Chapter 12

Analysis of BACI experiments

Contents

The suggested citation for this chapter of notes is:

Schwarz, C. J. (2015). *Analysis of BACI experiments*.
In *Course Notes for Beginning and Intermediate Statistics*.
Available at `http://www.stat.sfu.ca/~cschwarz/CourseNotes`. Retrieved 2015-08-20.

12.1 Introduction

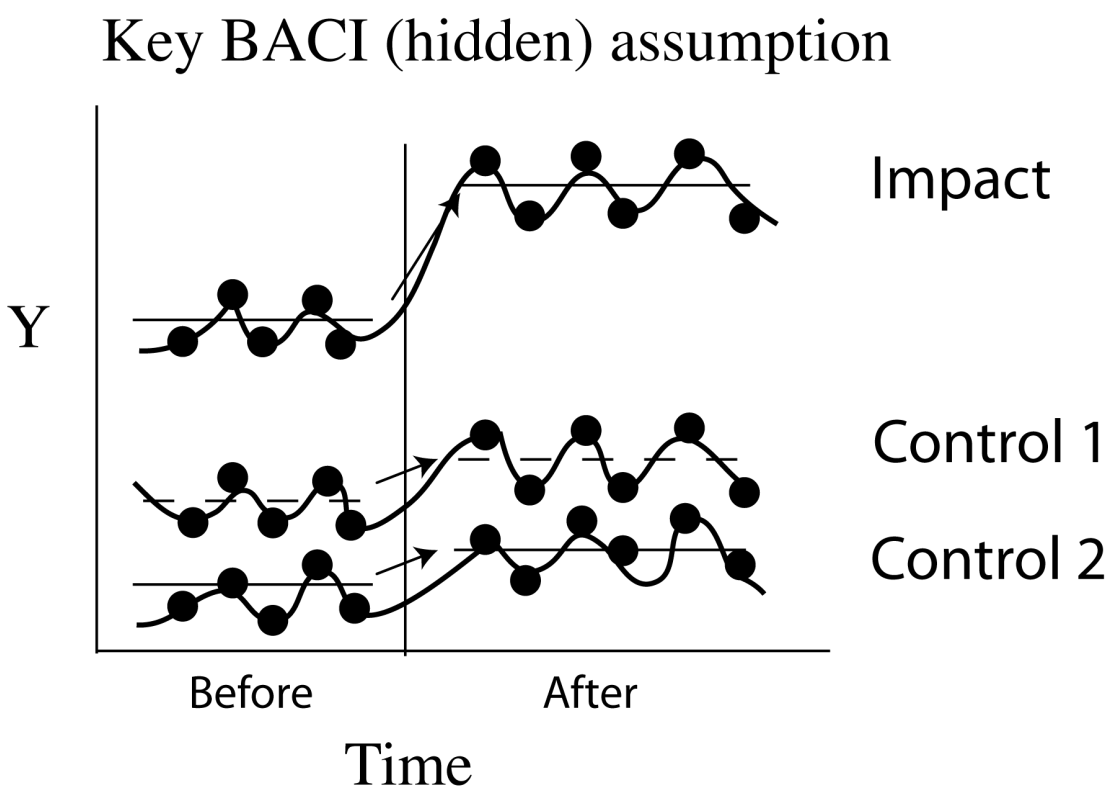
Environmental-impact assessment is often required as part of large projects. A very common design is the Before-After-Control-Impact (BACI) design. In this design, measurements are taken at the treatment (impacted) site and at a control site both before and after the impact occurs.

This design is preferred over a simple Before-After comparison as a change in the response may occur independently of any impact because of temporal effects. For example, precipitation levels may change between the before and after periods and the response may be related to precipitation rather than the impact. Or measurement devices may improve over time and the observed difference is simply an artefact of the measurement process.

By establishing a control site (where presumably no effect of the impact will be felt), the temporal change that occurs in the absence of the impact can be measured Then the differential change in the difference over time is evidence of an environmental impact.

There are many variants of the BACI design – this chapter will review some of the most common. This chapter looks at detecting changes in the mean, but similar designs can be used to detect changes in slopes or proportions or any other response measure. The analysis of these more complex experiments can be difficult – please contact me for details.

A key assumption of BACI designs is that the system is in equilibrium before and after the impact. Schematically, we assume that the overall mean before and after the impact is stable and that the change is quick and immediate in the mean. Schematically, we are assuming a system like:



where there is a step change in the mean immediately after the impact.

Consequently, BACI designs are good for:

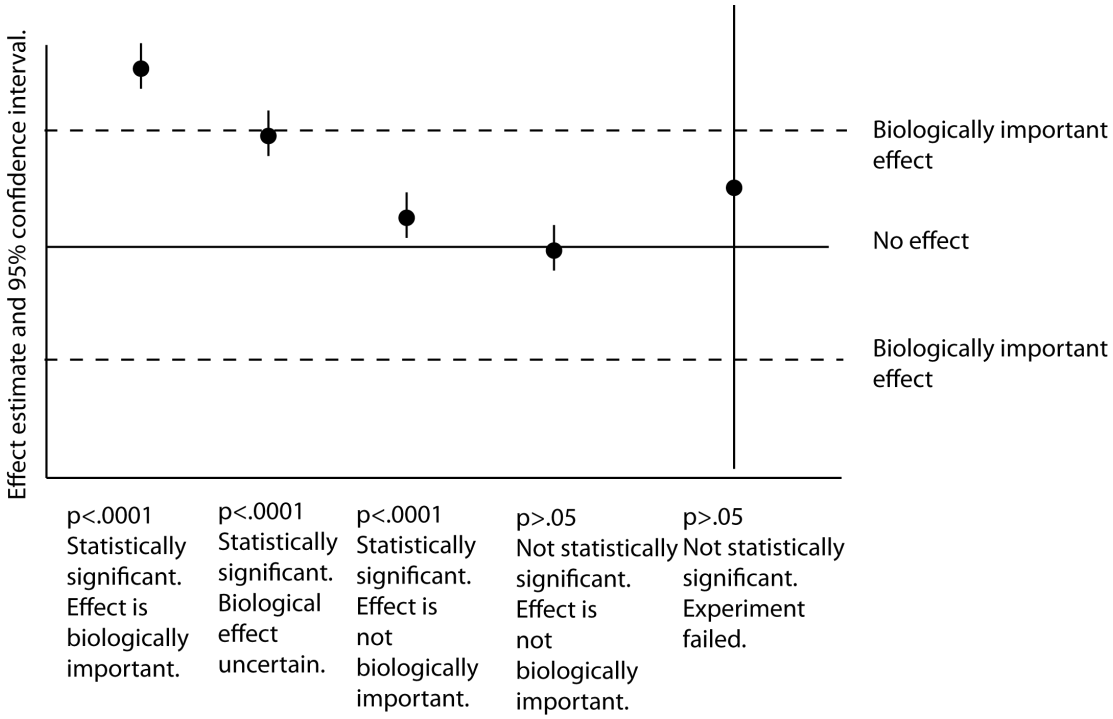
- Detecting large changes after impact.
- Detecting permanent changes after impact.
- Monitoring to protect against disasters.
- Monitoring for changes in the MEAN.

