος όρος ελαχίστων τετραγώνων. Τα EMM είναι προσαρμοσμένα για κάθε ποικιλία.

Για να γνωρίζουμε τις ποικιλίες με στατιστικά σημαντικές διαφορές, πραγματοποιούμε το post-hoc τεστ με τη μέθοδο για τον έλεγχο πολλαπλών υποθέσεων σε διάστημα 5%.

```
emmeans_test(data = cotton, formula = LENGTH ~ Variety, covariate = Year, p.adjust.method = "sidak")
```

```
> emmeans_test(data = cotton, formula = LENGTH ~ Variety, covariate = Year, p.adjust.method = "sidak")
# A tibble: 3 x 9
  term      .y.  group1 group2  df statistic      p    p.adj p.adj.signif
* <chr>    <chr> <chr>  <chr>  <dbl> <dbl> <dbl> <dbl> <chr>
1 Year*Variety LENGTH ARMONIA ELPIDA  11  1.66 0.124  0.329 ns
2 Year*Variety LENGTH ARMONIA PRG9811  11  7.02 0.0000222 0.0000666 ****
3 Year*Variety LENGTH ELPIDA PRG9811  11  5.35 0.000233 0.000698 ***
```

Δεν υπάρχει σημαντική διαφορά στο Μήκος ίνας μεταξύ των ARMONIA και ELPIDA ($p = 0.124$), αλλά υπάρχει σημαντική διαφορά μεταξύ των ARMONIA και PRG9811 ($p = 0.0000222$) και μεταξύ των ELPIDA και PRG9811 ($p = 0.000233$). Αρα, η ποικιλία ARMONIA έχει το υψηλότερο μέσο μήκος ίνας, με τις ποικιλίες ELPIDA και PRG9811 να είναι σημαντικά μικρότερες κατά μέσο όρο.

Δοκιμή υποθέσεων ANCOVA

Υποθέσεις κανονικότητας

Τα υπολείμματα θα πρέπει να είναι περίπου κανονικά καταμεμημένα. Η δοκιμή Shapiro-Wilk μπορεί να χρησιμοποιηθεί για τον έλεγχο της κανονικής κατανομής των υπολειμμάτων.

Μηδενική υπόθεση: τα δεδομένα προέρχονται από μια κανονική κατανομή.

```
shapiro.test(resid(aov(LENGTH ~ Variety + Year, data = cotton)))
```

```
> shapiro.test(resid(aov(LENGTH ~ Variety + Year, data = cotton)))

      Shapiro-Wilk normality test

data:  resid(aov(LENGTH ~ Variety + Year, data = cotton))
W = 0.82431, p-value = 0.007672
```

Τα αποτελέσματα της δοκιμής Shapiro-Wilk δείχνουν ότι τα υπολείμματα δεν κατανέμονται κανονικά, με τιμή $p = 0.007672$.

Υπάρχει περίπτωση δηλαδή να υπάρχουν κάποιοι άλλοι παράγοντες που επηρεάζουν τα δεδομένα που δεν καταγράφονται από το μοντέλο ANOVA. Αυτό μπορεί να περιλαμβάνει παράγοντες όπως ο τύπος του εδάφους, η άρδευση ή οι καιρικές συνθήκες.

Υποθέσεις ομοιογένειας διακυμάνσεων

Η διακύμανση πρέπει να είναι παρόμοια για όλες τις ποικιλίες. Το τεστ Bartlett μπορεί να χρησιμοποιηθεί για τον έλεγχο της ομοιογένειας των διακυμάνσεων.

Μηδενική υπόθεση: δείγματα από πληθυσμούς έχουν ίσες διακυμάνσεις.

```
bartlett.test(LENGTH ~ Variety, data = cotton)
```

```
> bartlett.test(LENGTH ~ Variety, data = cotton)
      Bartlett test of homogeneity of variances
data:  LENGTH by Variety
Bartlett's K-squared = 1.9915, df = 2, p-value = 0.3695
```

Οι διακυμάνσεις είναι ομοιογενείς, με τιμή $p = 0.3695$. Συνεπώς, δεν υπάρχει σημαντική διαφορά στο μέγεθος της διακύμανσης στη μεταβλητή μήκους ίνας μεταξύ των τριών ποικιλιών.

Υποθέσεις ομοιογένειας (συντελεστές συμμεταβλητών)

Αυτή είναι μια σημαντική υπόθεση στην ANCOVA. Δεν πρέπει να υπάρχει αλληλεπίδραση μεταξύ της κατηγορικής ανεξάρτητης μεταβλητής και της συμμεταβλητής. Αυτό μπορεί να ελεγχθεί χρησιμοποιώντας όρους αλληλεπίδρασης μεταξύ ποικιλίας και χρονιάς σε *anova*. Εάν παραβιαστεί αυτή η υπόθεση, το αποτέλεσμα της μεταχείρισης δεν θα είναι το ίδιο σε διάφορα επίπεδα της συμμεταβλητής.

```
Anova(aov(LENGTH ~ Variety * Year, data = cotton), type = 3)
```

```
> Anova(aov(LENGTH ~ Variety * Year, data = cotton), type = 3)
Anova Table (Type III tests)

Response: LENGTH
      Sum Sq Df F value Pr(>F)
(Intercept)  0.92212  1  2.8267 0.12701
Variety      2.70173  2  4.1410 0.05308 .
Year         1.01761  1  3.1194 0.11117
Variety:Year 2.71265  2  4.1577 0.05262 .
Residuals   2.93598  9
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

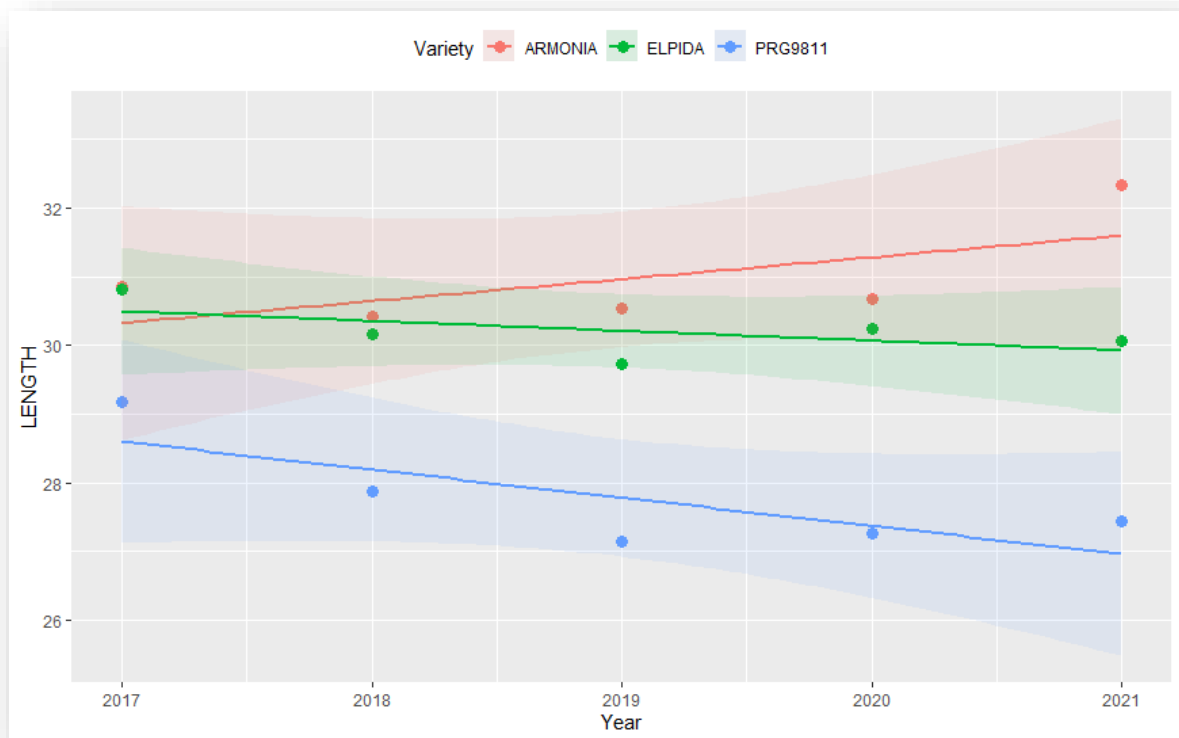
Από τον πίνακα βλέπουμε ότι τα αποτελέσματα της ποικιλίας και της αλληλεπίδρασης της ποικιλίας και του έτους, είναι και τα δύο σημαντικά, με τιμές $p = 0.05308$ και 0.05262 αντίστοιχα. Η επίδραση του Έτους δεν είναι σημαντική, με τιμή $p = 0.11117$.

Υποθέσεις γραμμικότητας

Η σχέση μεταξύ της συμμεταβλητής σε κάθε ομάδα της ανεξάρτητης μεταβλητής πρέπει να είναι γραμμική. Το διάγραμμα διασποράς της συμμεταβλητής και της εξαρτημένης μεταβλητής σε κάθε ομάδα της ανεξάρτητης μεταβλητής μπορεί να

χρησιμοποιηθεί για την αξιολόγηση αυτής της υπόθεσης. Τα σημεία δεδομένων πρέπει να βρίσκονται στην ευθεία γραμμή για να πληρούν την υπόθεση γραμμικότητας.

