Worksheet 8 - BACI designs

Question 1 - Simple BACI design

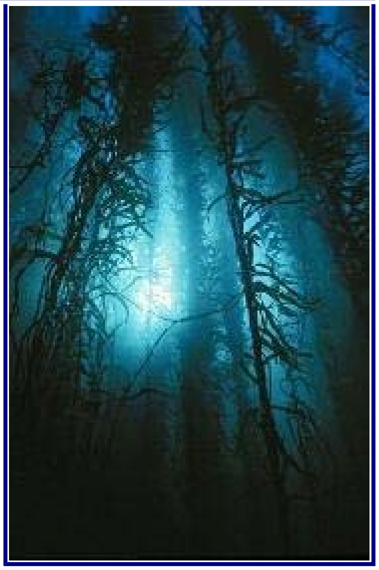
As part of assessing the impact of a nuclear power plant in southern California, data were collected from a location near the outfall pipe, which released heated seawater into the ocean, and at a control location a few km along the coast. The species that was of most concern when the plant was built, was the Giant Kelp, a large brown seaweed growing to a height of 20m, and providing a major habitat for a range of fish. The monitoring program collected data on the density of these algae, by counting the number of plants in large (10 m x 10 m) quadrats on the sea floor of two rocky reefs, one near the discharge point (Impact) and one situated approximately 10 km along the coast (Control). Monitoring started in 1981, before the outfall had been constructed, providing 9 quarterly samples Before the thermal discharge began, and continued for a further 11 quarters After the discharge.

Format of Nuclear power plant data set

Before the discharge									
Control	10.4	10.3	8.6	14.1	13.1	14.1	8.9	6.3	5.8
Impact	4.2	3.4	6.3	8.3	5.8	8.3	4.8	3.5	3.3

After the discharge											
Control	16.5	36.9	43.0	41.4	38.3	35.5	28.2	12.1	9.3	7.6	6.6
Impact	2.0	1.8	1.5	1.6	2.8	2.9	2.5	0.9	0.8	0.7	0.6

Each column represents the number of plants per $100m^2$ for 9 quarters before and 11 quarters after the impact.



Q1-1. One way to determine whether there has been an impact of the nuclear power plant is to examine whether the mean difference in algae density between control and impact sites differs between the before and after treatment.

a. What sort of statistical test would be appropriate for such an analysis?

b. What are the variables?

c. What are the replicates?

d. What are the assumptions of this test?

Q1-2. Generate the appropriate data set. Here is a suggested set of procedures;

a. Generate a factorial variable listing the 'Before' and 'After' for the 9 quarters before and 11 after the impact respectively (HINT)

b. Generate a numeric variable containing the algae densities within control sites(HINT)

c. Generate a numeric variable containing the algae densities within impact sites(HINT)

d. Combine the BA variable and the control and impact variables into a single data frame (data set) (HINT)

e. Generate a numeric variable containing the difference in algae densities between each pair of control and impact sites(HINT)

Q1-3. Recall the assumptions of a t-test or ANOVA and test and comment on the these assumptions. (HINT)

Q1-4. How could the data be made to conform to the test assumptions? Note, if transformations are required, it is probably because the original observations for both control and impact quarters were skewed. Therefore, transform the original variables (control and impact) and recreate the dependent variable used in the analysis (logCI).(HINT)

Q1-5. Perform the **single factor ANOVA** (HINT) or **t-test** (HINT) using the transformed data. Check the diagnostics (primarily residual plot, HINT). If these summaries do not reveal any additional assumption violations, examine the hypothesis test output (HINT) and generate a **bar graph**. Note, since graphical displays do not have underlying distributional assumptions (c.f. ANOVA model), the untransformed data should be used in the construction of the bar graph. It is easier for us humans to interpret raw data.

Q1-6. Based on the analysis and graph, did the power station have a significant impact on Giant Kelp plants?

Question 2 - Split-plot BACI design

In the previous question, we analysed the data from the nuclear power station example, using either a ttest or a one-way ANOVA, with the dependent variable as the difference in kelp density between Control and Impact locations (i.e., a BACI analysis). Now, we will revisit that analysis, using instead the full ANOVA model. The advantage of the full model is that we can deal with situations in which there are multiple control and/or impact locations.

Format of songs.csv data file

BA	TIME	CI	KELP
Before	1	Control	10.4
Before	2	Control	10.3

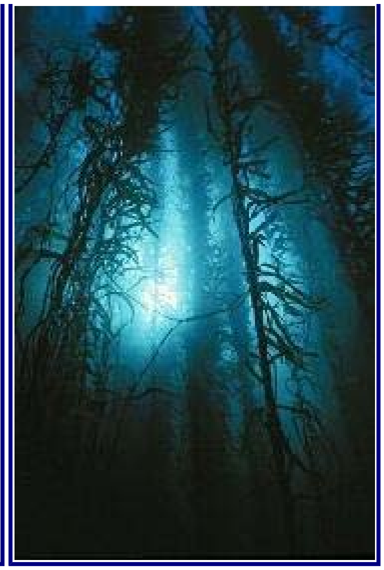
Before	3	Control	8.6
Before	4	Control	14.1
Before	5	Control	13.1
Before	6	Control	14.1
..	..	..	..

BA Categorical listing of whether the sampling quarter was before (Before) or after (After) the impact (introduction of the power plant). Factor A (between plot factor).

TIME Listing of the sampling quarters. These are the plots (Factor B) and are nested within the Before and After sampling quarters. Numbers in this column represent numerical labels given to each plate.

CI Categorical listing for the location (Control = control site, Impact = impact site). Factor C (within plot factor)

KELP Density of kelp (#/100m$_2$) measured. Response variable.



Open the songs data file. HINT. Notice that the TIME variable contains only numbers. Make sure that you define this variable as a factor (HINT)

Q2-1. This now represents a split-plot design. What are the null hypotheses being tested, and what are the correct MS terms to be used as the denominators in each of the F-ratio calculations?

a. H_0 Main Effect 1 (Factor A):

F-ratio = MS_{BA}/MS (choose correct option)

b. H_0 Nested effect (Factor B):

F-ratio = MS_{TIME}/MS (choose correct option)

c. H_0 Main Effect 2 (Factor C):

F-ratio = MS_{CI}/MS (choose correct option)

d. H_0 Main Effect 3 (Factor A:C):

F-ratio = MS_{CI}/MS (choose correct option)

Q2-3. Of these hypotheses, which is of greatest interest to the aims of this study? (BA, TIME, CI, BA:CI)

Q2-4. What are the assumptions associated with testing this hypothesis? How might they be tested? If need be transform data?

Q2-5. What are the replicates for this hypothesis?

Q2-6. Perform a **split-plot ANOVA** (HINT), and complete the following table (HINT). To obtain the hypothesis test for the random factor (Factor B: PLATE), examine the full anova table as if all factors were fixed and thus all terms are tested against the overall residuals, HINT)

Source of variation	df	Mean Sq	F-ratio	P-value
BA				
TIME				
CI				
BA:CI				
Residuals				

Note that the main test of interest is the interaction. You might not expect to find any difference in the density of kelp between control and impact sites before the impact (power station), but you might expect that there would be a difference after the impact - hence an interaction between before-after and control-impact. Note also that the test of this interaction gives the same degrees of freedom, F-ratio and P-value as is achieved via the simple ANOVA from Q1 above!

Q2-7. Construct an **interaction plot** to accompany these results. HINT

Q2-8. What conclusions would you draw from the analysis (and graph)? Does it concur with the outcome from question 1?

Q2-9. What options are available for increasing the power of this particular sampling program?

Welcome to the end of Worksheet8!