BACI designs are poor for:

- Small potential changes after impact.
- Gradual changes after impact (i.e. not a step change).
- Long term monitoring.
- Monitoring for changes in VARIABILITY.

It is important to recognize in advance that there will be some change over time and because of the impact. It is simply not scientifically defensible to believe that the mean response will be identical to 20 decimal places over time or in the absence of an impact. Rather, the key question is more than

simply hypothesis testing for evidence of any change – regardless of the final p-value from the statistical hypothesis testing, a estimate of the effect size, i.e. an estimate of what is known as the BACI contrast, is required along with a measure of precision. Then this can be examined to see if the effect is detected and if it is biologically important. A conceptual diagram of the possible outcomes from the analysis of an experiment is:



In the left three experiments, the effects are all statistically significant (i.e. the p-value will be less than the α level, usually 0.05). However, the first outcome implies that the effect is also biologically important. In the second outcome, the biological important is in doubt. In this case, more data needs to be collected or precautionary management actions should be implemented. In the third outcome, despite the effect being statistically significant, the actual effect is small and not of biological importance.

Conversely, in the right two experiment, the effect is not statistically significant. However, the fourth case really is no different than the third case. In the fourth case, the effect was not statistically significant, and the 95% confidence intervals indicate that even if the effect was different from zero, it is small. In both the third and fourth cases, no management action is needed. In the last experiment, the estimated effect size has such poor precision (i.e. a large standard error and wide confidence limits) that nothing can be said about the outcome. This failure of this experiment can be due to inadequate sample size, excessive variation in the data that wasn't expected, or other reasons, but really provides no information on how to proceed.

The most difficult part of the above conceptual diagram is the determination of the biologically important effect size! This is NOT a trivial exercise. For example, if you are looking at the impact of warm water from a power plant upon local crab populations, how much of a change in density is biologically important?

An important part of the study design is the question of power and sample size determination, i.e. how many samples or how many years of monitoring are required to detect the biologically important difference as noted above? Far too many studies have inadequate sample sizes and end up with results as in the far right entry in the previous diagram. The proper time for a determination of power and sample size is at the start of the study – the practise of retrospective power analysis provides no new

information (see the chapter on the introduction to ANOVA for details). A key (and most difficult) part of the power/sample size determination is the question of biologically important effect sizes.

Now a days, computer packages are used to analyze these complex designs. As noted in other chapters, simple spreadsheet programs (such as Excel) are NOT recommended. With the use of sophisticated packages such as *JMP* or *SAS*, care needs to be taken to ensure that the analysis requested matches the actual design of the experiment. Always draw a diagram of the experimental protocol. Make sure that your brain is engaged before putting the package in gear!

A review article on the design and analysis of BACI designs is available at:

Smith, E. P. (2002).
BACI design.
Encyclopedia of Environmetrics, 1, 141-148.
`http://www.web-e.stat.vt.edu/vining/smith/B001-_o.pdf`

The standard reference for the analysis of complex BACI designs is:

Underwood, A. J. (1993).
The mechanics of spatially replicated sampling programs to detect environmental impacts in a variable world.
Australian Journal of Ecology 18, 99-116.
`http://http://dx.doi.org/10.1111/j.1442-9993.1993.tb00437.x`

In this chapter, we will look at the design and analysis of four standard BACI designs:

1. Single impact site; single control site; one year before; one year after.
2. Single/multiple impact site; multiple control sites; one year before; one year after.
3. Single impact site; single control site; multiple years before; multiple years after.
4. Single impact site; multiple control sites; multiple years before; multiple years after.

12.2 Before-After Experiments - prelude to BACI designs

As a prelude to the analysis of BACI designs, let us consider a simple before-after experiment, where measurements are taken over time both before and then after an environmental impact.

For example, consider an experiment where two streams were measured at a small hydro electric power project. At each stream multiple quadrats were taken on the steam with a new set of quadrates chosen each year, and the number of invertebrates was counted in each quadrat. After 5 years, the hydro electric plant started up, and an additional 5 years of data were collected. Is there evidence of a change in mean abundance after the plant starts?

Note that this is NOT a BACI design, and is a VERY poor substitute for such. The key problem is that changes may have occurred over time unrelated to the impact, so a "change" may simply be natural (e.g. unrelated to the hydro project). A nice discussion of the perils of using a simple before-after design in non-ecological contexts is found at `http://ssmon.chb.kth.se/safebk/Chp_3.pdf`.

Also, these design typically involve a long time-series. As such, problems related to autocorrelation can creep in. Refer to the chapter on the analysis of trend data for more details. iIn this section, we will ignore autocorrelation which can be a serious problem in long-term studies.

This data will be analyzed several ways. First, we will look at only stream 1, and analyze the yearly averages or the individual data. In these cases, inference is limited to that particular stream and you cannot generalize to other streams in the study. If the same number of quadrates was measured in all years, the two analyses will give identical results.

Second, we will analyze both streams together and analyze the yearly averages or the individual values. Now, because of multiple streams, the results can be generalized to all streams in the project (assuming that the two streams were selected at random from all potentially affected streams. Again, if the number of quadrats was the same for all streams in all years, then the analysis of the yearly averages and the individual values will be identical.

Finally, we will demonstrate how to compute a power analysis of these designs to see how many years of monitoring would be required. The limiting feature for these designs is the year-to-year variation and no amount of subsampling (i.e. sampling more quadrats each year) will reduce this variation.

