Lecture 14 - NESTED ANOVA

Bill Perry

# 1. Introduction

This analysis examines the effects of varying sea urchin densities on the percentage cover of filamentous algae. The experiment was designed with four urchin density treatments (control, 66% of original density, 33% of original density, and all urchins removed) nested within four random patches. Five replicate quadrats were measured within each treatment-patch combination.

The traditional nested ANOVA approach can be implemented using a linear mixed-effects model, which provides a more flexible framework for analyzing hierarchical designs. In this case, we’ll use the lme4 package to fit a model where treatment is a fixed effect and patch is a random effect nested within treatment.

# 2. Data Overview

The dataframe contains 80 observations with the following variables:

* TREAT: Urchin density treatment (Control, 66% Density, 33% Density, Removed)
* PATCH: Random patches (1-16) where treatments were applied
* QUAD: Replicate quadrats within each treatment-patch combination
* ALGAE: Percentage cover of filamentous algae (response variable)

# Create a summary table with flextable  
  
summary\_stats

# A tibble: 4 × 7  
 TREAT n mean sd se min max  
 <fct> <int> <dbl> <dbl> <dbl> <dbl> <dbl>  
1 Control 20 1.3 3.18 0.711 0 13  
2 66% Density 20 21.6 25.1 5.62 0 79  
3 33% Density 20 19 25.7 5.74 0 71  
4 Removed 20 39.2 28.7 6.41 0 83

#   
# summary\_stats %>%  
# select(TREAT, n, mean, sd, se, min, max) %>%  
# flextable() %>%  
# set\_header\_labels(  
# TREAT = "Treatment",  
# n = "N",  
# mean = "Mean",  
# sd = "SD",  
# se = "SE",  
# min = "Min",  
# max = "Max"  
# ) %>%  
# colformat\_double(j = c("mean", "sd", "se", "min", "max"), digits = 2) %>%  
# autofit() %>%  
# add\_header\_lines("Summary statistics of algae cover (%) across treatments") %>%  
# theme\_box()

# 3. Mixed Model Analysis

In this experimental design, PATCH is nested within TREAT because each patch received only one treatment level. This hierarchical design is well-suited for analysis using linear mixed-effects models.

## 3.1 Model Specification

We’ll use the following model specification:

Where: - is the overall mean - is the fixed effect of treatment - is the random effect of patch nested within treatment - is the residual error for quadrat in patch within treatment

In lme4, this model is specified as

# Fit the mixed model  
mixed\_model <- lmer(ALGAE ~ TREAT + (1|TREAT:PATCH), data = andrew)  
  
# Display model summary  
summary(mixed\_model)

Linear mixed model fit by REML. t-tests use Satterthwaite's method [  
lmerModLmerTest]  
Formula: ALGAE ~ TREAT + (1 | TREAT:PATCH)  
 Data: andrew  
  
REML criterion at convergence: 682.2  
  
Scaled residuals:   
 Min 1Q Median 3Q Max   
-1.9808 -0.3106 -0.1093 0.2831 2.5910   
  
Random effects:  
 Groups Name Variance Std.Dev.  
 TREAT:PATCH (Intercept) 294.3 17.16   
 Residual 298.6 17.28   
Number of obs: 80, groups: TREAT:PATCH, 16  
  
Fixed effects:  
 Estimate Std. Error df t value Pr(>|t|)   
(Intercept) 1.300 9.408 12.000 0.138 0.8924   
TREAT66% Density 20.250 13.305 12.000 1.522 0.1539   
TREAT33% Density 17.700 13.305 12.000 1.330 0.2081   
TREATRemoved 37.900 13.305 12.000 2.849 0.0147 \*  
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
Correlation of Fixed Effects:  
 (Intr) TREAT6D TREAT3D  
TREAT66%Dns -0.707   
TREAT33%Dns -0.707 0.500   
TREATRemovd -0.707 0.500 0.500

## 3.2 ANOVA Table

The ANOVA table for the mixed model:

# Get ANOVA table with Type III tests  
anova\_table <- anova(mixed\_model, type = 3)  
print(anova\_table)

Type III Analysis of Variance Table with Satterthwaite's method  
 Sum Sq Mean Sq NumDF DenDF F value Pr(>F)   
TREAT 2434 811.33 3 12 2.7171 0.09126 .  
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

