Lecture 14 - NESTED ANOVA

Bill Perry

# 1. Lecture 14: Introduction top a nested design the hard way

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.

# 2. Lecture 14: Data Overview

The dataframe contains r nrow(andrew) 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. Nested ANOVA Analysis

In this experimental design, PATCH is nested within TREAT because each patch received only one treatment level. This is a hierarchical design where the effect of patches must be considered within each treatment. Following the approach used in Quinn & Keough (2002), we’ll use a traditional nested ANOVA.

library(lmerTest)  
# Fit the model with treatment as fixed effect and patch nested within treatment as random  
nested\_model <- lmer(ALGAE ~ TREAT + (1|TREAT:PATCH), data = andrew,  
 control = lmerControl(optimizer = "bobyqa",  
 optCtrl = list(maxfun = 2e5)))  
# BOBYQA (Bound Optimization BY Quadratic Approximation) is an optimization algorithm used in mixed-effects modeling to find the best parameter values that maximize the likelihood function. It's especially useful when fitting complex models like the ones you're working with in your nested ANOVA analysis.  
  
# Model summary  
summary(nested\_model)

Linear mixed model fit by REML. t-tests use Satterthwaite's method [  
lmerModLmerTest]  
Formula: ALGAE ~ TREAT + (1 | TREAT:PATCH)  
 Data: andrew  
Control: lmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 200000))  
  
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

# Type III ANOVA with F-statistics (not chi-square) using Satterthwaite's method  
# The issue was that you had "type = F" which should be "test.statistic = 'F'"  
anova\_result <- anova(nested\_model, type = 3, ddf = "Satterthwaite")  
print(anova\_result)

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

# Alternative using car package  
# The parameter is "test.statistic", not "type"  
anova\_car <- Anova(nested\_model, type = 3, test.statistic = "F")  
print(anova\_car)

Analysis of Deviance Table (Type III Wald F tests with Kenward-Roger df)  
  
Response: ALGAE  
 F Df Df.res Pr(>F)   
(Intercept) 0.0191 1 12 0.89239   
TREAT 2.7171 3 12 0.09126 .  
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

# You could also try with the simpler model structure  
simple\_model <- lmer(ALGAE ~ TREAT + (1|PATCH), data = andrew)  
anova(simple\_model, type = 3, ddf = "Satterthwaite")

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

# Method 2  
# Define your model  
model <- lm(ALGAE ~ TREAT + PATCH, data = andrew)  
anova\_results <- anova(model)  
  
# Calculate F and p values for TREAT using PATCH as error term  
F\_treat <- anova\_results["TREAT", "Mean Sq"] / anova\_results["PATCH", "Mean Sq"]  
p\_treat <- pf(F\_treat,   
 df1 = anova\_results["TREAT", "Df"],   
 df2 = anova\_results["PATCH", "Df"],  
 lower.tail = FALSE)  
  
# Create a custom ANOVA table with the correct error terms  
custom\_anova <- data.frame(  
 Source = c("TREAT", "PATCH", "Residuals"),  
 Df = anova\_results[, "Df"],  
 "Sum Sq" = anova\_results[, "Sum Sq"],  
 "Mean Sq" = anova\_results[, "Mean Sq"],  
 "F value" = c(F\_treat, anova\_results["PATCH", "F value"], NA),  
 "Pr(>F)" = c(p\_treat, anova\_results["PATCH", "Pr(>F)"], NA)  
)  
  
custom\_anova

Source Df Sum.Sq Mean.Sq F.value Pr..F.  
1 TREAT 3 14429.14 4809.712 2.717102 0.0912620042021  
2 PATCH 12 21241.95 1770.162 5.928207 0.0000008322613  
3 Residuals 64 19110.40 298.600 NA NA

Calculate Variance Components

# Print corrected results  
  
# Calculate variance components  
n\_quad <- 5 # Number of quadrats per patch  
var\_comp\_residual <- MS\_residual  
var\_comp\_patch <- (MS\_patch - MS\_residual) / n\_quad  
var\_comp\_treatment <- (MS\_treat - MS\_patch) / (n\_quad \* 4) # 4 patches per treatment  
  