The raw data is available in the *invert.csv* file available at the Sample Program Library at `http://www.stat.sfu.ca/~cschwarz/Stat-650/Notes/MyPrograms`. The data are imported into *R* in the usual way:

```
invert <- read.csv('invert.csv', header=TRUE, as.is=TRUE, strip.white=TRUE)
# Declare Period as an ordered factor so that sorts in proper order
invert$Period  <- factor(invert$Period, levels=c("Before","After"), ordered=TRUE)
# We also want Year to be categorical variable so make a new variable
invert$YearF   <- factor(invert$Year)
# Stream should also be defined as a factor
invert$StreamF <- factor(invert$Stream)

# It turns out that missing values cause subtle problems in some R
# functions that deal with the fitted object. Sigh... Usually, R
# is sensible enough, but not all package authors did a good job.
dim(invert)
invert <- invert[complete.cases(invert),]
dim(invert)
invert[1:12,]
```

Part of the raw data are shown below:

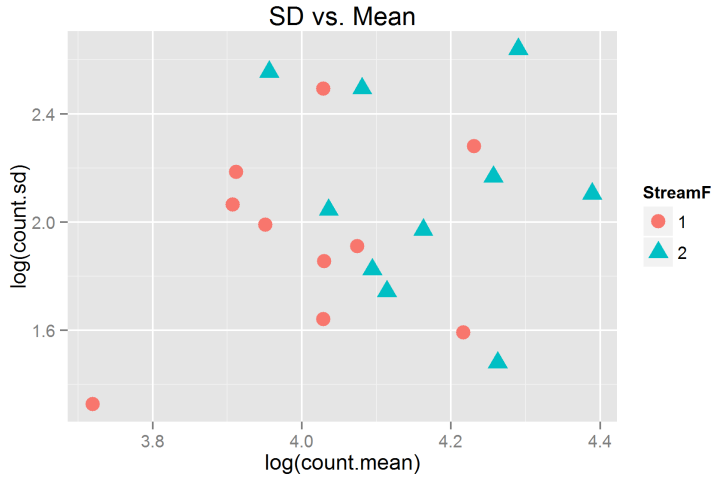
```
[1] 100   7
[1] 90  7
   Year Period Stream quadrat count YearF StreamF
1     1 Before      1       1    58     1       1
2     1 Before      1       2    62     1       1
3     1 Before      1       3    48     1       1
4     1 Before      1       4    72     1       1
5     1 Before      1       5    41     1       1
6     1 Before      2       1    76     1       2
7     1 Before      2       2    73     1       2
9     1 Before      2       4    69     1       2
```

```
10     1 Before       2        5     66      1        2
11     2 Before       1        1     61      2        1
12     2 Before       1        2     48      2        1
13     2 Before       1        3     60      2        1
```

We begin by computing some basic summary statistics in the usual fashion

```
   Year YearF Stream StreamF Period n n.miss count.mean  count.sd
1     1     1      1       1 Before 5      0      56.20 12.091319
2     1     1      2       2 Before 4      0      71.00  4.396969
3     2     2      1       1 Before 5      0      56.20  5.167204
4     2     2      2       2 Before 5      0      60.00  6.204837
5     3     3      1       1 Before 4      0      49.75  7.889867
6     3     3      2       2 Before 5      0      56.60  7.733046
7     4     4      1       1 Before 3      0      50.00  8.888194
8     4     4      2       2 Before 5      0      59.20 12.111978
9     5     5      1       1 Before 4      0      41.25  3.774917
10    5     5      2       2 Before 4      0      52.25 12.867919
11    6     6      1       1  After 5      0      68.80  9.782638
12    6     6      2       2  After 5      0      70.60  8.734987
13    7     7      1       1  After 5      0      52.00  7.314369
14    7     7      2       2  After 5      0      61.20  5.718391
15    8     8      1       1  After 5      0      58.80  6.760178
16    8     8      2       2  After 3      0      73.00 14.000000
17    9     9      1       1  After 5      0      67.80  4.919350
18    9     9      2       2  After 5      0      80.60  8.203658
19   10    10      1       1  After 4      0      56.25  6.396614
20   10    10      2       2  After 4      0      64.25  7.182154
```

The standard deviations appears to be approximate equal over time, and a plot of the $\ln(s)$ vs $\ln(\overline{Y})$:



doesn't show any evidence of relationship between the mean and the variance which is often found when dealing with count data, especially if the counts were smallish.

A plot of the time trend:

appears to show a shift in the mean starting in year 6, but there is large year-to-year variation in both streams. However, the trend lines in both streams appear to be parallel (indicating little evidence of a stream-year interaction).

12.2.1 Analysis of stream 1 - yearly averages

We start with the analysis of a single stream as it is quite common for Before-After studies to have only one "experimental" unit. Of course, inference will be limited to that specific stream chosen, and the results from this stream cannot be extrapolated to other streams in the study.

The multiple quadrats measured every year are pseudo-replicates and cannot be treated as independent observations. As in most cases of pseudo-replication, one approach is to take the average over the pseudo-replicates and analyze the averages. If the number of quadrats taken each year was the same, the analysis of the averages doesn't loose any information and is perfectly valid. If the number of quadrats varies from year-to-year, the analysis of the averages is only approximate, but usually is sufficient.

The averages are computed in the usual way and the actual code to do this is not shown. After the averages are taken, the data has been reduced to 10 values (one for each year in the study) with five measurements taken before the impact and five measurements taken after the impact. Consequently, the simple two-sample t-test can be used to analyze the data. As noted earlier, autocorrelation in the measurements over time has been ignored.

We use the *lm()* function to fit the model. The *lm()* function assumes equal variance for both periods. This could be examined and the *test()* function could be used for the unequal variance t-test but is not shown here.

```
result100 <- lm( count.mean ~ Period, data=mean.invert1)
anova(result100)
```

The ANOVA table from the *lm()* fit is:

```
Analysis of Variance Table

Response: count.mean
          Df Sum Sq Mean Sq F value Pr(>F)
Period     1 252.51 252.506  5.5146 0.0468 *
Residuals  8 366.31  45.789
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

In this case, the two approaches (the equal variance and unequal variance t-test) yield virtuality the same results – weak evidence of an effect with a two-sided p-value just under 0.05.

The multiple comparison procedure is used to estimate the effect size:

```
result100.lsmo <- lsmeans::lsmeans(result100, ~Period)
cat("\n\n Estimated marginal means for each Period \n\n")
summary(result100.lsmo)
cat("Estimated difference between means in Before and After periods\n")
confint(pairs(result100.lsmo, infer=TRUE))
```

