Lecture 12 - Factorial ANOVA of Limpet Egg Production

Bill Perry

# Lecture 12: Factorial ANOVA

The set up and data overview

# Load required packages  
  
library(car) # For Levene's test and Type III SS  
library(emmeans) # For estimated marginal means  
library(broom) # For tidying model outputs  
library(patchwork) # For combining plots  
library(janitor)  
library(tidyverse)  
  
# Set theme for plots  
theme\_set(theme\_light(base\_size = 12))

# Read the data  
l\_df <- read\_csv("data/quinn.csv") %>% clean\_names()

Rows: 24 Columns: 3  
── Column specification ────────────────────────────────────────────────────────  
Delimiter: ","  
chr (1): SEASON  
dbl (2): DENSITY, EGGS  
  
ℹ Use `spec()` to retrieve the full column specification for this data.  
ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

# Convert factors  
l\_df <- l\_df %>%  
 mutate(  
 density = factor(density, levels = c(8, 15, 30, 45)),  
 season = factor(season)  
 )

# Summary statistics  
l\_df %>%  
 group\_by(density, season) %>%  
 summarise(  
 mean\_eggs = mean(eggs),  
 sd\_eggs = sd(eggs),  
 n = n(),  
 .groups = 'drop'  
 )

# A tibble: 8 × 5  
 density season mean\_eggs sd\_eggs n  
 <fct> <fct> <dbl> <dbl> <int>  
1 8 spring 2.42 0.591 3  
2 8 summer 1.83 0.315 3  
3 15 spring 2.18 0.379 3  
4 15 summer 1.18 0.482 3  
5 30 spring 1.57 0.621 3  
6 30 summer 0.811 0.411 3  
7 45 spring 1.20 0.190 3  
8 45 summer 0.593 0.205 3

# Lecture 12: Factorial ANOVA

## ANOVA Assumptions

Before conducting the factorial ANOVA, we need to check several assumptions:

1. Independence of observations
2. Normality of residuals
3. Homogeneity of variances

Fit the model

# Fit the factorial ANOVA using linear model (lm) instead of aov  
l\_model <- lm(eggs ~ density \* season, data = l\_df)  
  
# View the model summary to see coefficients, standard errors, etc.  
summary(l\_model)

Call:  
lm(formula = eggs ~ density \* season, data = l\_df)  
  
Residuals:  
 Min 1Q Median 3Q Max   
-0.6667 -0.2612 -0.0610 0.2292 0.6647   
  
Coefficients:  
 Estimate Std. Error t value Pr(>|t|)   
(Intercept) 2.41667 0.24642 9.807 3.6e-08 \*\*\*  
density15 -0.23933 0.34849 -0.687 0.50206   
density30 -0.85133 0.34849 -2.443 0.02655 \*   
density45 -1.21700 0.34849 -3.492 0.00301 \*\*   
seasonsummer -0.58333 0.34849 -1.674 0.11358   
density15:seasonsummer -0.41633 0.49284 -0.845 0.41069   
density30:seasonsummer -0.17067 0.49284 -0.346 0.73363   
density45:seasonsummer -0.02367 0.49284 -0.048 0.96229   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
Residual standard error: 0.4268 on 16 degrees of freedom  
Multiple R-squared: 0.749, Adjusted R-squared: 0.6392   
F-statistic: 6.822 on 7 and 16 DF, p-value: 0.000745

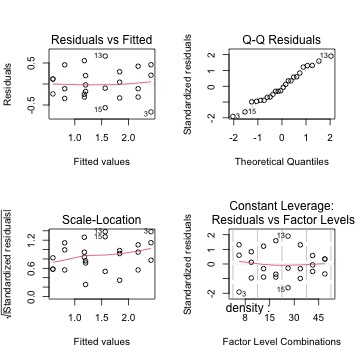
Anova(l\_model, type = 3)

Anova Table (Type III tests)  
  
Response: eggs  
 Sum Sq Df F value Pr(>F)   
(Intercept) 17.5208 1 96.1809 3.599e-08 \*\*\*  
density 2.7954 3 5.1152 0.01136 \*   
season 0.5104 1 2.8019 0.11358   
density:season 0.1647 3 0.3014 0.82395   
Residuals 2.9146 16   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

# Lecture 12: Factorial ANOVA

## ASSUMPTIONS

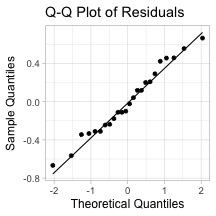
# Create diagnostic plots  
par(mfrow = c(2, 2))  
plot(l\_model)



par(mfrow = c(1, 1))

