

# Group 2: Bokashi/Compost Bioremediation

Marc Los Huertos

2025-04-26

## Introduction

This document performs a Before-After-Control-Impact (BACI) analysis using simulated data. The goal is to detect whether an environmental impact causes a change relative to a control site.

## Hypotheses

## Data Set

```
# Load the data
fakedata.csv = "/home/mwl04747/RTricks/00_Project_Group_Demos/Group2_FakeData.csv"
group1data = read.csv(fakedata.csv, header = TRUE)
str(group1data)

## 'data.frame':    54 obs. of  5 variables:
## $ Sample.ID      : chr  "B1ZB" "B2ZB" "B3ZB" "B1FB" ...
## $ Treatment      : chr  "Bokashi" "Bokashi" "Bokashi" "Bokashi" ...
## $ Before.After   : chr  "Before" "Before" "Before" "Before" ...
## $ Concentraion..ppm.: int  0 0 0 5 5 5 10 10 10 0 ...
## $ Absorption     : num  0.045 0.021 0.034 0.241 0.223 0.234 0.461 0.452 0.443 0.056 ...
names(group1data)

## [1] "Sample.ID"          "Treatment"          "Before.After"
## [4] "Concentraion..ppm." "Absorption"
unique(group1data$Treatment)

## [1] "Bokashi"           "Compost"           "Water (control)"
unique(group1data$Before.After)

## [1] "Before" "After"
```

## Removing Water – since that is a correction, not a factor

```
# Remove water
group1data <- group1data %>%
  filter(Treatment != "Water (control)")
# Remove the "Water" treatment
```

```

# Check the data
unique(group1data$Treatment)

## [1] "Bokashi" "Compost"

# Make sure site and time are factors
group1data <- group1data %>%
  mutate(Treatment = factor(Treatment),
         Before.After = factor(Before.After),
         Concentraion..ppm. = factor(Concentraion..ppm.))

# Check the data
str(group1data)

## 'data.frame': 36 obs. of 5 variables:
## $ Sample.ID : chr "B1ZB" "B2ZB" "B3ZB" "B1FB" ...
## $ Treatment : Factor w/ 2 levels "Bokashi","Compost": 1 1 1 1 1 1 1 1 1 2 ...
## $ Before.After : Factor w/ 2 levels "After","Before": 2 2 2 2 2 2 2 2 2 2 ...
## $ Concentraion..ppm.: Factor w/ 3 levels "0","5","10": 1 1 1 2 2 2 3 3 3 1 ...
## $ Absorption : num 0.045 0.021 0.034 0.241 0.223 0.234 0.461 0.452 0.443 0.056 ...

group1data[sample(1:nrow(group1data), 8), ]

## Sample.ID Treatment Before.After Concentraion..ppm. Absorption
## 31 C1FA Compost After 5 0.261
## 15 C3FB Compost Before 5 0.316
## 14 C2FB Compost Before 5 0.354
## 3 B3ZB Bokashi Before 0 0.034
## 10 C1ZB Compost Before 0 0.056
## 18 C3TB Compost Before 10 0.539
## 22 B1FA Bokashi After 5 0.153
## 11 C2ZB Compost Before 0 0.067

```

## Summary Stats

Fake data isn't really working – since you don't have any variance with ppm. We are not going reporting absorbance values, but I'll do this now, since it looks like there is some variance there.

```

# Summarize the data
summary_stats <- group1data %>%
  group_by(Treatment, Before.After, Concentraion..ppm.) %>%
  summarise(
    mean = mean(Absorption),
    sd = sd(Absorption),
    n = n()
  ) %>%
  ungroup()

## `summarise()` has grouped output by 'Treatment', 'Before.After'. You can
## override using the `.groups` argument.

summary_stats

## # A tibble: 12 x 6
## Treatment Before.After Concentraion..ppm. mean sd n
## <fct> <fct> <fct> <dbl> <dbl> <int>

```

##	1	Bokashi	After	0	0.04	0.0135	3
##	2	Bokashi	After	5	0.142	0.0263	3
##	3	Bokashi	After	10	0.146	0.0108	3
##	4	Bokashi	Before	0	0.0333	0.0120	3
##	5	Bokashi	Before	5	0.233	0.00907	3
##	6	Bokashi	Before	10	0.452	0.00900	3
##	7	Compost	After	0	0.0693	0.0215	3
##	8	Compost	After	5	0.249	0.0157	3
##	9	Compost	After	10	0.329	0.0116	3
##	10	Compost	Before	0	0.056	0.011	3
##	11	Compost	Before	5	0.347	0.0286	3
##	12	Compost	Before	10	0.552	0.0121	3