```


 Estimated marginal means for each Period

 Period lsmean       SE df lower.CL upper.CL
 Before  50.68 3.026182  8 43.70161 57.65839
 After   60.73 3.026182  8 53.75161 67.70839

Confidence level used: 0.95
Estimated difference between means in Before and After periods
 contrast       estimate       SE df  lower.CL   upper.CL
 Before - After   -10.05 4.279667  8 -19.91893 -0.1810701

Confidence level used: 0.95
```

As expected, the confidence interval for the effect size just barely excludes the value of zero. A power analysis could be conducted to see how many years of monitoring would be required to detect a reasonably sized effect.

12.2.2 Analysis of Stream 1 - individual values

The number of quadrats does vary a bit over time, and so the analysis on the averages for each year is only approximate. A more refined analysis that also provides estimates of the year-to-year and the quadrat-to-quadrat variation is possible.

Again note that quadrats are pseudo-replicates taken in each year. This analysis is functionally equivalent to the analysis of means with sub-sampling discussed in earlier chapters.

A mixed-model analysis is performed. The statistical model is

$$Y = Period\ \ Year(Period)(R)$$

If each year is labelled with a separate value (e.g. using the actual calendar year as the label) rather than year 1 within the before period and year 1 within the after period, the nesting of year within period can be "dropped" and the model can be written as

$$Y = Period\ \ Year(R)$$

Note that these two models are equivalent – we are just letting the software deal with the nesting of years within each period. The *Period* refers to the before vs. after periods and represents the effect of interest.

This model can be fit using the *lme()* function from the *nlme* package, or the *lmer()* function from the *lme4* package. Note that *lmer()* does not report *p*-values as noted in elsewhere in these notes. Be sure to verify that *Year* is defined as factor prior to the fit.

```
# Don't forget that you need a Year as a factor.
# We created YearF as a factor when we readin the data
# We don't need to transform counts because they are largish.
result200 <- lmer(count ~ Period + (1|YearF), data=stream1)
anova(result200, ddf="Kenward-Roger")
```

The test for a period effect (i.e. is there evidence of a difference in the mean response between the before and after period) is found as:

```
Analysis of Variance Table of type 3  with  Kenward-Roger
approximation for degrees of freedom
       Sum Sq Mean Sq NumDF  DenDF F.value  Pr(>F)
Period 310.24  310.24     1 7.9842  5.2373 0.05145 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Because the design is unbalanced, *lme()* estimates the appropriate *df* for the test using a different algorithm than *SAS* and the result can differ slightly.

The *p*-value indicates weak evidence of an effect, similar to that found with the analysis of the averages. Don't live and die by the 0.05 rule – surely the two analyses are saying similar things!

The estimated means in the two periods can be found.
In *R* use the *popMeans()* function in the *doBy* package:

```
# Need to use the lsmeans::lsmeans() notation because there is a
# lsmeans() function in both the lsmeans and lmerTest package.
result200.lsmo <- lsmeans::lsmeans(result200, ~Period)
summary(result200.lsmo)
```

which then gives:

```
 Period   lsmean       SE   df lower.CL upper.CL
 Before 50.83452 3.108632 8.31 43.71214 57.95691
 After  60.78523 3.040144 7.66 53.72071 67.84974

Confidence level used: 0.95
```

Again, *R* uses a different algorithm to estimate the appropriate df for each mean and the results will differ from *SAS*.

The estimated difference in the means between the two periods can also be estimated, and usually is of most interest.
In *R* this is given in the summary table and can be extracted:

```
# Estimate the period effect along with a se
cat("Estimated difference between means in Before and After periods\n")
confint(pairs(result200.lsmo, infer=TRUE))
```

which then gives:

```
Estimated difference between means in Before and After periods
 contrast        estimate       SE   df  lower.CL   upper.CL
 Before - After -9.950706 4.348111 7.98 -19.98092 0.07951333

Confidence level used: 0.95
```

As noted earlier, the df are estimated using a different method in *R* than in *SAS* and results can differ slightly. Both results are similar to that from the analysis of the averages.

There are two sources of variation in this experiment. First the year-to-year variation represents the effects of year-specific factors such as total rainfall, or other random events. Second, the residual variation (or quadrat-to-quadrat) variation measures the variation over quadrats in different parts of the stream. The two estimated variance components are:

In *R* this is extracted using the *VarCorr()* extractor function:

```
# Extract the variance components.
vc <- VarCorr(result200)
vc
```

which then gives:

```
 Groups   Name        Std.Dev.
 YearF    (Intercept) 5.8137
 Residual             7.6922
```

As you will see later when the power analysis is discussed, you can reduce the impact of the quadrat-to-quadrat variation on the estimates by sampling more quadrats each year. However, sampling more quadrats has no impact on the year-to-year random variation and this is often the limiting features of simple before/after designs. As a head up, the year-to-year variation can be controlled by conducting a paired-baci experiment where both impact and control sites are measured on each year of the study.

It should be noted that in the case of a single stream, that year-to-year variation really is a combination of the actual year-specific effects and a year-stream interaction where the time-trend of this particular stream may not be completely parallel with other streams in the study. Of course, with only a single-stream measured, these two sources could be disentangled.

Diagnostic plots (not shown) show no evidence of problems.

12.2.3 Analysis of all streams - yearly averages

We now move to the analysis of both streams. By sampling more than one stream, the inference is greatly strengthened as now the conclusions are more general. Rather than referring to one specific stream, the results now refer to all streams in the population of streams on this project.

As before, we start by averaging over the pseudo-replicates (the multiple-quadrats) measured in each stream-year combination. The statistical model is

$$YearlyAvg = \textit{Period}\ \ \textit{Stream(R)}\ \ \textit{Year(Period)(R)}$$

Once again, if each year is uniquely labelled, the model can be written as:

$$YearlyAvg = \textit{Period}\ \ \textit{Stream(R)}\ \ \textit{Year(R)}$$

As before, the *Period* term represents the difference in the mean between the before and after periods and the other terms represent the random effects of streams and years. The residual variation represents a combination of the stream-year interaction and the quadrat-noise terms, but because the averages over the quadrats were taken, cannot be separated.