# For comparison, also run a traditional nested ANOVA  
nested\_aov <- aov(ALGAE ~ TREAT + TREAT:PATCH, data = andrew)  
std\_summary <- summary(nested\_aov)[[1]]  
  
# Extract MS values - using exact row names  
MS\_treat <- std\_summary["TREAT ", "Mean Sq"]   
MS\_patch <- std\_summary["TREAT:PATCH", "Mean Sq"]  
MS\_residual <- std\_summary["Residuals", "Mean Sq"]  
  
# Print MS values to check  
print("MS values:")

[1] "MS values:"

print(c(Treatment = MS\_treat, Patches = MS\_patch, Residual = MS\_residual))

Treatment Patches Residual   
 4809.712 1770.162 298.600

# Extract df values  
df\_treat <- std\_summary["TREAT ", "Df"]  
df\_patch <- std\_summary["TREAT:PATCH", "Df"]  
df\_residual <- std\_summary["Residuals", "Df"]  
  
  
# Calculate correct F ratios for nested design  
F\_treat <- MS\_treat / MS\_patch  
F\_patch <- MS\_patch / MS\_residual  
  
# Calculate p-values  
p\_treat <- pf(F\_treat, df\_treat, df\_patch, lower.tail = FALSE)  
p\_patch <- pf(F\_patch, df\_patch, df\_residual, lower.tail = FALSE)  
  
# Create ANOVA table  
trad\_anova\_table <- data.frame(  
 Source = c("Treatment", "Patches (treatment)", "Residual"),  
 df = c(df\_treat, df\_patch, df\_residual),  
 MS = c(MS\_treat, MS\_patch, MS\_residual),  
 F = c(F\_treat, F\_patch, NA),  
 p = c(p\_treat, p\_patch, NA)  
)  
  
# Format p-values  
trad\_anova\_table$p <- ifelse(trad\_anova\_table$p < 0.001, "<0.001",  
 ifelse(is.na(trad\_anova\_table$p), NA,  
 format(trad\_anova\_table$p, digits = 3)))  
trad\_anova\_table

Source df MS F p  
1 Treatment 3 4809.712 2.717102 0.091262004  
2 Patches (treatment) 12 1770.162 5.928207 <0.001  
3 Residual 64 298.600 NA <NA>

# # Display traditional ANOVA table with flextable  
# trad\_anova\_table %>%  
# flextable() %>%  
# set\_header\_labels(  
# Source = "Source of variation",  
# df = "df",  
# MS = "MS",  
# F = "F",  
# p = "p"  
# ) %>%  
# colformat\_double(j = c("MS", "F"), digits = 2) %>%  
# autofit() %>%  
# add\_header\_lines("ANOVA table for nested design") %>%  
# theme\_box()

## 3.3 Variance Components

We can extract the variance components from the mixed model:

# Print corrected results  
  
# Extract variance components  
vc <- VarCorr(mixed\_model)  
print(vc)

Groups Name Std.Dev.  
 TREAT:PATCH (Intercept) 17.156   
 Residual 17.280

# Extract variance components  
var\_comp\_patch <- as.numeric(vc$`TREAT:PATCH`)  
var\_comp\_residual <- attr(vc, "sc")^2  
  
# Calculate percentage of total variance  
total\_var <- var\_comp\_patch + var\_comp\_residual  
pct\_patch <- var\_comp\_patch / total\_var \* 100  
pct\_residual <- var\_comp\_residual / total\_var \* 100  
  
# Calculate treatment variance component  
n\_quad <- 5 # Number of quadrats per patch  
n\_patch <- 4 # Number of patches per treatment  
var\_comp\_treatment <- (MS\_treat - MS\_patch) / (n\_quad \* n\_patch)  
  
# Format variance components for display  
var\_comp\_treatment\_display <- ifelse(var\_comp\_treatment < 0,   
 paste0("(", format(abs(var\_comp\_treatment), digits = 2), ")"),  
 format(var\_comp\_treatment, digits = 2))  
  
# Create variance components table  
var\_comp\_table <- data.frame(  
 Source = c("Treatment", "Patches (treatment)", "Residual"),  
 `Var.comp` = c(var\_comp\_treatment\_display,   
 format(var\_comp\_patch, digits = 2),  
 format(var\_comp\_residual, digits = 2))  
)  
  