```
ggplot(cotton, aes(Year, LENGTH, colour = Variety)) + geom_point(size = 3) +  
geom_smooth(method = "lm", aes(fill = Variety), alpha = 0.1) +  
theme(legend.position="top")
```



Διάγραμμα 3-19: διάγραμμα διασποράς

Αυτή η γραφική παράσταση δείχνει ένα διάγραμμα διασποράς του έτους σε σχέση με το μήκος της ίνας, με κάθε ποικιλία να αντιπροσωπεύεται από διαφορετικό χρώμα και μια γραμμή γραμμικής παλινδρόμησης για κάθε ποικιλία. Παρατηρούμε ότι υπάρχει θετική σχέση μεταξύ του έτους και του μήκους της ίνας για τη ποικιλία Armonia, ενώ για τις ποικιλίες PRG9811 και Elpida έχουμε αρνητική συσχέτιση εμφανίζοντας χαμηλότερες τιμές με το πέρασμα των ετών.

Από το διάγραμμα, μπορούμε να διακρίνουμε ότι το μήκος των ινών βαμβακιού ποικίλλει ελαφρώς από έτος σε έτος, με τις υψηλότερες τιμές να εμφανίζονται το 2020 και το 2021. Επίσης βλέπουμε ότι η ποικιλία ARMONIA έχει το υψηλότερο μέσο μήκος ίνας, με τις ELPIDA και PRG9811 ποικιλίες να είναι ελαφρώς μικρότερες κατά μέσο όρο.

ΚΕΦΑΛΑΙΟ 4: ΣΤΑΤΙΣΤΙΚΗ ΑΝΑΛΥΣΗ ΠΟΙΟΤΙΚΩΝ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ ΙΝΑΣ ΒΑΜΒΑΚΙΟΥ ΜΕ ΒΑΣΗ ΤΟ ΠΡΑΓΜΑΤΙΚΟ ΠΛΑΝΟ ΔΙΕΞΑΓΩΓΗΣ ΕΡΕΥΝΑΣ

Ανάλυση δεδομένων

Σε αυτό το κεφάλαιο θα γίνει η στατιστική επεξεργασία και ανάλυση των συνολικών δεδομένων που συλλέχθηκαν και θα παρουσιαστούν τα αποτελέσματα σχετικά με τα ποιοτικά χαρακτηριστικά ενδιαφέροντος, τις ποικιλίες, τις περιοχές τις χρονιές από τις οποίες λάβαμε τα δείγματα καθώς και τις καιρικές συνθήκες (ανά περιοχή για κάθε χρονιά). Επίσης, θα εξεταστεί το ενδεχόμενο οι παράγοντες να είναι ανεξάρτητοι ή να αλληλεπιδρούν.

4.1 Για τη μεταβλητή: MICRONAIRE

Αναπτύχθηκε γραμμικό μοντέλο με εξαρτημένη μεταβλητή MIC (micronaire) και ανεξάρτητες μεταβλητές ποικιλία, έτος, μέση θερμοκρασία, βροχόπτωση, πόλη και ταχύτητα ανέμου. Το τελικό γραμμικό μοντέλο περιείχε τους παράγοντες ποικιλία, έτος και μέση θερμοκρασία και τις αντίστοιχες αλληλεπιδράσεις τους. Η επιλογή τους βασίστηκε στο κριτήριο AIC.

Βήματα για την επιλογή μοντέλου:

Αρχικά αναπτύχθηκε το πιο απλό μοντέλο χωρίς κάποιο παράγοντα:

```
> model_MIC_intercept<-lm(MIC~1,data=data)
```

Μετά συμπεριλάβαμε όλους τους πιθανούς παράγοντες χωρίς αλληλεπίδραση:

```
> model_MIC_full<-lm(MIC~., data=data)
```

Έγινε χρήση της συνάρτησης step() που επιστρέφει το καλύτερο δυνατό μοντέλο με βάση το κριτήριο AIC:

```
> step(model_MIC_intercept, direction='both', scope=formula(model_MIC_full), trace=0)
```

```
> lm(MIC~variety+year+mean_temp,data=data)
```

Στο επόμενο βήμα εισάγουμε στο μοντέλο μία-μία τις αλληλεπιδράσεις και ελέγχουμε τη σημαντικότητα των συντελεστών του μοντέλου. Το τελικό μοντέλο με τις αντίστοιχες αλληλεπιδράσεις παράχθηκε με τον παρακάτω κώδικα.

```
> lm(MIC~variety+year+mean_temp+variety*year +year*mean_temp, data=data)
```

Πίνακας 4-1: Το τελικό μοντέλο με τις αλληλεπιδράσεις.

```

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.392e+03  2.911e+02  -4.783 4.43e-06 ***
varietyELPIDA  2.038e+02  4.008e+01   5.084 1.20e-06 ***
varietyPRG 9811  3.144e+02  4.008e+01   7.844 1.15e-12 ***
year          6.916e-01  1.441e-01   4.799 4.14e-06 ***
mean_temp     3.918e+01  1.075e+01   3.646 0.000378 ***
varietyELPIDA:year -1.008e-01  1.985e-02  -5.080 1.22e-06 ***
varietyPRG 9811:year -1.555e-01  1.985e-02  -7.835 1.21e-12 ***
year:mean_temp -1.941e-02  5.321e-03  -3.649 0.000375 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1087 on 136 degrees of freedom
Multiple R-squared:  0.758,    Adjusted R-squared:  0.7456
F-statistic: 60.86 on 7 and 136 DF,  p-value: < 2.2e-16

```

Πίνακας 4-2: Επιδράσεις των μεταχειρίσεων

```

Anova Table (Type III tests)

Response: MIC
              Sum Sq Df F value    Pr(>F)
(Intercept)  0.27036  1  22.877 4.433e-06 ***
variety      0.74841  2  31.664 5.128e-12 ***
year        0.27216  1  23.030 4.142e-06 ***
mean_temp   0.15713  1  13.296 0.0003777 ***
variety:year 0.74672  2  31.593 5.385e-12 ***
year:mean_temp 0.15733  1  13.312 0.0003747 ***
Residuals   1.60724 136
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