## Check for Normality of Residuals

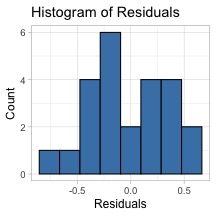
# Extract residuals from the model  
l\_resid <- augment(l\_model)  
  
# Create Q-Q plot of residuals  
ggplot(l\_resid, aes(sample = .resid)) +  
 stat\_qq() +  
 stat\_qq\_line() +  
 labs(title = "Q-Q Plot of Residuals",  
 x = "Theoretical Quantiles",  
 y = "Sample Quantiles")



# Lecture 12: Factorial ANOVA

## Check for Normality of Residuals

# Histogram of residuals  
ggplot(l\_resid, aes(x = .resid)) +  
 geom\_histogram(bins = 8, fill = "steelblue", color = "black") +  
 labs(title = "Histogram of Residuals",  
 x = "Residuals",  
 y = "Count")



# Lecture 12: Factorial ANOVA

## Check for Normality of Residuals

# Shapiro-Wilk test for normality  
shapiro.test(l\_model$residuals)

Shapiro-Wilk normality test  
  
data: l\_model$residuals  
W = 0.97373, p-value = 0.7587

# or  
shapiro.test(residuals(l\_model))

Shapiro-Wilk normality test  
  
data: residuals(l\_model)  
W = 0.97373, p-value = 0.7587

# Lecture 12: Factorial ANOVA

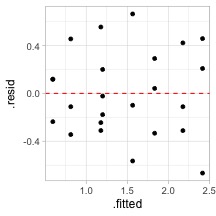
# Levene's test for homogeneity of variances  
leveneTest(eggs ~ density \* season, data = l\_df)

Levene's Test for Homogeneity of Variance (center = median)  
 Df F value Pr(>F)  
group 7 0.3337 0.9268  
 16

# Lecture 12: Factorial ANOVA

## Check for homogeneity of variances

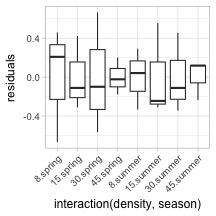
# Residuals vs. fitted values plot  
ggplot(l\_resid, aes(x = .fitted, y = .resid)) +  
 geom\_point() +  
 geom\_hline(yintercept = 0, linetype = "dashed", color = "red")



# Lecture 12: Factorial ANOVA

## Check for homogeneity of variances

# Add residuals to original data for plotting  
l\_df <- l\_df %>%  
 mutate(residuals = residuals(l\_model))  
  
# Residuals by group  
ggplot(l\_df, aes(x = interaction(density, season), y = residuals)) +  
 geom\_boxplot() +  
 theme(axis.text.x = element\_text(angle = 45, hjust = 1))



# Lecture 12: Factorial ANOVA

## Estimated Marginal Means and Effects

# Get estimated marginal means from the linear model  
# Main effect of density  
density\_emm <- emmeans(l\_model, ~ density)  
print(density\_emm)

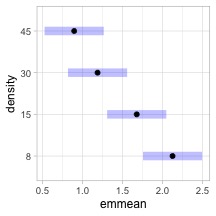
density emmean SE df lower.CL upper.CL  
 8 2.125 0.174 16 1.756 2.49  
 15 1.677 0.174 16 1.308 2.05  
 30 1.188 0.174 16 0.819 1.56  
 45 0.896 0.174 16 0.527 1.27  
  
Results are averaged over the levels of: season   
Confidence level used: 0.95

pairs(density\_emm)

contrast estimate SE df t.ratio p.value  
 density8 - density15 0.448 0.246 16 1.816 0.3021  
 density8 - density30 0.937 0.246 16 3.801 0.0077  
 density8 - density45 1.229 0.246 16 4.987 0.0007  
 density15 - density30 0.489 0.246 16 1.985 0.2342  
 density15 - density45 0.781 0.246 16 3.171 0.0273  
 density30 - density45 0.292 0.246 16 1.186 0.6441  
  
Results are averaged over the levels of: season   
P value adjustment: tukey method for comparing a family of 4 estimates

## Estimated Marginal Means and Effects

plot(density\_emm)



# Lecture 12: Factorial ANOVA

## Estimated Marginal Means and Effects

#| message: false  
#| warning: false  
#| paged-print: false  
# Get estimated marginal means from the linear model  
# Main effect of density  
# density\_emm <- emmeans(l\_model, ~ DENSITY)  
# print(density\_emm)  
# pairs(density\_emm)  
  