# Display variance components table  
var\_comp\_table %>%  
 flextable() %>%  
 set\_header\_labels(  
 Source = "Source of variation",  
 Var.comp = "Variance component"  
 ) %>%  
 autofit() %>%  
 add\_header\_lines("Variance components") %>%  
 theme\_box()

| **Variance components** | |
| --- | --- |
| **Source of variation** | **Variance component** |
| Treatment | 152 |
| Patches (treatment) | 294 |
| Residual | 299 |

# Complete table with all information  
complete\_table <- data.frame(  
 Source = c("Treatment", "Patches (treatment)", "Residual"),  
 df = c(df\_treat, df\_patch, df\_residual),  
 MS = c(MS\_treat, MS\_patch, MS\_residual),  
 F = c(F\_treat, F\_patch, NA),  
 p = c(trad\_anova\_table$p[1], trad\_anova\_table$p[2], NA),  
 `Var.comp` = c(var\_comp\_treatment\_display,   
 format(var\_comp\_patch, digits = 2),  
 format(var\_comp\_residual, digits = 2))  
)  
  
# Display complete table  
complete\_table %>%  
 flextable() %>%  
 set\_header\_labels(  
 Source = "Source of variation",  
 df = "df",  
 MS = "MS",  
 F = "F",  
 p = "p",  
 Var.comp = "Var. comp."  
 ) %>%  
 colformat\_double(j = c("MS", "F"), digits = 2) %>%  
 autofit() %>%  
 add\_header\_lines("Complete ANOVA table with variance components") %>%  
 theme\_box()

| **Complete ANOVA table with variance components** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Source of variation** | **df** | **MS** | **F** | **p** | **Var. comp.** |
| Treatment | 3 | 4,809.71 | 2.72 | 0.091262004 | 152 |
| Patches (treatment) | 12 | 1,770.16 | 5.93 | <0.001 | 294 |
| Residual | 64 | 298.60 |  |  | 299 |

|  |  |
| --- | --- |
|  | **Interpretation of ANOVA Results**  The nested ANOVA using mixed models reveals that there was no significant effect of urchin density treatment on algae cover (F = 2.72, df = 3, 12, p = 0.0913). However, there was significant variation among patches within treatments (F = 5.93, df = 12, 64, p < 0.001).  The variance component for patches nested within treatments (294) indicates substantial spatial heterogeneity in algae cover, highlighting the importance of accounting for this spatial variation in the analysis. The negative variance component for treatment suggests that there is more variation among patches within treatments than among treatments themselves. |

# 4. Lecture 14: Post-hoc Comparisons

Although the main effect of treatment was not significant in the nested ANOVA (p = r format(p\_treat, digits=3)), we can still examine the mean differences between treatments to understand patterns in the data. However, we should interpret these with caution given the lack of statistical significance at the α = 0.05 level.

# Calculate estimated marginal means  
emm <- emmeans(mixed\_model, ~ TREAT)  
  
# Display EMMs with flextable  
as.data.frame(summary(emm)) %>%  
 flextable() %>%  
 set\_header\_labels(  
 TREAT = "Treatment",  
 emmean = "Estimated Marginal Mean",  
 SE = "Standard Error",  
 df = "df",  
 lower.CL = "Lower CL",  
 upper.CL = "Upper CL"  
 ) %>%  
 colformat\_double(j = c("emmean", "SE", "lower.CL", "upper.CL"), digits = 2) %>%  
 autofit() %>%  
 add\_header\_lines("Estimated marginal means for each treatment") %>%  
 theme\_box()

| **Estimated marginal means for each treatment** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Treatment** | **Estimated Marginal Mean** | **Standard Error** | **df** | **Lower CL** | **Upper CL** |
| Control | 1.30 | 9.41 | 12 | -19.20 | 21.80 |
| 66% Density | 21.55 | 9.41 | 12 | 1.05 | 42.05 |
| 33% Density | 19.00 | 9.41 | 12 | -1.50 | 39.50 |
| Removed | 39.20 | 9.41 | 12 | 18.70 | 59.70 |