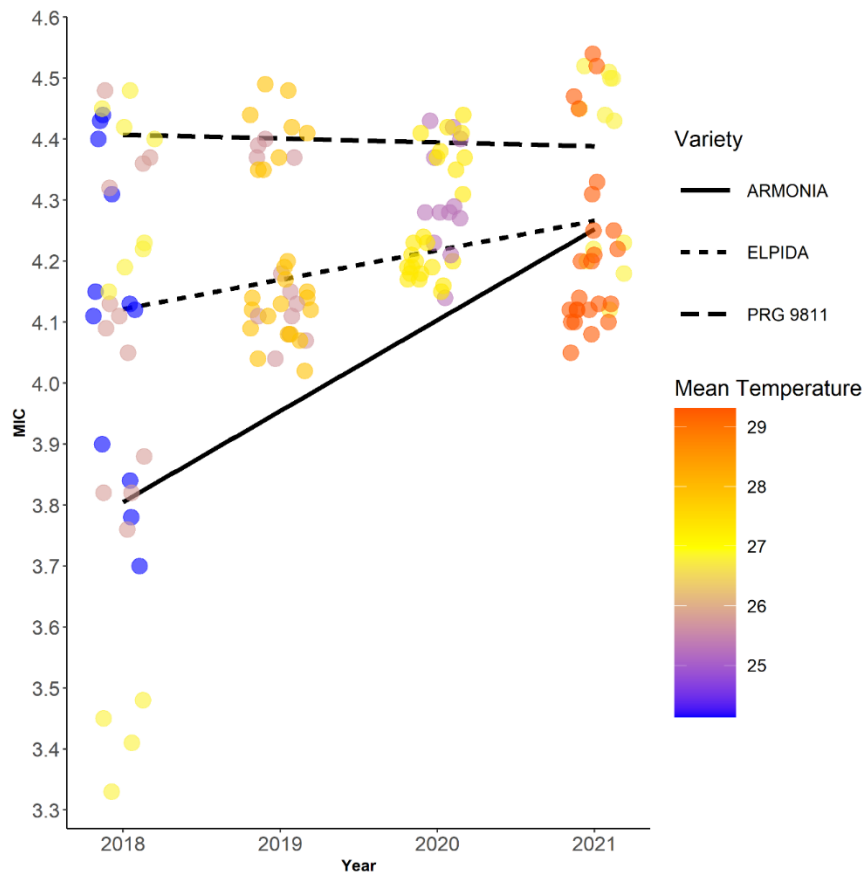
Ο παρακάτω κώδικας παράγει γράφημα που συμπεριλαμβάνει όλες τις μεταβλητές του μοντέλου. Χρησιμοποιήθηκε το πακέτο ggplot2. Η πρώτη γραμμή του κώδικα ορίζει τα χαρακτηριστικά της εικόνας (2500 x 2500 με 300dpi ανάλυση).

```
png("interaction_plot_year_var_mean_temp.png", pointsize=7, width=2500, height=2500,
res=300)

ggplot(data=data,aes(x=year,y=MIC))+
geom_smooth(method="lm",aes(linetype=variety),se=F,color="black",size=1.5)+
  geom_point(aes(color=mean_temp), size=5,alpha=0.6)+
  scale_color_gradient2(midpoint=mean(data$mean_temp),low = "blue", mid="yellow", high
= "red", name="Mean Temperature")+
  theme(legend.position = "right",
        legend.key.size=unit(1.5,"cm"),
        legend.title = element_text(size=15),
        legend.text = element_text(size=12),
        axis.text.x = element_text(size=14),
        axis.text.y = element_text(size=14 ),
        axis.title.y = element_text(face="bold",size=15) ,
        axis.title.x = element_text(face="bold",size=15),
        legend.key = element_rect(fill = "transparent", colour = "transparent"),
        plot.background=element_rect(fill="transparent",colour=NA),
        legend.background = element_rect(fill="white"),panel.background = element_blank(),
        axis.line = element_line(colour = "black"))+xlab("Year")+
  scale_linetype_discrete(name="Variety")+
  scale_y_continuous(breaks=c(seq(3.3,4.5,by=0.1)))

dev.off()
```


Αυτό το διάγραμμα αλληλεπίδρασης δείχνει τη σχέση μεταξύ των micronaire (MIC), της ποικιλίας του φυτού και της μέσης θερμοκρασίας σε ένα δεδομένο έτος.



Διάγραμμα 4-1: Normal Q-Q plot για τα υπόλοιπα του μοντέλου

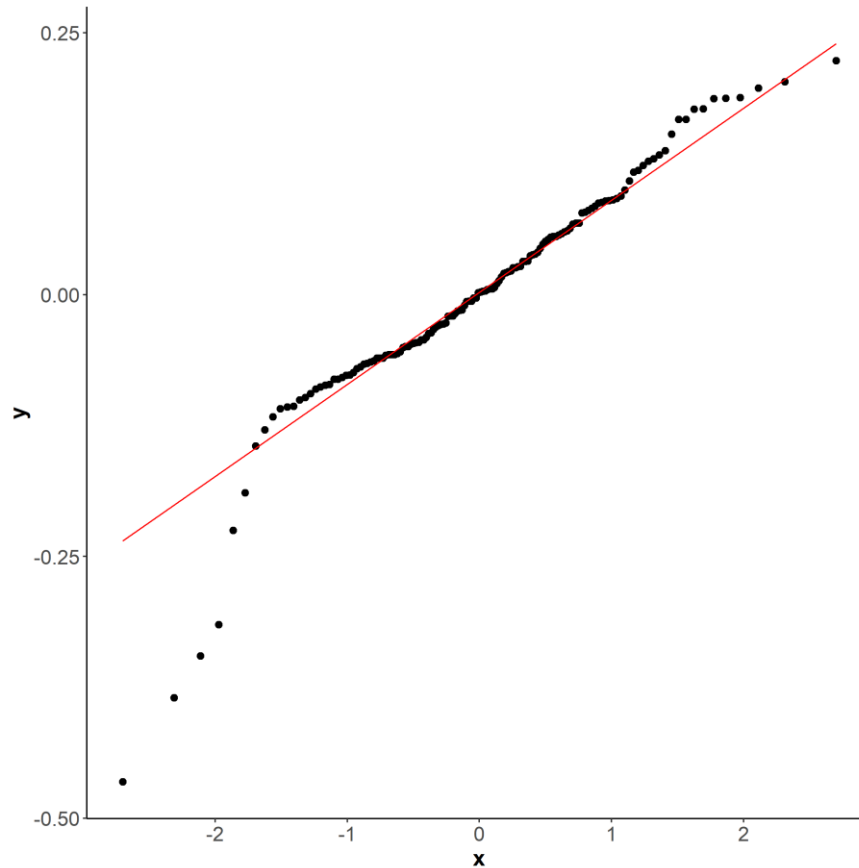
Τα παρακάτω συμπεράσματα αφορούν τις 3 ποικιλίες που μελετώνται όσον αφορά την απόκρισή τους στις διαφορετικές μέσες θερμοκρασίες για τα έτη ενδιαφέροντος, με βάση τις τιμές του micronaire.

Αρχικά, η χρονιά 2018 έχει τις χαμηλότερες μέσες τιμές θερμοκρασίας, η χρονιά 2021 τις υψηλότερες και οι υπόλοιπες βρίσκονται σε ενδιάμεσες τιμές μέσω θερμοκρασιών.

Η ποικιλία PRG9811 παρατηρούμε ότι εμφανίζει σταθερά υψηλότερες τιμές micronaire σε όλες τις χρονιές σε σύγκριση με τις υπόλοιπες ποικιλίες, υποδεικνύοντας ότι η συγκεκριμένη ποικιλία δεν επηρεάζεται από τις μεταβολές της θερμοκρασίας όσον αφορά στις τιμές του micronaire της ίνας.

Εν αντιθέσει, όσον αφορά στις ποικιλίες ELPIDA και ARMONIA, όσο υψηλότερες είναι οι μέσες τιμές θερμοκρασίας, τόσο καλύτερα είναι τα αποτελέσματά τους στις τιμές του micronaire, παρατηρώντας μία γραμμική σχέση ανάμεσά τους.