## Hypothesis Tests

waiting to see hypotheses!

```
# Fit a linear model
model <- lm(Absorption ~ Treatment + Concentraion..ppm. * Before.After, data = group1data)
# Summarize the model
anova(model)
```

```
## Analysis of Variance Table
##
## Response: Absorption
##
##              Df Sum Sq Mean Sq F value    Pr(>F)
## Treatment      1  0.07719  0.077191   67.876 4.396e-09 ***
## Concentraion..ppm.  2  0.62341  0.311705  274.088 < 2.2e-16 ***
## Before.After      1  0.12192  0.121917  107.204 2.980e-11 ***
## Concentraion..ppm.:Before.After  2  0.11515  0.057574   50.626 3.470e-10 ***
## Residuals      29  0.03298  0.001137
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

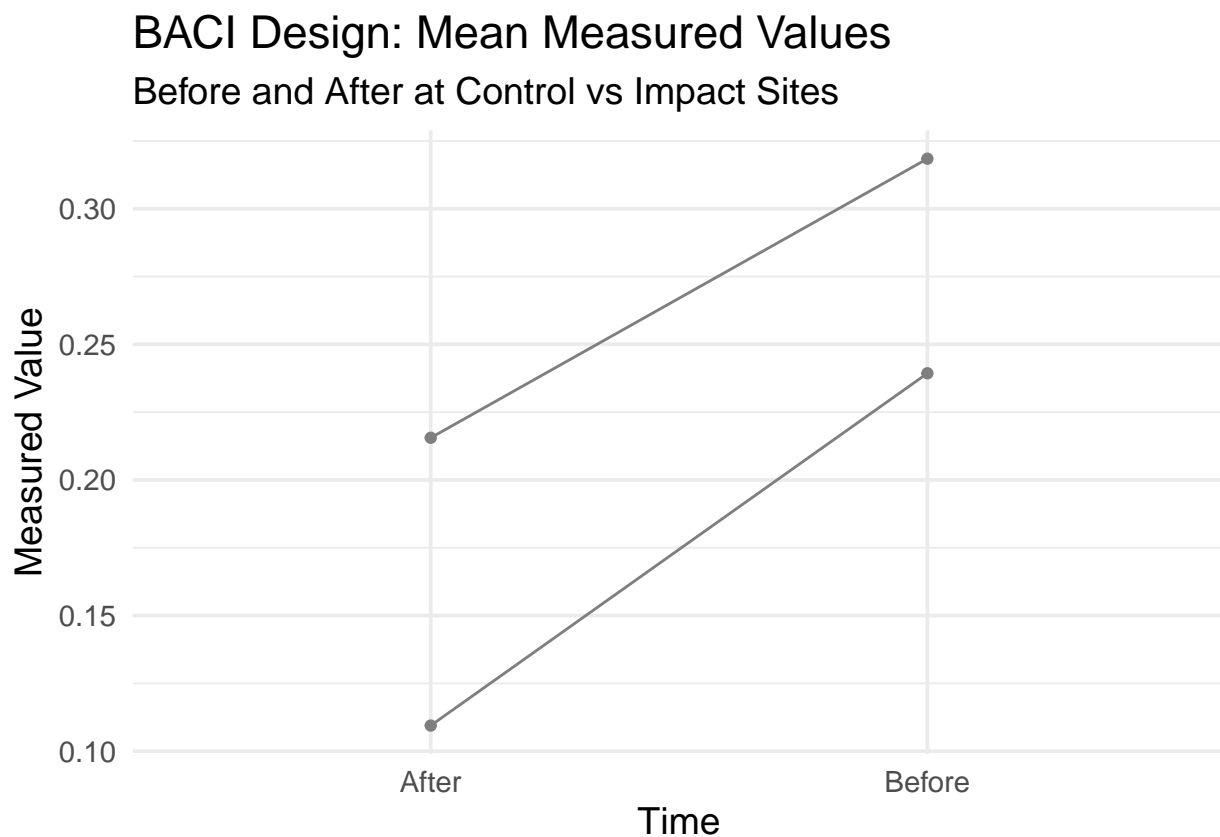
```
model_summary <- tidy(model)
model_summary
```

```
## # A tibble: 7 x 5
##   term                                estimate std.error statistic  p.value
##   <chr>                                <dbl>     <dbl>     <dbl>    <dbl>
## 1 (Intercept)                        0.00836    0.0149      0.562 5.78e- 1
## 2 TreatmentCompost                    0.0926    0.0112      8.24  4.40e- 9
## 3 Concentraion..ppm.5                  0.141    0.0195     7.22  5.91e- 8
## 4 Concentraion..ppm.10                  0.183    0.0195     9.39  2.69e-10
## 5 Before.AfterBefore                   -0.0100    0.0195    -0.514 6.11e- 1
## 6 Concentraion..ppm.5:Before.AfterBefore  0.105    0.0275     3.80  6.84e- 4
## 7 Concentraion..ppm.10:Before.AfterBefore 0.274    0.0275     9.97  7.08e-11
```

```
# Key interaction term:
key_interaction <- coef(model)["Treatmentimpact:Before.Afterafter"]
```

## Plots

```
ggplot(group1data, aes(x = Before.After, y = Absorption, color = Treatment, group = Treatment)) +  
  stat_summary(fun = mean, geom = "line") +  
  stat_summary(fun = mean, geom = "point") +  
  labs(  
    title = "BACI Design: Mean Measured Values",  
    subtitle = "Before and After at Control vs Impact Sites",  
    y = "Measured Value",  
    x = "Time",  
    color = "Treatment"  
  ) +  
  scale_color_manual(values = c("control" = "#1f77b4", "impact" = "#d62728")) +  
  theme_minimal(base_size = 14) +  
  theme(legend.position = "bottom")
```



5. Conclusion The analysis suggests:

There was a decrease in the measured value at the impact site after the event.

The significant site  $\times$  time interaction supports a likely effect of the environmental disturbance.

6. Appendix (Optional) You can improve the model by considering:

Mixed models (lmer) if you have random effects (e.g., multiple sites)

Repeated measures ANOVA

Adding covariates (e.g., weather, seasonality)