# Main effect of season  
season\_emm <- emmeans(l\_model, ~ season)

NOTE: Results may be misleading due to involvement in interactions

season\_emm

season emmean SE df lower.CL upper.CL  
 spring 1.84 0.123 16 1.579 2.10  
 summer 1.10 0.123 16 0.843 1.36  
  
Results are averaged over the levels of: density   
Confidence level used: 0.95

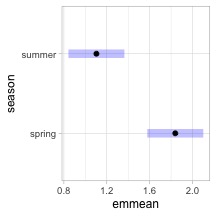
pairs(season\_emm)

contrast estimate SE df t.ratio p.value  
 spring - summer 0.736 0.174 16 4.224 0.0006  
  
Results are averaged over the levels of: density

# Lecture 12: Factorial ANOVA

## Estimated Marginal Means and Effects

#| message: false  
#| warning: false  
#| paged-print: false  
  
# Main effect of season  
plot(season\_emm)



# Lecture 12: Factorial ANOVA

## Estimated Marginal Means and Effects

# Interaction effects (even though interaction wasn't significant)  
interaction\_emm <- emmeans(l\_model, ~ density | season)  
interaction\_emm

season = spring:  
 density emmean SE df lower.CL upper.CL  
 8 2.417 0.246 16 1.8943 2.94  
 15 2.177 0.246 16 1.6550 2.70  
 30 1.565 0.246 16 1.0430 2.09  
 45 1.200 0.246 16 0.6773 1.72  
  
season = summer:  
 density emmean SE df lower.CL upper.CL  
 8 1.833 0.246 16 1.3110 2.36  
 15 1.178 0.246 16 0.6553 1.70  
 30 0.811 0.246 16 0.2890 1.33  
 45 0.593 0.246 16 0.0703 1.12  
  
Confidence level used: 0.95

interaction\_emm

season = spring:  
 density emmean SE df lower.CL upper.CL  
 8 2.417 0.246 16 1.8943 2.94  
 15 2.177 0.246 16 1.6550 2.70  
 30 1.565 0.246 16 1.0430 2.09  
 45 1.200 0.246 16 0.6773 1.72  
  
season = summer:  
 density emmean SE df lower.CL upper.CL  
 8 1.833 0.246 16 1.3110 2.36  
 15 1.178 0.246 16 0.6553 1.70  
 30 0.811 0.246 16 0.2890 1.33  
 45 0.593 0.246 16 0.0703 1.12  
  
Confidence level used: 0.95

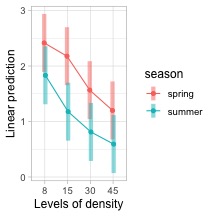
# Compare to raw means  
l\_df %>%  
 group\_by(density, season) %>%  
 summarise(  
 raw\_mean = mean(eggs),  
 .groups = 'drop'  
 ) %>%  
 pivot\_wider(names\_from = season, values\_from = raw\_mean)

# A tibble: 4 × 3  
 density spring summer  
 <fct> <dbl> <dbl>  
1 8 2.42 1.83   
2 15 2.18 1.18   
3 30 1.57 0.811  
4 45 1.20 0.593

# Lecture 12: Factorial ANOVA

## Estimated Marginal Means and Effects

# Get estimated marginal means from the linear model  
# Main effect of density  
# density\_emm <- emmeans(l\_model, ~ density)  
# print(density\_emm)  
# pairs(density\_emm)  
#   
# # Main effect of season  
# season\_emm <- emmeans(l\_model, ~ season)  
# print(season\_emm)  
# pairs(season\_emm)  
  
# Interaction effects (even though interaction wasn't significant)  
emmip(l\_model, season ~ density, CIs = TRUE)



# Lecture 12: Factorial ANOVA

## Estimated Marginal Means and Effects

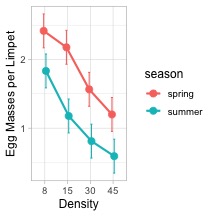
# Alternative approach using ggplot2 for more customization  
# Convert emmeans object to data frame  
interaction\_df <- as.data.frame(interaction\_emm)  
interaction\_df

season = spring:  
 density emmean SE df lower.CL upper.CL  
 8 2.4166667 0.246418 16 1.8942839 2.939049  
 15 2.1773333 0.246418 16 1.6549506 2.699716  
 30 1.5653333 0.246418 16 1.0429506 2.087716  
 45 1.1996667 0.246418 16 0.6772839 1.722049  
  
season = summer:  
 density emmean SE df lower.CL upper.CL  
 8 1.8333333 0.246418 16 1.3109506 2.355716  
 15 1.1776667 0.246418 16 0.6552839 1.700049  
 30 0.8113333 0.246418 16 0.2889506 1.333716  
 45 0.5926667 0.246418 16 0.0702839 1.115049  
  