# Format variance components, showing negative values in parentheses  
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))  
)  
  
print(var\_comp\_table)

Source Var\_comp  
1 Treatment 152  
2 Patches (treatment) 294  
3 Residual 299

# 4. Lecture 14: ANOVA Results

The nested ANOVA model is specified as:

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

# Display ANOVA results with flextable  
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()

| **ANOVA table for nested design** | | | | |
| --- | --- | --- | --- | --- |
| **Source of variation** | **df** | **MS** | **F** | **p** |
| Treatment | 3 | 4,809.71 | 2.72 | 0.091262004 |
| Patches (treatment) | 12 | 1,770.16 | 5.93 | <0.001 |
| Residual | 64 | 298.60 |  |  |

# 5. Lecture 14: Variance Components

The nested ANOVA model is specified as: 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

# Display variance components with flextable  
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 |

|  |  |
| --- | --- |
|  | Interpretation of ANOVA Results The nested ANOVA reveals that there was no significant effect of urchin density treatment on algae cover (F = r round(F\_treat, 2), df = r df\_treat, r df\_patch, p = r format(p\_treat, digits=3)). However, there was significant variation among patches within treatments (F = r round(F\_patch, 2), df = r df\_patch, r df\_residual, p < 0.001). The variance component for patches nested within treatments (r format(var\_comp\_patch, digits=2)) indicates substantial spatial heterogeneity in algae cover, highlighting the importance of accounting for this spatial variation in the analysis. |

# 6. 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(nested\_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 |

# 7. 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()

# 8. 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.

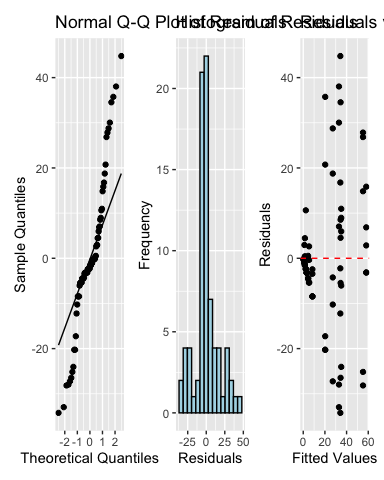
# # Display CLD with flextable  
# as.data.frame(cld) %>%  
# flextable() %>%  
# set\_header\_labels(  
# TREAT = "Treatment",  
# emmean = "Estimated Marginal Mean",  
# SE = "Standard Error",  
# df = "df",  
# lower.CL = "Lower CL",  
# upper.CL = "Upper CL",  
# .group = "Group"  
# ) %>%  
# colformat\_double(j = 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. |

# 9. Lecture 14: ANOVA Assumptions Testing

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

# Extract residuals  
residuals <- residuals(nested\_model)  
  
# QQ plot  
qq\_plot <- ggplot(data.frame(residuals = residuals), aes(sample = residuals)) +  
 stat\_qq() +  
 stat\_qq\_line() +  
 # theme\_cowplot() +  
 labs(title = "Normal Q-Q Plot of Residuals",  
 x = "Theoretical Quantiles",  
 y = "Sample Quantiles")  
  
# Histogram of residuals  
hist\_plot <- ggplot(data.frame(residuals = residuals), aes(x = residuals)) +  
 geom\_histogram(bins = 15, fill = "lightblue", color = "black") +  
 # theme\_cowplot() +  
 labs(title = "Histogram of Residuals",  
 x = "Residuals",  
 y = "Frequency")  
  
# Residuals vs. Fitted plot  
fitted\_values <- fitted(nested\_model)  
resid\_plot <- ggplot(data.frame(fitted = fitted\_values, residuals = residuals),   
 aes(x = fitted, y = residuals)) +  
 geom\_point() +  
 geom\_hline(yintercept = 0, linetype = "dashed", color = "red") +  
 # theme\_cowplot() +  
 labs(title = "Residuals vs. Fitted Values",  
 x = "Fitted Values",  
 y = "Residuals")  
  