Ειδικότερα, η ποικιλία ARMONIA φαίνεται να είναι πιο ευαίσθητη στις μεταβολές της θερμοκρασίας, δίνοντας τις χαμηλότερες τιμές micronaire σε θερμοκρασίες μεταξύ 25-26°C, ενώ εμφανίζει ικανοποιητικές τιμές σε υψηλότερες θερμοκρασίες.



Διάγραμμα 4-2: Q-Q Plot Διάγραμμα κανονικότητας

Σύμφωνα με το διάγραμμα, βλέπουμε ότι οι μεταβλητές δεν αποκλίνουν σημαντικά από το αναμενόμενο.

4.2 Για τη μεταβλητή Μήκος ίνας (LENGTH)

Αναπτύχθηκε γραμμικό μοντέλο με εξαρτημένη μεταβλητή LEN και ανεξάρτητες μεταβλητές ποικιλία, έτος, μέση θερμοκρασία, βροχόπτωση, πόλη και ταχύτητα ανέμου. Το τελικό γραμμικό μοντέλο περιείχε τους παράγοντες ποικιλία, έτος, μέση θερμοκρασία και πόλη. Η επιλογή τους βασίστηκε στο κριτήριο AIC.

Βήματα για την επιλογή μοντέλου:

Αρχικά αναπτύχθηκε το πιο απλό μοντέλο χωρίς κάποιο παράγοντα:

```
> model_LEN_intercept<-lm(LEN~1,data=data)
```

Μετά συμπεριλάβαμε όλους τους πιθανούς παράγοντες χωρίς αλληλεπίδραση:

```
> model_LEN_full<-lm(LEN~., data=data)
```

Έγινε χρήση της συνάρτησης step() που επιστρέφει το καλύτερο δυνατό μοντέλο με βάση το κριτήριο AIC:

```
> both<- step(model_LEN_intercept, direction='both', scope=formula(model_LEN_full), trace=0)
```

```
> both$coefficients
```

> lm(LEN~variety+year+mean_temp+city,data=data)

Πίνακας 4-3: Το τελικό μοντέλο με τις αλληλεπιδράσεις.

```

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -636.01670   161.25479   -3.944 0.000127 ***
varietyELPIDA -0.74687     0.11327   -6.593 8.58e-10 ***
varietyPRG 9811 -3.37104     0.11327  -29.760 < 2e-16 ***
year          0.33445     0.08077    4.141 6.02e-05 ***
mean_temp    -0.33505     0.08526   -3.930 0.000134 ***
citylamia     0.74464     0.22962    3.243 0.001486 **
citytrikala   0.65651     0.21134    3.106 0.002302 **
-----
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5549 on 137 degrees of freedom
Multiple R-squared:  0.879,    Adjusted R-squared:  0.8737
F-statistic: 165.8 on 6 and 137 DF,  p-value: < 2.2e-16
    
```

Πίνακας 4-4: Επιδράσεις των μεταχειρίσεων

```

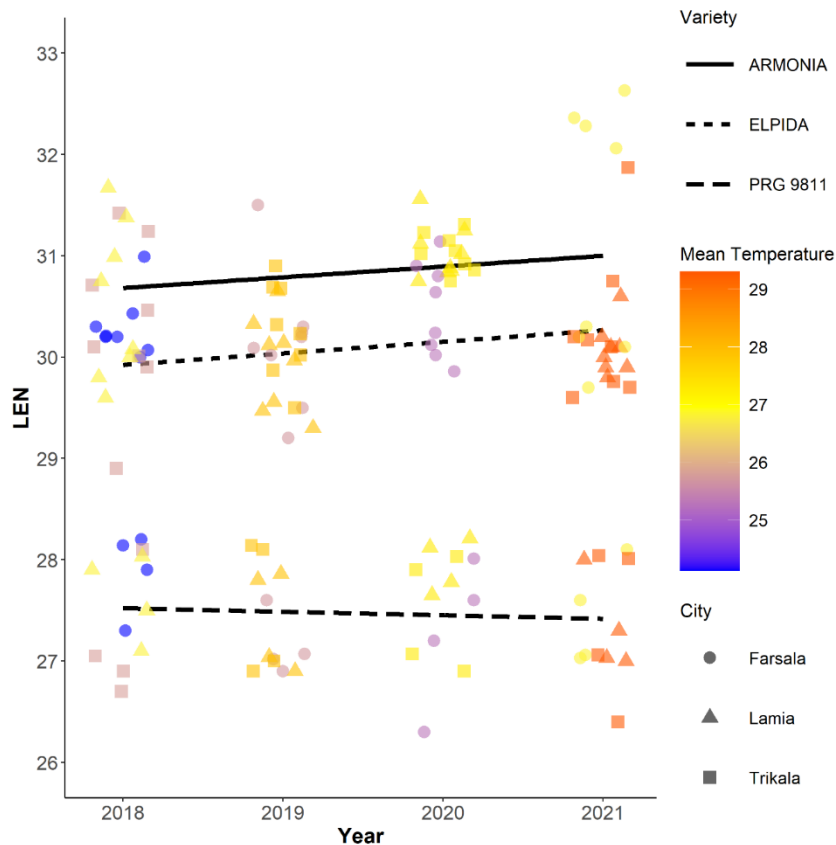
Anova Table (Type III tests)

Response: LEN
              Sum Sq Df F value    Pr(>F)
(Intercept)  4.791   1  15.5565 0.0001273 ***
variety      300.928  2 488.6029 < 2.2e-16 ***
year         5.280   1  17.1456 6.02e-05 ***
mean_temp    4.755   1  15.4411 0.0001344 ***
city         3.347   2   5.4346 0.0053548 **
Residuals   42.189 137
-----
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
    
```

Ο παρακάτω κώδικας παράγει γράφημα που συμπεριλαμβάνει όλες τις μεταβλητές του μοντέλου. Χρησιμοποιήθηκε το πακέτο ggplot2. Η πρώτη γραμμή του κώδικα ορίζει τα χαρακτηριστικά της εικόνας (2500 x 2500 με 300dpi ανάλυση).

```
png("plot LEN.png", pointsize=7, width=2500, height=2500, res=300)
ggplot(data=data,aes(x=year,y=LEN))+
  geom_smooth(method="lm",aes(linetype=variety),se=F,color="black",size=1.5)+
  scale_color_gradient2(midpoint=mean(data$mean_temp),low = "blue",mid="yellow", high
= "red",
                        name="Mean Temperature")+
  theme(legend.position = "right",
        legend.key.size=unit(1.5,"cm"),
        legend.title = element_text(size=13)
        legend.text = element_text(size=12)
        axis.text.x = element_text(size=14)
        axis.text.y = element_text(size=14)
        axis.title.y = element_text(face="bold",size=15) ,
        axis.title.x = element_text(face="bold",size=15),
        legend.key = element_rect(fill = "transparent", colour = "transparent"),
        plot.background=element_rect(fill="transparent",colour=NA),
        legend.background = element_rect(fill="white"),panel.background = element_blank(),
        axis.line = element_line(colour = "black"))+xlab("Year")+
  scale_linetype_discrete(name="Variety")+
  scale_shape_discrete(name="City")+
  scale_y_continuous(breaks=c(seq(26,34,by=1)), limits = c(26,33))+
  geom_jitter(aes(shape=city,color=mean_temp),size=4,alpha=0.6,width=0.2,height=0)
dev.off()
```

Η γραφική παράσταση δείχνει ότι το μήκος της ίνας αυξάνεται με την αύξηση της μέσης θερμοκρασίας, με την ισχυρότερη αύξηση του μήκους ίνας να εμφανίζεται σε ποικιλίες με υψηλότερη μέση θερμοκρασία.