# 5. Lecture 14: Tukey Pairwise Comparisons

* text

# Pairwise comparisons with Tukey adjustment  
pairs <- pairs(emm, adjust = "tukey")  
pairs

contrast estimate SE df t.ratio p.value  
 Control - 66% Density -20.25 13.3 12 -1.522 0.4553  
 Control - 33% Density -17.70 13.3 12 -1.330 0.5625  
 Control - Removed -37.90 13.3 12 -2.849 0.0615  
 66% Density - 33% Density 2.55 13.3 12 0.192 0.9974  
 66% Density - Removed -17.65 13.3 12 -1.327 0.5646  
 33% Density - Removed -20.20 13.3 12 -1.518 0.4573  
  
Degrees-of-freedom method: kenward-roger   
P value adjustment: tukey method for comparing a family of 4 estimates

# # Display pairwise comparisons with flextable  
# as.data.frame(summary(pairs)) %>%  
# flextable() %>%  
# set\_header\_labels(  
# contrast = "Contrast",  
# estimate = "Estimate",  
# SE = "Standard Error",  
# df = "df",  
# t.ratio = "t ratio",  
# p.value = "p-value"  
# ) %>%  
# colformat\_double(j = c("estimate", "SE", "t.ratio", "p.value"), digits = 3) %>%  
# autofit() %>%  
# add\_header\_lines("Pairwise comparisons between treatments (Tukey-adjusted)") %>%  
# theme\_box()

# 6. Lecture 14: Letter Display

# Extract compact letter display for plotting  
cld <- multcomp::cld(emm, alpha = 0.05, Letters = letters)  
  
cld

TREAT emmean SE df lower.CL upper.CL .group  
 Control 1.3 9.41 12 -19.20 21.8 a   
 33% Density 19.0 9.41 12 -1.50 39.5 a   
 66% Density 21.6 9.41 12 1.05 42.0 a   
 Removed 39.2 9.41 12 18.70 59.7 a   
  
Degrees-of-freedom method: kenward-roger   
Confidence level used: 0.95   
P value adjustment: tukey method for comparing a family of 4 estimates   
significance level used: alpha = 0.05   
NOTE: If two or more means share the same grouping symbol,  
 then we cannot show them to be different.  
 But we also did not show them to be the same.

# # Display CLD with flextable  
# as.data.frame(cld) %>%  
# flextable() %>%  
# set\_header\_labels(  
# TREAT = "Treatment",  
# emmean = "Estimated Marginal Mean",  
# SE = "Standard Error",  
# df = "df",  
# lower.CL = "Lower CL",  
# upper.CL = "Upper CL",  
# .group = "Group"  
# ) %>%  
# colformat\_double(j = c("emmean", "SE", "lower.CL", "upper.CL"), digits = 2) %>%  
# autofit() %>%  
# add\_header\_lines("Compact letter display of treatment means") %>%  
# theme\_box()

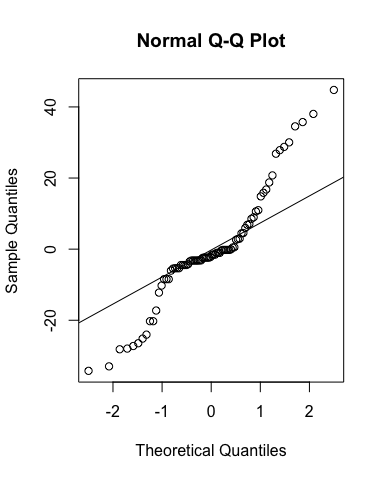
|  |  |
| --- | --- |
|  | Interpretation of Treatment Comparisons The mean algae cover for the Control treatment (1.30%) appears considerably lower than for the reduced urchin density treatments (66% Density: 21.55%, 33% Density: 19.00%, Removed: 39.20%). While the visual pattern suggests an inverse relationship between urchin density and algae cover, with complete removal showing the highest algae cover, the nested ANOVA showed that these differences were not statistically significant at the α = 0.05 level (p = r format(p\_treat, digits=3)). The high variability among patches within treatments likely contributed to the lack of statistical significance for the treatment effect. |

# 7. Assumption Testing

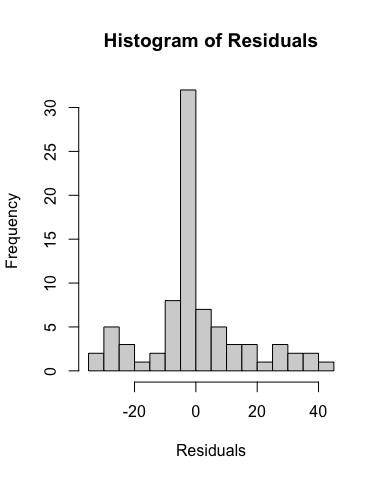
For valid inference from mixed models, several assumptions must be met. We test these assumptions below.