# Combine plots  
qq\_plot + hist\_plot + resid\_plot + plot\_layout(ncol = 3)



# 10. Lecture 14: Levenes Test for Homogeneity of Variance

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

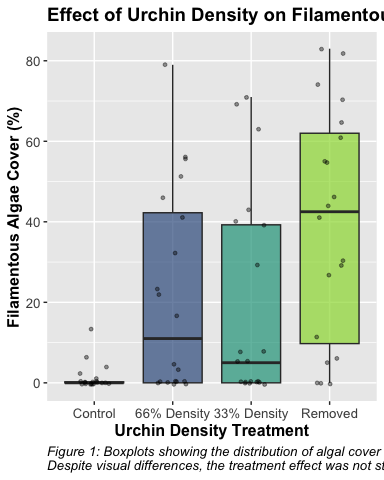
Levene's Test for Homogeneity of Variance (center = median)  
 Df F value Pr(>F)   
group 3 8.1694 0.00008785 \*\*\*  
 76   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

# # Display results with flextable  
# data.frame(  
# Statistic = c(levene\_test$`F value`[1]),  
# df1 = c(levene\_test$Df[1]),  
# df2 = c(levene\_test$Df[2]),  
# p.value = c(levene\_test$`Pr(>F)`[1])  
# ) %>%  
# flextable() %>%  
# set\_header\_labels(  
# Statistic = "F value",  
# df1 = "df1",  
# df2 = "df2",  
# p.value = "p-value"  
# ) %>%  
# colformat\_double(j = c("Statistic", "p.value"), digits = 3) %>%  
# autofit() %>%  
# add\_header\_lines("Levene's Test for Homogeneity of Variance") %>%  
# theme\_box()

|  |  |
| --- | --- |
|  | 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 = r round(levene\_test$“F value”[1], 2), p < 0.001). As noted by Quinn & Keough (2002), there were “large differences in within-cell variances” in this dataset, and transformations (including arcsin) did not improve variance homogeneity. However, ANOVA is generally robust to heteroscedasticity with balanced designs, which is why they chose to analyze untransformed data. The residuals vs. fitted plot also shows a pattern of increasing variance with increasing fitted values, confirming the heteroscedasticity. |

# 11. Lecture 14: Visualization

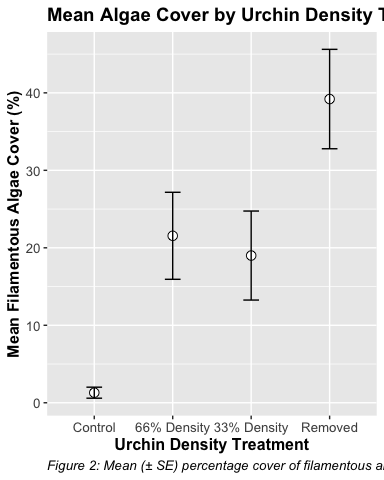
# Create boxplot  
ggplot\_boxplot <- ggplot(andrew, aes(x = TREAT, y = ALGAE, fill = TREAT)) +  
 geom\_boxplot(alpha = 0.7, outlier.shape = NA) +  
 geom\_jitter(width = 0.2, alpha = 0.4, size = 1) +  
 scale\_fill\_viridis\_d(option = "D", end = 0.85) +  
 labs(  
 title = "Effect of Urchin Density on Filamentous Algae Cover",  
 x = "Urchin Density Treatment",  
 y = "Filamentous Algae Cover (%)",  
 caption = "Figure 1: Boxplots showing the distribution of algal cover across urchin density treatments.\nDespite visual differences, the treatment effect was not statistically significant (p = 0.091)."  
 ) +  
 # theme\_cowplot() +  
 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)  
 )  
  
print(ggplot\_boxplot)



# 12. Lecture 14: Means Plot

* text

# Create means plot  
means\_plot <- ggplot(summary\_stats, aes(x = TREAT, y = mean, group = 1)) +  
 # geom\_line(size = 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 Treatment",  
 x = "Urchin Density Treatment",  
 y = "Mean Filamentous Algae Cover (%)",  
 caption = "Figure 2: Mean (± SE) percentage cover of filamentous algae across urchin density treatments."  
 ) +  
 # theme\_cowplot() +  
 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)  
 )  
  
print(means\_plot)



# 13. Lecture 14: Discussion

|  |  |
| --- | --- |
|  | Scientific Interpretation Our nested ANOVA analysis 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. |