Διάγραμμα 4-3: Normal Q-Q plot για τα υπόλοιπα του μοντέλου

Από το διάγραμμα 4-3 εξάγουμε αρκετά συμπεράσματα. Για τις ποικιλίες Αρμονία και Ελπίδα οι υψηλότερες μέσες τιμές θερμοκρασιών εμφανίζουν μεγαλύτερες τιμές του μήκους της ίνας, κάτι το οποίο δεν ισχύει για τη ποικιλία PRG9811. Βέβαια, η συσχέτιση ανάμεσα στη μέση τιμή της θερμοκρασίας και του μήκους ίνας δε φαίνεται να είναι σημαντική.

Η PRG9811 εμφανίζει αρκετά μικρότερες τιμές μήκους ίνας για όλες τις θερμοκρασίες σε σχέση με τις υπόλοιπες 2 ποικιλίες, όπως επίσης η ARMONIA εμφανίζει μεγαλύτερες τιμές μήκους ίνας σε σχέση με την ELPIDA σε όλες τις θερμοκρασίες.

Το μόνο αξιοσημείωτο όσον αφορά τη συσχέτιση του μήκους ίνας και των περιοχών συλλογής δειγμάτων είναι οι ιδιαίτερα υψηλές τιμές που εμφανίζει η περιοχή των Φαρσάλων στις μέσες τιμές θερμοκρασιών κοντά στους 27°C.

png("LEN qqplot.png", pointsize=7, width=2500, height=2500, res=300)

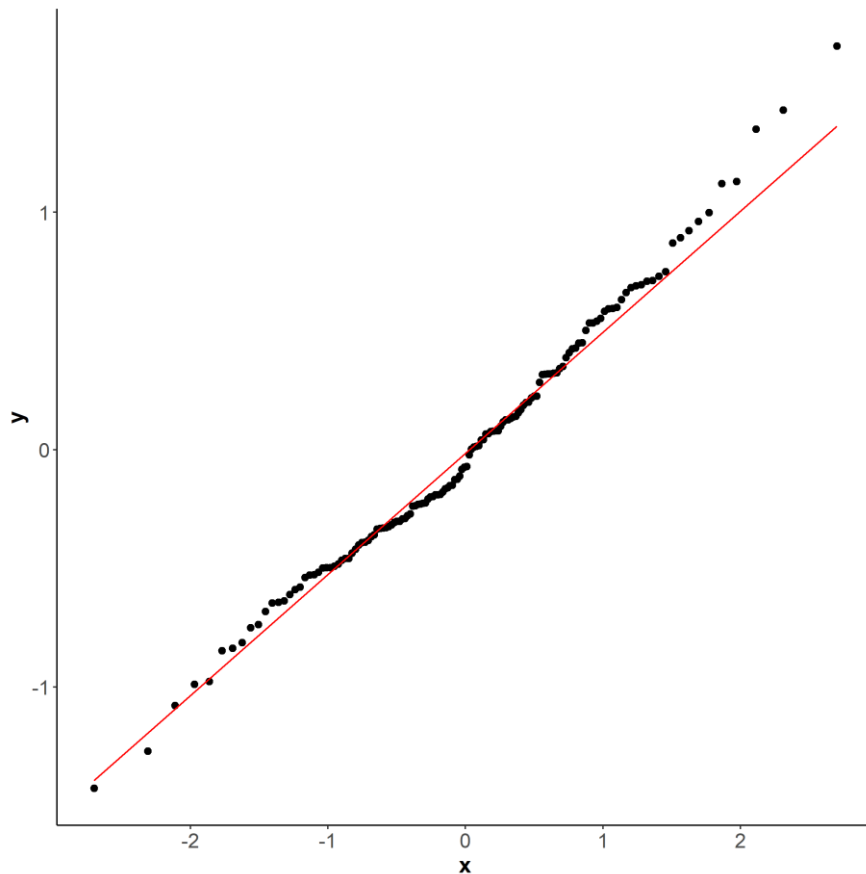
```
ggplot(NULL, aes(sample = residuals(model_LEN_full_inter)),alpha=0.5) + # Create
QQplot with ggplot2 package

stat_qq(size=2) +

theme(axis.text.x = element_text(size=14),
      axis.text.y = element_text(size=14),
      axis.title.y = element_text(face="bold",size=15) ,
      axis.title.x = element_text(face="bold",size=15),
      panel.background = element_blank(),
      plot.background=element_rect(fill="transparent",colour=NA),
      axis.line = element_line(colour = "black"))+

stat_qq_line(col = "red")

dev.off()
```



Διάγραμμα 4-4: Q-Q Plot Διάγραμμα κανονικότητας

Γραφικά φαίνεται ότι τα σημεία του διαγράμματος είναι πολύ κοντά στην ευθεία γραμμή και εντός της ζώνης που ορίζεται η κανονικότητα. Αυτό μας επιτρέπει να ισχυριστούμε ότι δεν αποκλίνουν σημαντικά από το αναμενόμενο.

4.3 Για τη μεταβλητή : Αντοχή της ίνας (STRENGTH)

Αναπτύχθηκε γραμμικό μοντέλο με εξαρτημένη μεταβλητή STR και ανεξάρτητες μεταβλητές ποικιλία, έτος, μέση θερμοκρασία, βροχόπτωση, πόλη και ταχύτητα ανέμου. Το τελικό γραμμικό μοντέλο περιείχε τους παράγοντες έτος και μέση θερμοκρασία, πόλη, Βροχόπτωση και η αλληλεπίδραση της χρονιάς και της βροχόπτωσης. Η επιλογή τους βασίστηκε στο κριτήριο AIC.

Βήματα για την επιλογή μοντέλου:

Αρχικά αναπτύχθηκε το πιο απλό μοντέλο χωρίς κάποιο παράγοντα:

```
> model_STR_intercept<-lm(STR~1,data=data)
```

Μετά συμπεριλάβαμε όλους τους πιθανούς παράγοντες χωρίς αλληλεπίδραση:

```
> model_STR_full<-lm(STR~., data=data)
```

Έγινε χρήση της συνάρτησης step() που επιστρέφει το καλύτερο δυνατό μοντέλο με βάση το κριτήριο AIC:

```
> both<- step(model_STR_intercept, direction='both', scope=formula(model_STR_full), trace=0)
```

```
> both$coefficients
```

```
> lm(STR~year+mean_temp,data=data)
```

```

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.973e+03  3.095e+02 -9.606 < 2e-16 ***
year          1.497e+00  1.546e-01  9.688 < 2e-16 ***
mean_temp    -8.008e-01  1.285e-01 -6.230 5.37e-09 ***
cityLamia     1.208e+00  3.189e-01  3.788 0.000226 ***
cityTrikala   1.617e+00  3.066e-01  5.274 5.09e-07 ***
`Rain_total (mm)` 2.826e+01  4.780e+00  5.912 2.56e-08 ***
year:`Rain_total (mm)` -1.399e-02  2.367e-03 -5.910 2.58e-08 ***
-----
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6973 on 137 degrees of freedom
Multiple R-squared:  0.4686,    Adjusted R-squared:  0.4453
F-statistic: 20.13 on 6 and 137 DF,  p-value: < 2.2e-16

```

Πίνακας 4-5: Το τελικό μοντέλο με τις αλληλεπιδράσεις.