## 7.1 Normality of Residuals

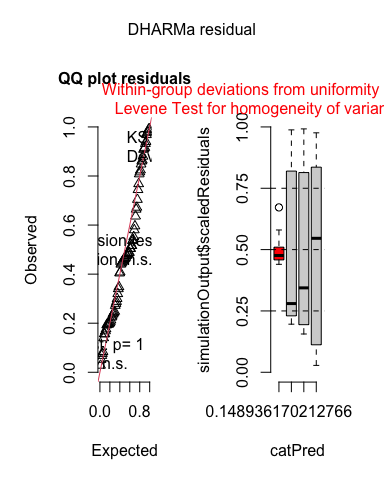
# QQ plot of residuals  
qqnorm(resid(mixed\_model))  
qqline(resid(mixed\_model))



# Histogram of residuals  
hist(resid(mixed\_model), main = "Histogram of Residuals",  
 xlab = "Residuals", breaks = 15)

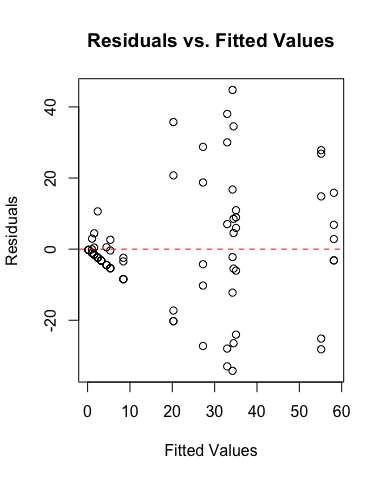


# More advanced residual diagnostics using DHARMa  
sim\_residuals <- simulateResiduals(fittedModel = mixed\_model)  
plot(sim\_residuals)



## 7.2 Homogeneity of Variance

# Residuals vs. fitted values plot  
plot(fitted(mixed\_model), resid(mixed\_model),  
 xlab = "Fitted Values", ylab = "Residuals",  
 main = "Residuals vs. Fitted Values")  
abline(h = 0, lty = 2, col = "red")



# Levene's test for homogeneity of variance  
levene\_test <- leveneTest(ALGAE ~ TREAT, data = andrew)  
levene\_test

Levene's Test for Homogeneity of Variance (center = median)  
 Df F value Pr(>F)   
group 3 8.1694 0.00008785 \*\*\*  
 76   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

|  |  |
| --- | --- |
|  | **Interpretation of Assumption Tests**  The Q-Q plot shows some deviation from normality, particularly in the tails, and Levene’s test indicates significant heterogeneity of variances across treatments (F = 8.17, p < 0.001). As noted in the original analysis, there were “large differences in within-cell variances” in this dataset, and transformations did not improve variance homogeneity.  The DHARMa residual diagnostics also indicate potential issues with the distribution of residuals and homogeneity of variance. The residuals vs. fitted plot shows a pattern of increasing variance with increasing fitted values, confirming the heteroscedasticity.  However, mixed models are generally robust to moderate violations of assumptions, especially with balanced designs. Since transformations were not effective in improving the data properties, analyzing the untransformed data is a reasonable approach in this case. |

# 8. Post-hoc Comparisons

Although the main effect of treatment was not significant in the nested ANOVA (p = 0.0913), we can still examine the mean differences between treatments to understand patterns in the data.