This model can be fit using the *lme()* function from the *nlme* package, or the *lmer()* function from the *lme4* package. Note that *lmer()* does not report *p*-values as noted in elsewhere in these notes. Be sure to verify that both *Year* and *Stream* are defined as factors prior to the fit. The *lme()* function uses a different method to estimate the df so results may differ as noted previously. Notice the awkward way in which crossed random factors must be specified!

```
# Note that Stream, Period, and Year all need to be factors
# These were defined as such when the data were read in.
result300 <- lmer(count.mean~ Period + (1|Stream) + (1|YearF), data=mean.invert)
anova(result300, ddf='Kenward-Roger')
```

The test for a period effect (i.e. is there evidence of a difference in the mean response between the before and after period) is found as:

```
Analysis of Variance Table of type 3  with  Kenward-Roger
approximation for degrees of freedom
       Sum Sq Mean Sq NumDF DenDF F.value  Pr(>F)
```

```
Period 52.034  52.034     1     8  5.7309 0.04359 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Because the design is unbalanced, *lme()* estimates the appropriate df for the test using a different algorithm than *SAS* and the result can differ slightly.

The p-value indicates weak evidence of an effect.

The estimated marginal means in the two periods are:

In *R* use the *popMeans()* function in the *doBy* package:

```
# Need to use the lsmeans::lsmeans() notation because there is a
# lsmeans() function in both the lsmeans and lmerTest package.
result300.lsmo <- lsmeans::lsmeans(result300, ~Period)
cat("\n\n Estimated marginal means for each Period \n\n")
summary(result300.lsmo)
```

which then gives:

```


 Estimated marginal means for each Period

 Period lsmean       SE   df lower.CL upper.CL
 Before 55.245 5.423916 1.92 30.94685 79.54315
 After  65.330 5.423916 1.92 41.03185 89.62815

Confidence level used: 0.95
```

Again, *R* uses a different algorithm to estimate the appropriate df for each mean and the results will differ from *SAS*.

The estimated difference in the marginal means between the two periods is usually of most interest. In *R* this is given in the summary table and can be extracted:

```
# Estimate the Period effect
cat("Estimated difference between means in Before and After periods\n")
pairs(result300.lsmo)
confint(pairs(result300.lsmo))
```

which then gives:

```
Estimated difference between means in Before and After periods
 contrast       estimate       SE df t.ratio p.value
```

```
 Before - After  -10.085 4.212725  8  -2.394  0.0436

 contrast       estimate       SE df  lower.CL   upper.CL
 Before - After  -10.085 4.212725  8 -19.79956 -0.3704392

Confidence level used: 0.95
```

As noted earlier, the df are estimated using a different method in *R* than in *SAS* and results can differ slightly.

The confidence interval just barely excludes the value of zero.

There are four sources of variation in this experiment, and not all all can be separated when the analysis is done on the yearly averages. First the year-to-year variation represents the effects of year-specific factors such as total rainfall, or other random events. Second, the stream-to-stream variation measures the effect of the different streams on the mean response. The plot produced earlier shows that there is some stream effect as the lines for the streams are consistently separated. Third, is the stream-year interaction where the individual streams do not have exactly parallel trend lines. Lastly, is the quadrat-to-quadrat variation measures the variation over quadrats in different parts of a stream. Unfortunately, when the yearly averages are analyzed, the last two variance components cannot be separated.

The estimated variance components are:

In *R* this is extracted using the *VarCorr()* extractor function:

```
# Extract the variance components
vc <-VarCorr(result300)
vc
```

which then gives:

```
 Groups   Name        Std.Dev.
 YearF    (Intercept) 6.3109
 Stream   (Intercept) 6.4102
 Residual             3.0132
```

Notice that the residual variance component represents a combination of the stream-year interaction and the quadrat-to-quadrat variation (divided by the average number of quadrats sampled per year). The stream and year variance component are substantially larger than the residual variance!

Diagnostic plots (not shown) show no evidence of problems.

12.2.4 Analysis of all streams - individual values

Finally, we do the analysis of both streams using the individual values.. By sampling more than one stream, the inference is greatly strengthened as now the conclusions are more general. Rather than referring to one specific stream, the results now refer to all streams in the population of streams on this project.

The statistical model is

$$Y = Period\ \ Stream(R)\ \ Year(Period)(R)\ \ Stream*Year(Period)(R)$$

Once again, if each year is uniquely labelled, the model can be written as:

$$Y = Period\ \ Stream(R)\ \ Year(R)\ \ Stream*Year(R)$$

As before, the *Period* term represents the difference in the mean between the before and after periods and the other terms represent the random effects of streams and years. We now can separate out the stream-year interaction from the quadrat-to-quadrat variation.

This model can be fit using the *lme()* function from the *nlme* package, or the *lmer()* function from the *lme4* package. Note that *lmer()* does not report *p*-values as noted in elsewhere in these notes. Be sure to verify that both *Year* and *Stream* are defined as factors prior to the fit. The *lme()* function uses a different method to estimate the *df* so results may differ as noted previously. Notice the awkward way in which crossed random factors must be specified!

```
# Year, Period, Stream must all be defined as factors.
# This was done when the data were read in.
result400 <- lmer(count ~ Period + (1|StreamF)+(1|YearF)+(1|StreamF:YearF), data=inve
anova(result400, ddf="Kenward-Roger")
```

The test for a period effect (i.e. is there evidence of a difference in the mean response between the before and after period) is found as:

```
Analysis of Variance Table of type 3  with  Kenward-Roger
approximation for degrees of freedom
       Sum Sq Mean Sq NumDF  DenDF F.value  Pr(>F)
Period 376.76  376.76     1 7.9986  5.6712 0.04445 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Because the design is unbalanced, *lme()* estimates the appropriate *df* for the test using a different algorithm than *SAS* and the result can differ slightly.

The *p*-value indicates weak evidence of an effect and the result is similar to that found when the yearly averages were analyzed.

The estimated marginal means in the two periods are:

In *R* use the *popMeans()* function in the *doBy* package:

```
# Need to use the lsmeans::lsmeans() notation because there is a
# lsmeans() function in both the lsmeans and lmerTest package.
result400.lsmo <- lsmeans::lsmeans(result400, ~Period)
cat("\n\n Estimated marginal means for each Period \n\n")
summary(result400.lsmo)
```

which then gives:

```


 Estimated marginal means for each Period

 Period   lsmean       SE   df lower.CL upper.CL
 Before 55.21874 5.293016 1.93 31.66222 78.77526
 After  65.22664 5.287031 1.92 41.59450 88.85878

Confidence level used: 0.95
```

Again, *R* uses a different algorithm to estimate the appropriate df for each mean and the results will differ from *SAS*.

and the estimated difference in the marginal means between the two periods is:

In *R* this is given in the summary table and can be extracted:

```
# Estimate the Period effect
cat("Estimated difference between means in Before and After periods\n")
pairs(result400.lsmo)
confint(pairs(result400.lsmo))
```

which then gives:

```
Estimated difference between means in Before and After periods
 contrast       estimate       SE df t.ratio p.value
 Before - After -10.0079 4.202475  8  -2.381  0.0445

 contrast       estimate       SE df  lower.CL   upper.CL
 Before - After -10.0079 4.202475  8 -19.69912 -0.3166744

Confidence level used: 0.95
```

As noted earlier, the df are estimated using a different method in *R* than in *SAS* and results can differ slightly.