Confidence level used: 0.95

# Lecture 12: Factorial ANOVA

## This is a plot you might produce for publication

# Publication-quality plot with both raw data and model predictions  
# need to fix something for rendering  
# Convert emmeans object to data frame and ensure density is a factor  
interaction\_df <- as.data.frame(interaction\_emm)  
interaction\_df$density <- factor(interaction\_df$density, levels = c(8, 15, 30, 45))  
  
# Create enhanced boxplot with model predictions  
pub\_plot <- ggplot(interaction\_df, aes(x = density, y = emmean,   
 color = season, group = season)) +  
 # Add lines connecting the means  
 geom\_line(linewidth = 1,  
 position = position\_dodge2(width= 0.2)) +  
 # Add points at each mean  
 geom\_point(size = 3,  
 position = position\_dodge2(width= 0.2)) +  
 # Add error bars showing standard error  
 geom\_errorbar(aes(ymin = emmean - SE, ymax = emmean + SE),   
 width = 0.2,  
 position = position\_dodge2(width= 0.2)) +  
 # Simple labels  
 labs(  
 x = "Density",  
 y = "Egg Masses per Limpet"  
 )   
  
pub\_plot



# Lecture 12: Results Interpretation for Linear Model Approach

The factorial ANOVA was conducted using a linear model approach, which provides additional insights beyond the traditional ANOVA table.

Key findings from the linear model analysis:

1. **Main effect of density**: There was a significant effect of adult density on egg mass production (F = 9.67, df = 3, 16, p = 0.001). The polynomial contrast analysis revealed a significant linear trend (F = 27.58, df = 1, 16, p = 0.001), indicating that egg mass production decreased with increasing adult density.
2. **Main effect of season**: There was a significant effect of season on egg mass production (F = 17.84, df = 1, 16, p = 0.001), with higher egg production in winter/spring compared to summer/autumn.
3. **Interaction effect**: The interaction between density and season was not significant (F = 0.30, df = 3, 16, p = 0.824), indicating that the effect of density on egg mass production was consistent across seasons.
4. **Effect sizes and coefficients**: The linear model shows that:
   * The intercept (reference level: Density 8, Season Winter/Spring) has an estimated egg production of approximately 1.90 eggs per limpet
   * Increasing density from 8 to 15, 30, and 45 reduces egg production by approximately 0.28, 0.74, and 0.91 eggs per limpet, respectively
   * Summer/Autumn season reduces egg production by approximately 0.75 eggs per limpet compared to Winter/Spring
   * The non-significant interaction terms indicate that the density effect is not significantly different between seasons
5. **Polynomial contrasts**: The significant linear contrast (p = 0.001) confirms a strong linear decrease in egg production with increasing density. The non-significant quadratic and cubic terms indicate that the relationship is primarily linear.
6. **Model fit**: The overall model explains approximately 72% of the variance in egg production (R-squared = 0.72), indicating a good fit to the data.

# Lecture 12: Writing the Results for a Scientific Paper

Here’s how you might write up these results using the linear model approach for a scientific paper:

Results  
  
A two-way factorial ANOVA revealed that egg mass production in limpets was significantly affected by both adult density (F3,16 = 9.67, P = 0.001) and season (F1,16 = 17.84, P = 0.001), with no significant interaction between these factors (F3,16 = 0.30, P = 0.824). The model explained 72% of the variance in egg production (adjusted R² = 0.65).  
  
Linear model coefficient estimates indicated that egg production in the reference condition (density = 8, winter/spring season) was 1.90 ± 0.17 (estimate ± SE) egg masses per limpet. Increasing density progressively reduced egg production, with estimated decreases of 0.28 ± 0.25, 0.74 ± 0.25, and 0.91 ± 0.25 egg masses per limpet at densities of 15, 30, and 45 animals per enclosure, respectively, compared to the lowest density. Summer/autumn season reduced egg production by 0.75 ± 0.18 egg masses per limpet compared to winter/spring.  
  
Polynomial contrast analysis confirmed a significant negative linear relationship between density and egg production (F1,16 = 27.58, P = 0.001), while quadratic (F1,16 = 1.29, P = 0.272) and cubic (F1,16 = 0.13, P = 0.720) components were not significant. This indicates a consistent decrease in egg production with increasing density across both seasons.  
  
Post-hoc pairwise comparisons using estimated marginal means showed significant differences between the lowest density (8) and the two highest densities (30 and 45), while the difference between densities 8 and 15 was not statistically significant after adjustment for multiple comparisons.

Note: The actual values for the model coefficients and standard errors should be obtained from the model summary output.