Anova Table (Type III tests)

Response: STR

	Sum Sq	Df	F value	Pr(>F)	
(Intercept)	32.679	1	55.6607	8.955e-12	***
year	33.506	1	57.0690	5.392e-12	***
mean_temp	3.055	1	5.2037	0.024081	*
city	6.211	2	5.2894	0.006126	**
`Rain_total (mm)`	3.349	1	5.7046	0.018284	*
mean_temp: `Rain_total (mm)`	3.167	1	5.3935	0.021686	*
Residuals	80.433	137			

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Πίνακας 4-6:Επιδράσεις των μεταχειρίσεων

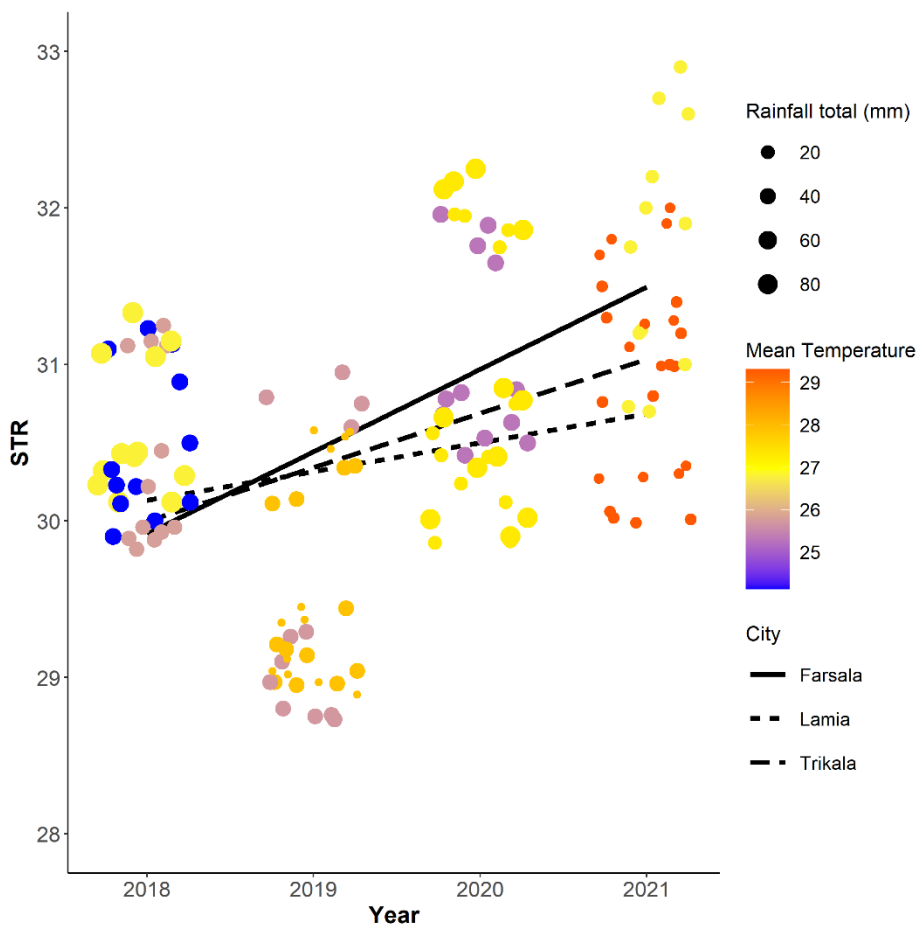
Ο παρακάτω κώδικας παράγει γράφημα που συμπεριλαμβάνει όλες τις μεταβλητές του μοντέλου. Χρησιμοποιήθηκε το πακέτο ggplot2. Η πρώτη γραμμή του κώδικα ορίζει τα χαρακτηριστικά της εικόνας (2500 x 2500 με 300dpi ανάλυση).

```
png("plot STR.png", pointsize=7, width=2500, height=2500, res=300)
ggplot(data=data,aes(x=year,y=STR))+
  geom_smooth(method="lm",aes(linetype=city),se=F,color="black",size=1.5)+
  scale_color_gradient2(midpoint=mean(data$mean_temp),low = "blue",mid="yellow",
high = "red",
                        name="Mean Temperature")+
theme(legend.position = "right",
      legend.key.size=unit(1,"cm"),
      legend.title = element_text(size=13),
      legend.text = element_text(size=12),
      axis.text.x = element_text(size=14),
      axis.text.y = element_text(size=14 ),
      axis.title.y = element_text(face="bold",size=15) ,
      axis.title.x = element_text(face="bold",size=15),
      legend.key = element_rect(fill = "transparent", colour = "transparent"),
      plot.background=element_rect(fill="transparent",colour=NA),
      legend.background = element_rect(fill="white"),panel.background = element_blank(),
      axis.line = element_line(colour = "black"))+xlab("Year")+
```



```
scale_linetype_discrete(name="City")+  
scale_y_continuous(breaks=c(seq(28,33,by=1)), limits = c(28,33))+  
geom_jitter(aes(color=mean_temp, size=`Rain_total (mm)`),width=0.3,height=0)+  
scale_size_continuous(name="Rainfall total",range=c(2,6),breaks=c(seq(0,100,by=20)))  
dev.off()
```

Αυτό το διάγραμμα αλληλεπίδρασης δείχνει τη σχέση μεταξύ της αντοχής της ίνας (STR) του φυτού, της πόλης, της μέσης θερμοκρασίας σε ένα δεδομένο έτος και της συνολικής βροχόπτωσης.



Διάγραμμα 4-5: Normal Q-Q plot για τα υπόλοιπα του μοντέλου

Το πρώτο σημαντικό συμπέρασμα που μπορούμε να εξάγουμε από το διάγραμμα 4-5 είναι ότι το έτος 2018 οι τιμές της αντοχής της ίνας είναι πολύ κοντά για όλες τις περιοχές, με τη περιοχή της Λαμίας να εμφανίζει λίγο καλύτερα αποτελέσματα.

Αντίθετα για τα έτη 2020, 2021 η περιοχή της Λαμίας εμφανίζει τις χαμηλότερες τιμές αντοχής, με τη περιοχή των Φαρσάλων να έχει σημαντικά υψηλότερες τιμές και τα Τρίκαλα να ακολουθούν.

Η διαφορά στις τιμές της αντοχής της ίνας για τα παραπάνω 2 έτη ανάμεσα στις 3 περιοχές είναι αρκετά μεγάλη.

Για τη συσχέτιση ανάμεσα στη μέση τιμή της θερμοκρασίας και στην αντοχή της ίνας ένα συμπέρασμα που μπορούμε να εξαγάγουμε είναι ότι η τιμή 27^ο C φαίνεται να είναι η ιδανική ανεξαρτήτως της ποσότητας των βροχοπτώσεων ή της περιοχής.

Οι χαμηλότερες τιμές της αντοχής της ίνας εμφανίζονται το έτος 2019, χρονιά που παρατηρήθηκε η ελάχιστη μέση ποσότητα βροχόπτωσης με τιμή < 20mm.

Οι υψηλότερες τιμές αντοχής της ίνας εμφανίζονται το έτος 2021, χρονιά που η μέση τιμή θερμοκρασίας ήταν υψηλότερη σε σχέση με τις υπόλοιπες και οι βροχοπτώσεις κάτω του μέσου όρου.

```
png("STR qqplot.png", pointsize=7, width=2500, height=2500, res=300)
```

```
ggplot(NULL, aes(sample = residuals(model_STR_full_inter_2)),alpha=0.5) + # Create  
QQplot with ggplot2 package
```

```
stat_qq(size=2) +
```

```
theme(axis.text.x = element_text(size=14),
```

```
axis.text.y = element_text(size=14),
```

```
axis.title.y = element_text(face="bold",size=15) ,
```

```
axis.title.x = element_text(face="bold",size=15),
```