The confidence interval just barely excludes the value of zero.

There are four sources of variation in this experiment, but all can separated when the analysis is done on the individual values. First the year-to-year variation represents the effects of year-specific factors

such as total rainfall, or other random events. Second, the stream-to-stream variation measures the effect of the different streams on the mean response. The plot produced earlier shows that there is some stream effect as the lines for the streams are consistently separated. Third, is the stream-year interaction where the individual streams do not have exactly parallel trend lines. Lastly, is the quadrat-to-quadrat variation measures the variation over quadrats in different parts of a stream. These can now be separated:

The estimated variance components are:

In *R* this is extracted using the *VarCorr()* extractor function:

```
# Extract the variance components
vc <- VarCorr(result400)
vc
```

which then gives:

```
 Groups        Name        Std.Dev.
 StreamF:YearF (Intercept) 0.0000
 YearF         (Intercept) 6.0537
 StreamF       (Intercept) 6.1896
 Residual                  8.1459
```

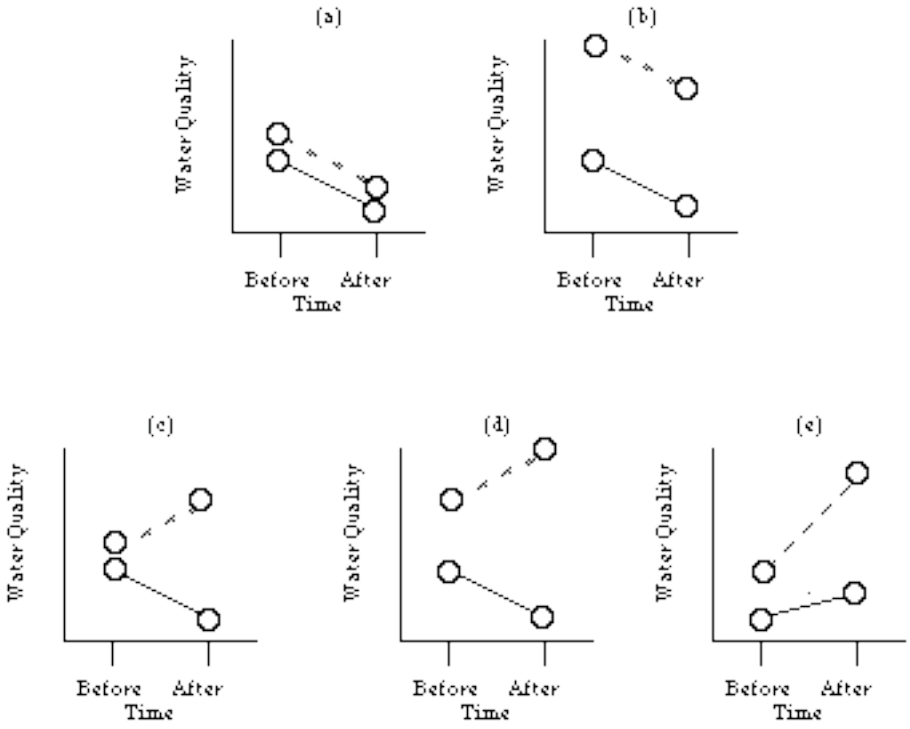
We see that the stream-year interaction variance component is estimated to be zero. *R* does not have an option to allow unbounded variance component estimates as in *SAS*.

We see that the stream-year interaction is very small – this not surprising given the parallel responses of the two steams over the years in the study seen in the preliminary plots.

The usual diagnostic plots (not shown) fail to show any problem with the model fit.

12.3 Simple BACI - One year before/after; one site impact; one site control

The very simplest BACI design has the treatment and a single control site measured before and after the impact occurs. As noted earlier, the evidence of an environmental impact is the non-parallelism of the response between the control and the treatment site. Several examples of possible response are shown below:



The results in the first row above both show no environmental impact; the results in the bottom row all show evidence of an environmental impact.

The number of replicates at each site in each year does not have to be equal, however, power to detect non-parallelism is maximized when the sample sizes are equal in each site-time combination.

The usually assumptions for ANOVA are made:

- the response variable is interval or ratio scale (i.e. continuous in *JMP* parlance). This assumption can be modified, e.g. responses are categorical, but this is not covered in this course. Please contact me for details.
- the standard deviations at each site-year combination should be equal. This is checked by looking at the sample standard deviations. If these are reasonably equal (rough rule of thumb is that the ratio of the largest to smallest standard deviation should be less than 5:1), then the analysis can proceed. If the standard deviation are grossly unequal, then consider a transformation (e.g. taking the logarithm of each observation).
- the observations are independent within each site-year combination and across years. For example, if the measurement at each site are from quadrats, then the quadrats are sufficiently far apart and the same location isn't measured after impact. [If the same spot is measured before and after impact, this is an example of blocking and will be discussed later.]
- the distribution of responses within each site has an approximate normal distribution. This assumption is not crucial – it is far more important that the standard deviations be approximately equal and the observations are independent. In any case, most impact studies have too small of a sample size to detect anything but gross violations of Normality.