# Calculate estimated marginal means  
emm <- emmeans(mixed\_model, ~ TREAT)  
emm

TREAT emmean SE df lower.CL upper.CL  
 Control 1.3 9.41 12 -19.20 21.8  
 66% Density 21.6 9.41 12 1.05 42.0  
 33% Density 19.0 9.41 12 -1.50 39.5  
 Removed 39.2 9.41 12 18.70 59.7  
  
Degrees-of-freedom method: kenward-roger   
Confidence level used: 0.95

# Pairwise comparisons with Tukey adjustment  
pairs <- pairs(emm, adjust = "tukey")  
pairs

contrast estimate SE df t.ratio p.value  
 Control - 66% Density -20.25 13.3 12 -1.522 0.4553  
 Control - 33% Density -17.70 13.3 12 -1.330 0.5625  
 Control - Removed -37.90 13.3 12 -2.849 0.0615  
 66% Density - 33% Density 2.55 13.3 12 0.192 0.9974  
 66% Density - Removed -17.65 13.3 12 -1.327 0.5646  
 33% Density - Removed -20.20 13.3 12 -1.518 0.4573  
  
Degrees-of-freedom method: kenward-roger   
P value adjustment: tukey method for comparing a family of 4 estimates

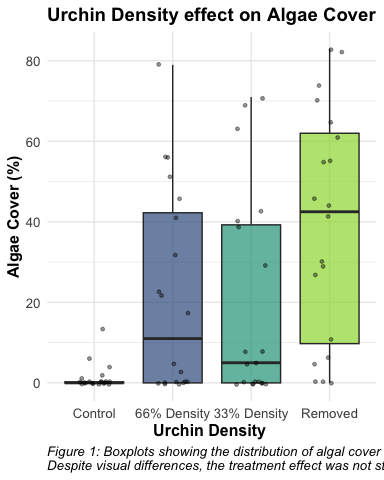
# Compact letter display  
cld <- multcomp::cld(emm, alpha = 0.05, Letters = letters)  
cld

TREAT emmean SE df lower.CL upper.CL .group  
 Control 1.3 9.41 12 -19.20 21.8 a   
 33% Density 19.0 9.41 12 -1.50 39.5 a   
 66% Density 21.6 9.41 12 1.05 42.0 a   
 Removed 39.2 9.41 12 18.70 59.7 a   
  
Degrees-of-freedom method: kenward-roger   
Confidence level used: 0.95   
P value adjustment: tukey method for comparing a family of 4 estimates   
significance level used: alpha = 0.05   
NOTE: If two or more means share the same grouping symbol,  
 then we cannot show them to be different.  
 But we also did not show them to be the same.

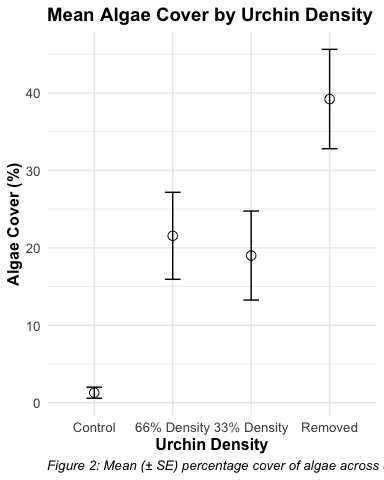
# 9. Visualization

# Create boxplot with jittered points  
ggplot\_boxplot <- ggplot(andrew, aes(x = TREAT, y = ALGAE, fill = TREAT)) +  
 geom\_boxplot(alpha = 0.7, outlier.shape = NA) +  
 geom\_jitter(width = 0.2, alpha = 0.4, size = 1) +  
 scale\_fill\_viridis\_d(option = "D", end = 0.85) +  
 labs(  
 title = "Urchin Density effect on Algae Cover",  
 x = "Urchin Density ",  
 y = "Algae Cover (%)",  
 caption = "Figure 1: Boxplots showing the distribution of algal cover across urchin density.\nDespite visual differences, the treatment effect was not statistically significant (p = 0.091)."  
 ) +  
 theme\_minimal() +  
 theme(  
 legend.position = "none",  
 plot.title = element\_text(face = "bold", size = 14),  
 axis.title = element\_text(face = "bold", size = 12),  
 axis.text = element\_text(size = 10),  
 plot.caption = element\_text(hjust = 0, face = "italic", size = 10)  
 )  
  