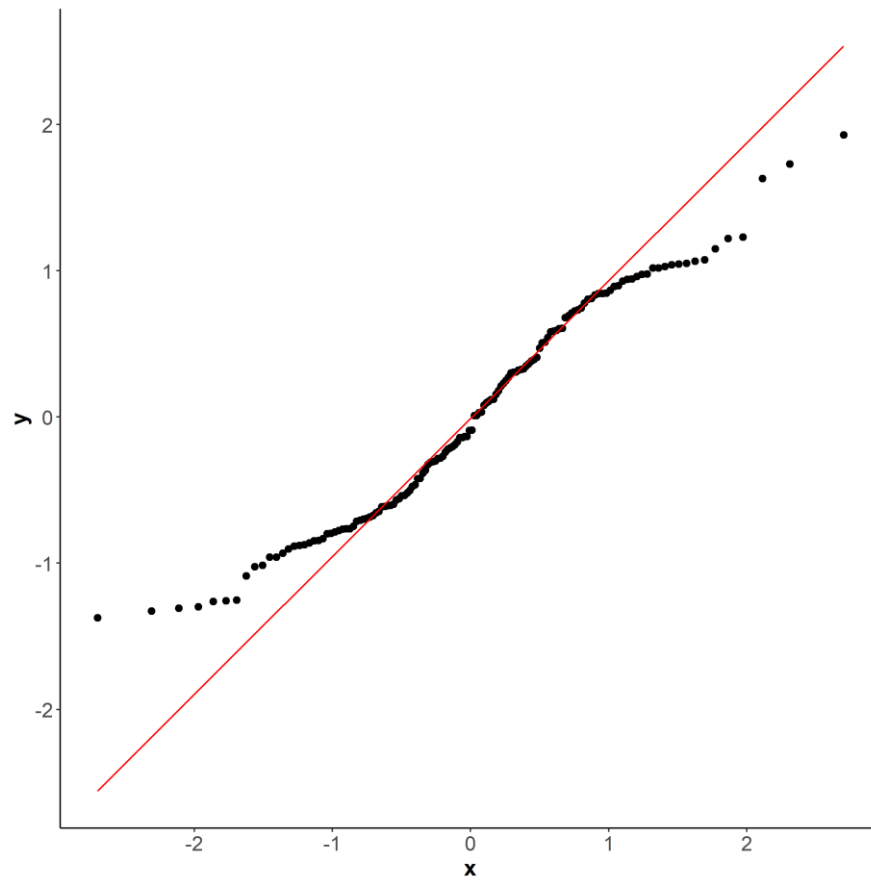
```
panel.background = element_blank(),
```

```
plot.background=element_rect(fill="transparent",colour=NA),
```

```
axis.line = element_line(colour = "black")+
```

```
stat_qq_line(col = "red")
```

```
dev.off()
```



Διάγραμμα 4-6: Q-Q Plot Διάγραμμα κανονικότητας

Σύμφωνα με το διάγραμμα , βλέπουμε ότι οι μεταβλητές δεν αποκλίνουν σημαντικά από το αναμενόμενο υποδεικνύοντας ότι το μοντέλο πλήρους αλληλεπίδρασης είναι επαρκές για τα δεδομένα.

ΚΕΦΑΛΑΙΟ 5: ΣΥΜΠΕΡΑΣΜΑΤΑ- ΠΡΟΤΑΣΕΙΣ

5.1 ΣΥΖΗΤΗΣΗ

Σε αυτό το κεφάλαιο, θα αναλύσουμε και θα συγκρίνουμε τα αποτελέσματα από τα Κεφάλαια 3 και 4, συμπεριλαμβανομένων των βιβλιογραφικών μας δεδομένων, για να εξάγουμε, όσο είναι δυνατό, κάποια συμπεράσματα.

5.1.1 Χαρακτηριστικό ποιότητας : Micronaire

Η ποικιλία PRG9811 φαίνεται να έχει σταθερά υψηλότερες τιμές micronaire από τις άλλες δύο ποικιλίες, υποδεικνύοντας ότι αυτή η ποικιλία δεν επηρεάζεται από αλλαγές θερμοκρασίας όσον αφορά τις τιμές micronaire των ινών. Αυτό πιθανότατα οφείλεται στο γεγονός ότι είναι μια υβριδική ποικιλία και ως εκ τούτου έχει δημιουργηθεί για να είναι πιο ανθεκτική στις διακυμάνσεις της θερμοκρασίας. Μελέτες έχουν δείξει ότι οι υβριδικές ποικιλίες είναι γενικά πιο σταθερές όσον αφορά την απόδοση και την ποιότητα των ινών από τις μη υβριδικές ποικιλίες (Meher et al., 2016).

Η ποικιλία Eirida παρατηρήθηκε ότι έχει αυξανόμενες τιμές με την πάροδο των ετών, με υψηλότερους αριθμούς micronaire σε υψηλότερες θερμοκρασίες. Γενετικά, Η ποικιλία βαμβακιού Eirida έχει ορισμένα χαρακτηριστικά που το καθιστούν πιο ανθεκτικό στις αλλαγές θερμοκρασίας και υγρασίας. Αυτά τα χαρακτηριστικά επιτρέπουν στις ίνες να διατηρούν υψηλότερο αριθμό micronaire σε υψηλότερες θερμοκρασίες (Cotton Incorporated, 2019).

Τέλος, η ποικιλία Armonia από την ανάλυση εμφανίζει χαμηλότερες τιμές micronaire σε θερμοκρασίες μεταξύ 25-26°C, ενώ παρουσιάζει ικανοποιητικές τιμές σε υψηλότερες θερμοκρασίες. Αυτό μπορεί να οφείλεται στο γεγονός ότι η Armonia είναι μια παλαιότερη ποικιλία και είναι λιγότερο ανθεκτική στις αλλαγές των περιβαλλοντικών συνθηκών.

Ως εκ τούτου, η Armonia είναι καταλληλότερη για πιο κλίματα με πιο μέτριες μέσες θερμοκρασίες, ενώ η Eirida και το PRG9811 είναι καλύτερες για περιοχές με πιο μεταβλητές θερμοκρασίες, όσον αφορά την απόδοση σε micronaire.

5.1.2 Χαρακτηριστικό ποιότητας : Αντοχή της ίνας

Μετά από την ανάλυση των δεδομένων, ένα από τα συμπεράσματα στα οποία καταλήγουμε είναι ότι υπάρχει συσχέτιση μεταξύ των τιμών της αντοχής της ίνας και των μέσων θερμοκρασιών. Όσο υψηλότερες είναι οι μέσες τιμές θερμοκρασιών τόσο υψηλότερες τιμές αντοχής της ίνας για τις ποικιλίες Eirida και Armonia ,ενώ η ποικιλία PRG9811 εμφανίζει τα αντίθετα αποτελέσματα.

Ιδιαίτερο ενδιαφέρον παρουσιάζουν οι χρονιές 2018 και 2021. Το 2021 έχουμε τα χαμηλότερα ύψη βροχής και τις υψηλότερες μέσες θερμοκρασίες και παρατηρούμε μια μεγάλη διαφοροποίηση των τιμών της αντοχής της ίνας ανάμεσα στις περιοχές

μελέτης, ενώ το 2018 όπου εμφανίζονται μεγάλα ύψη βροχής οι τιμές της αντοχής για όλες τις περιοχές μελέτης είναι αρκετά κοντά.

Η περιοχή των Φαρσάλων παρουσιάζει τις υψηλότερες τιμές αντοχής ίνας και τα τρία χρόνια με θερμοκρασίες άνω του μέσου όρου κ ιδιαίτερα το 2021 χρονιά με την υψηλότερη μέση θερμοκρασία από τις χρονιές μελέτης, και τα μικρότερα ύψη βροχής. Αυτό αποδεικνύει ότι υπάρχει ισχυρή συσχέτιση ανάμεσα στη περιοχή μελέτης και την αντοχή της ίνας.

Λαμβάνοντας υπόψη ότι αυτό γίνεται πιο έντονο τις χρονιές όπου είναι πιο εύκολες και πιο σημαντικές οι επεμβάσεις στη καλλιέργεια (άρδευση, ψεκασμοί, λίπανση), οδηγούμαστε στο συμπέρασμα ότι οι καλές καλλιεργητικές τεχνικές έχουν ισχυρή συσχέτιση με την αντοχή της ίνας του τελικού προϊόντος.