The analysis is straight forward. This is a classical two-way anova, a.k.a. a two-factor completely

randomized design. The model to be fit (in the standard simplified syntax) is:

$$Y = Treatment \;\; Time \;\; Treatment*Time$$

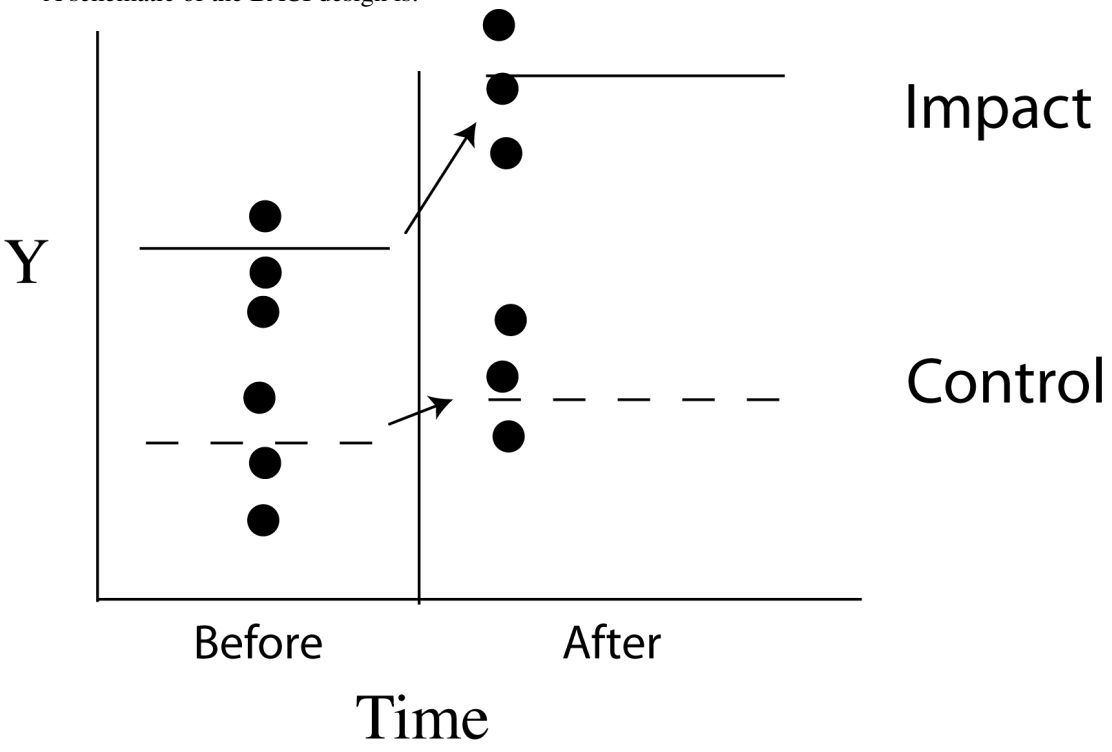
where *Treatment* takes the values of *control* or *impact*, and *Time* takes the value of *before* or *after*. The random variation among the multiple measurements is the only source of random variation and is implicit at the lowest level of the design.

Most standard statistical packages can analyze this design without problems.

12.4 Example: Change in density in crabs near a power plant - one year before/after; one site impact; one site control

A simple BACI design was used to assess the impact of cooling water discharge on the density of shore crabs. The beach near the outlet of the cooling water was sampled using several quadrats before and after the plant started operation, as was a control beach on the other side of the body of water.

A schematic of the BACI design is:



The raw data is available in the *baci-crabs.csv* file available at the Sample Program Library at `http://www.stat.sfu.ca/~cschwarz/Stat-650/Notes/MyPrograms`. The data are imported into *R* in the usual way:

```
cat(" BACI design measuring crab Density \n\n")

crabs <- read.csv("baci-crabs.csv", header=TRUE, as.is=TRUE, strip.white=TRUE)
crabs$trt <- interaction(crabs$SiteClass, crabs$Period)
```

```
crabs$SiteClass <- factor(crabs$SiteClass)
crabs$Period    <- factor(crabs$Period, levels=c("Before","After"), ordered=TRUE) # s
crabs$trt       <- factor(crabs$trt)
cat("Listing of part of the raw data \n")
crabs[1:10,]
```

Part of the raw data are shown below:

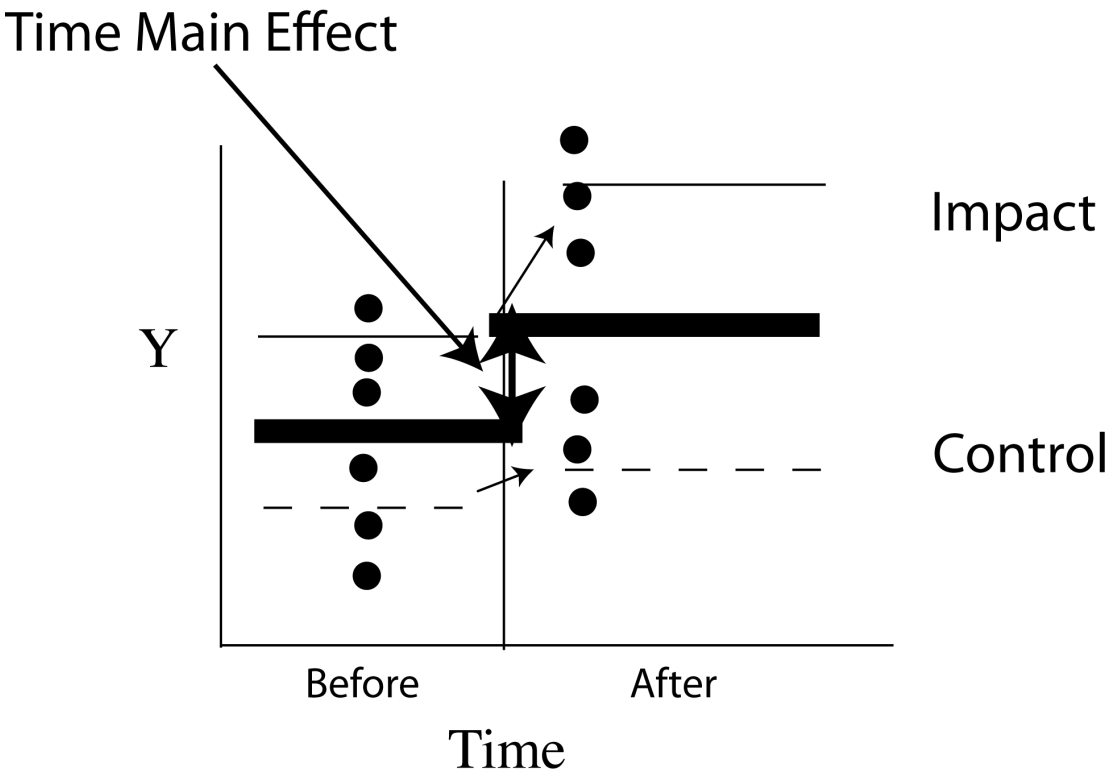
```
 BACI design measuring crab Density

Listing of part of the raw data
   SiteClass Site Period Quadrat Density            trt
1     Impact   I1 Before      Q1      23  Impact.Before
2     Impact   I1 Before      Q2      30  Impact.Before
3     Impact   I1 Before      Q3      27  Impact.Before
4     Impact   I1 Before      Q4      25  Impact.Before
5     Impact   I1  After      Q5      17   Impact.After
6     Impact   I1  After      Q6      19   Impact.After
7     Impact   I1  After      Q7      23   Impact.After
8     Impact   I1  After      Q8      25   Impact.After
9     Impact   I1  After      Q9      17   Impact.After
10   Control   C1 Before     Q10      32 Control.Before
```