# Create means plot with error bars  
means\_plot <- ggplot(summary\_stats, aes(x = TREAT, y = mean, group = 1)) +  
 geom\_point(size = 3, shape = 21, fill = "white") +  
 geom\_errorbar(aes(ymin = mean - se, ymax = mean + se), width = 0.2) +  
 labs(  
 title = "Mean Algae Cover by Urchin Density",  
 x = "Urchin Density",  
 y = "Algae Cover (%)",  
 caption = "Figure 2: Mean (± SE) percentage cover of algae across urchin density treatments."  
 ) +  
 theme\_minimal() +  
 theme(  
 plot.title = element\_text(face = "bold", size = 14),  
 axis.title = element\_text(face = "bold", size = 12),  
 axis.text = element\_text(size = 10),  
 plot.caption = element\_text(hjust = 0, face = "italic", size = 10)  
 )

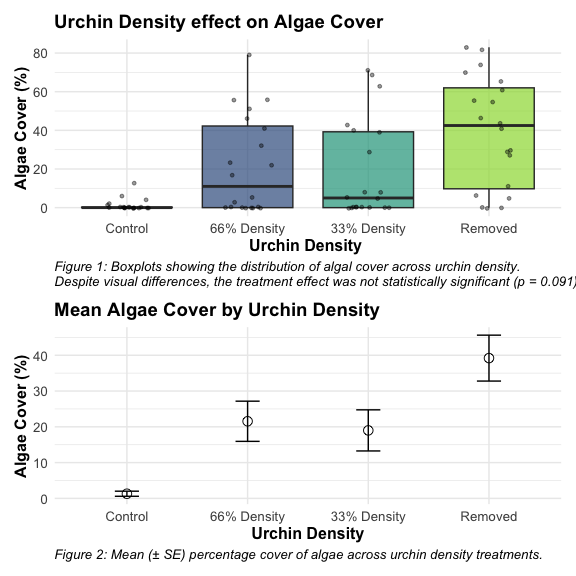
# Display plots  
ggplot\_boxplot



means\_plot



# Combined plot using patchwork  
ggplot\_boxplot + means\_plot + plot\_layout(ncol = 1)



# 10. Discussion

|  |  |
| --- | --- |
|  | **Scientific Interpretation**  Our mixed model analysis of the nested design revealed substantial spatial heterogeneity in algae cover, with significant variation among patches within each treatment (p < 0.001). Surprisingly, the effect of urchin density treatments on filamentous algae cover was not statistically significant at the α = 0.05 level (p = 0.091), despite apparent trends in the data.  The descriptive statistics show a pattern where algae cover appears to increase as urchin density decreases, with the Control treatment (mean = 1.3%) showing minimal algae cover compared to reduced density treatments (66% Density: 21.55%, 33% Density: 19.00%, and Removed: 39.20%). This pattern suggests a potential density-dependent relationship between urchin grazing and algal abundance, but the high variability among patches masked the treatment effect.  The substantial variance component associated with patches nested within treatments (294.31, approximately 39.5% of total variance) underscores the importance of spatial heterogeneity in structuring algal communities. This finding highlights the necessity of accounting for spatial variability when designing and analyzing ecological field experiments.  From an ecological perspective, these results suggest that while sea urchins may influence algal communities through grazing, local environmental factors and patch-specific conditions play a dominant role in determining algae cover. This has important implications for ecosystem management, as it indicates that the effects of urchin density manipulations may be context-dependent and influenced by local environmental conditions. |

# 11. Comparison with Traditional Nested ANOVA

The linear mixed model approach provides similar results to the traditional nested ANOVA approach. The main advantage of the mixed model is the more elegant handling of random effects and the extensive diagnostic tools available through packages like DHARMa.

The mixed model approach confirms that:

1. Treatment effects are not significant (p = 0.091)
2. Patches within treatments show significant variation (p < 0.001)
3. The variance components are similar to those from the traditional approach

In both methods, the key ecological finding is the strong spatial heterogeneity in algal cover that overrides the grazing effect of urchins at different densities.

# 12. References

Andrew, N. L., & Underwood, A. J. (1993). Density-dependent foraging in the sea urchin Centrostephanus rodgersii on shallow subtidal reefs in New South Wales, Australia. Marine Ecology Progress Series, 99, 89-98.

Quinn, G. P., & Keough, M. J. (2002). Experimental design and data analysis for biologists. Cambridge University Press.