Επιπροσθέτως τα ύψη βροχής δε φαίνεται να επηρεάζουν ιδιαίτερα τις τιμές της αντοχής της ίνας, το οποίο κρίνεται φυσιολογικό καθότι η πλειονότητα των εκτάσεων στις περιοχές μελέτης είναι αρδευόμενες.

5.1.3 Χαρακτηριστικό ποιότητας : Μήκος της ίνας

Παρόλο που τα αποτελέσματα της ανάλυσης μας δεν έδειξαν ισχυρή συσχέτιση του μήκους της ίνας, του έτους, και των μέσων θερμοκρασιών, υπάρχουν αρκετά σημεία με ιδιαίτερο ενδιαφέρον. Παρατηρήθηκε μια μικρή συσχέτιση μεταξύ των μέσων θερμοκρασιών και του μήκους της ίνας για τη πλειονότητα των ποικιλιών της μελέτης. Οι υψηλότερες μέσες θερμοκρασίες φαίνεται να συμβάλουν στη δημιουργία τελικού προϊόντος με μεγαλύτερο μήκος ίνας.

Το πιο ενδιαφέρον συμπέρασμα στο οποίο καταλήγουμε είναι ότι η ποικιλία παίζει το σημαντικότερο ρόλο στο μέγεθος του μήκους ίνας ενώ όλοι οι υπόλοιποι περιβαλλοντικοί παράγοντες και η περιοχή μελέτης δε φαίνεται να επηρεάζουν ιδιαίτερα τον παράγοντα μήκος.

Συμπερασματικά, η Elpida και η Armonia έδειξαν θετική συσχέτιση μεταξύ του μήκους των ινών και της θερμοκρασίας, ενώ το PRG9811 έδειξε αρνητική συσχέτιση. Αυτό υποδηλώνει ότι η Elpida και η Armonia ίσως ταιριάζουν καλύτερα σε θερμότερα κλίματα, ενώ το PRG9811 είναι καταλληλότερο σε πιο ψυχρά κλίματα.

5.2 Προτάσεις για περαιτέρω έρευνα

Για να υπάρχει συνέχεια και να αξιοποιηθούν οι πληροφορίες της εργασίας αυτής, θα πρέπει να συνεχιστεί η στατιστική επεξεργασία των δεδομένων των χαρακτηριστικών ποιότητας της ίνας του βαμβακιού για τα επόμενα έτη σε περισσότερες περιοχές που καλλιεργείται το βαμβάκι. Προτείνεται να ληφθούν υπ' όψη στην αξιολόγηση των χαρακτηριστικών ποιότητας οι παρακάτω παράγοντες:

- καλλιεργητικές τεχνικές,
- μέθοδος συγκομιδής,
- μέθοδος εκκοκκισμού.

Επίσης προτείνεται να επεκταθεί η έρευνα και σε επιπλέον χαρακτηριστικά ποιότητας της ίνας, όπως:

- το κυτίο,
- η ομοιομορφία,
- ο χρωματισμός,
- τα neps

ΒΙΒΛΙΟΓΡΑΦΙΚΕΣ ΑΝΑΦΟΡΕΣ

- 1) Askin, R.G. (2012). *Textile Science* (9th ed.). Hoboken, NJ: John Wiley & Sons, Inc.
- 2) Bao, G., Liu, S., Guo, W., Zhang, X., and Li, Y. (2009). The role of lignin deposition in the mechanical strength of plant cell wall. *Plant Cell Reports*, 28(5), 795-804.
- 3) Babu, S., M., et al. (2018). Cotton Quality: Fiber and Fabric Properties. In *Cotton Production Manual* (pp. 39-42). Academic Press.
- 4) Chamblee, T. J., C. R. Hall, J. W. Madden, and D. D. Tyler. (2018). Analysis of Completely Randomized Designs in Animal Science Research. *Animal Science*, 98: 345–359.
- 5) Chou, C. C. (2010). Cost of production and cotton quality management. *Cotton International*, 32(6), 18–20.
- 6) Cotton Incorporated. (2019). *Elpida: New Cotton Variety in the U.S.*
- 7) Dong, J., Zhang, H., and Zhang, X. (2010). Cell wall assembly and secondary wall formation in plants. *Plant Signaling & Behavior*, 5(7), 815-818.
- 8) Duncan, D. B. (1977). On the application of ANCOVA to the analysis of randomized block designs. *Journal of Educational Statistics*, 2(3), 327-339.
- 9) El-Kommos, Y. (2013). *Classification of Fibers: Natural, Regenerated and Synthetic*. New York, NY: Woodhead Publishing Ltd.
- 10) Feng, Z., Liu, J., and Li, X. (2012). Abiotic and biotic stress effects on plant cell wall biosynthesis and remodeling. *Plant Physiology*, 158(3), 1063-1073.
- 11) Fischer, M. (2006). High-Volume Instruments and Cotton Fiber Quality. *Cotton International*, 28(4), 28–31.
- 12) Gupta, A., Kalia, P., & Sharma, D. (2018). Effect of different agronomic practices on cotton fiber quality. *International Journal of Chemical Studies*, 6(1), 20-22.

- 13) Holland, B. (2020) Completely Randomized Design (CRD). Statistical Solutions. Accessed May 7, 2020. <https://www.statisticalsolutions.net/completely-randomized-design-crd/>
- 14) Khan, M. U., Marwat, K. B., Abro, R. A., & Ali, M. (2019). Effect of temperature on quality of cotton fibre. *Journal of Animal and Plant Sciences*, 29(2), 522-530.
- 15) Kovac, V., de Vicente, M. C., & Bautista, I. (2009). Temperature effects on cotton fiber properties. *Industrial Crops and Products*, 29(3), 543-548.
- 16) Kumar, M. S., Singh, G., & Gupta, V. (2017). Genetic variability and heritability of seed cotton yield and its components in *Gossypium hirsutum* L. *Crop Research*, 45(3), 331-337.
- 17) Kumar, N., et al. (2019). Fiber properties and their effects on fabric properties. *Indian Journal of Fibre & Textile Research*.
- 18) Ladisch, M. and Lippman, Z.B. (2007). Plant cell wall biochemistry and its application to biofuel production. *Current Opinion in Biotechnology*, 18(2), 113-120.
- 19) Li, Y., Wu, Y., & Wang, X. (2017). Estimation of cotton fiber quality and its related traits under different temperature conditions in different growth stages. *Field Crops Research*, 206, 144-153.
- 20) Liu, Y., Li, L., Gao, J., and Liu, X. (2011). Auxin and cytokinin control of fiber development. *Plant and Cell Physiology*, 52(7), 1172-1179.
- 21) Lu, P., Li, J., Li, Z., Li, H., & Zhang, Y. (2018). Effect of environmental factors on fiber quality of cotton: A review. *Field Crops Research*, 225, 93-103.
- 22) Montgomery, D. C. (2001). *Design and analysis of experiments* (5th ed.). New York, NY: John Wiley & Sons.
- 23) Meher, P. K., Shukla, N., & Hiremath, P. (2016). Yield and fiber quality of some hybrid and non-hybrid cotton cultivars. *Indian Journal of Fiber & Textile Research*, 41(2), 195-199.
- 24) Muskat, M. (2019). *Completely Randomized Design in Agricultural Experiments*. *Crop Management*. Accessed May 7, 2020.

- 25) Nanda, P. K., Jha, S. K., & Singh, R. K. (2018). Cotton Fiber Quality: A Review. *Agriculture and Agricultural Science Procedia*, 18, 17-28.
- 26) Ramirez, R. E., Condon, B. D., & Smith, D. J. (2009). Temperature effects on cotton fiber quality. *Industrial Crops and Products*, 30(1), 13-17.
- 27) Zheng, L., et al. (2020). Characterization and comparison of physical and chemical properties of cotton fibers from different varieties. *Journal of Natural Fibers*.