Notice that only one site was measured at both the impact and control areas, and multiple, independent quadrats were measured at each site. We have labelled the quadrats using separate identifiers to remind ourselves that new quadrats were measured in the both the before and after phases of the experiment. A common notation, which is not recommended, is to "reuse" the labels, e.g. you would have a Q1 in each site-time combination. This notation is NOT recommended as it not clear that separate quadrats were measured at each opportunity.

The variability in observations from a simple BACI (and more complex BACI as well) can be partitioned into several components, as will be reported in the ANOVA table.

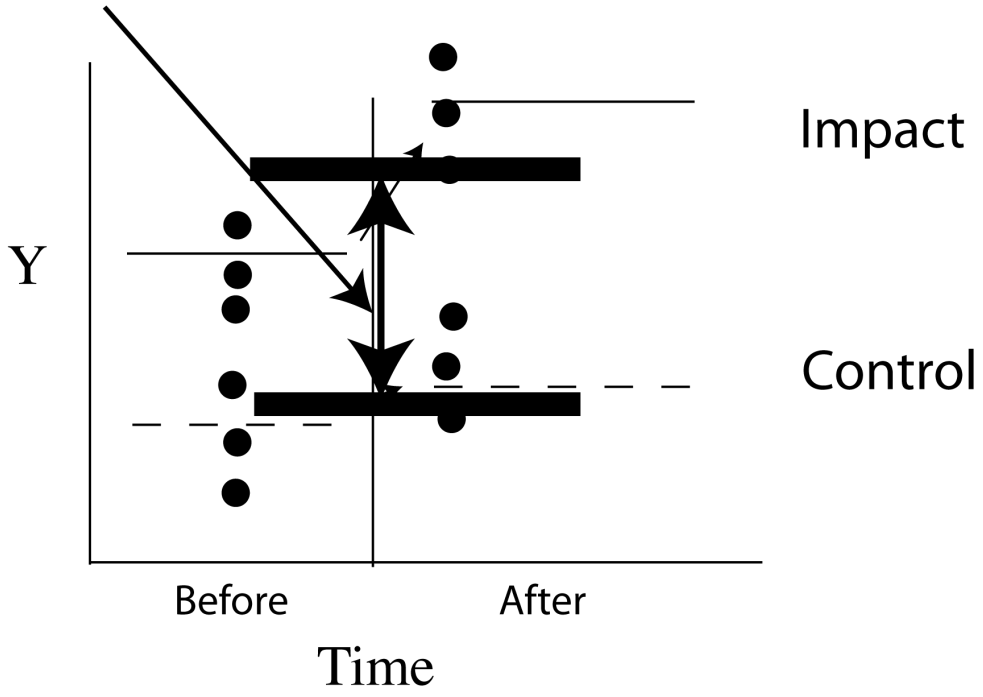
First is the main effect of time. This is simply the different in the mean (over all sites and quadrats) before and after the impact took place:



This temporal effect is not of interest as it may represent the effects of changes in the global conditions unrelated to the impact.

Next, not all sites have exactly the same mean even before the impact takes place. The Site Main Effect represents the differents in the mean among the sites when averaged over all times and quadrats:

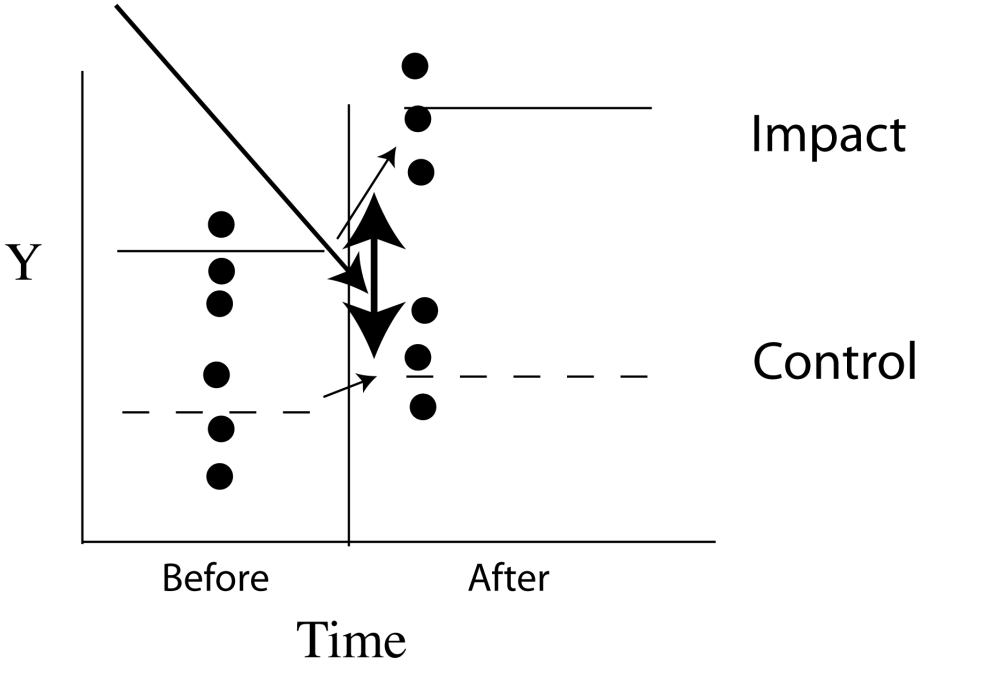
Site Main Effect



The Site effect is not of interest as it represents the effects of site-specific factors that we can neither measure nor control.

Finally, and of greatest intest, is the BACI effect, which is the DIFFERENTIAL CHANGE, i.e. the difference of the two changes in the means, that occurs between the before and after periods.:

BACI (Interaction) Effect = DIFFERENTIAL CHANGE



The BACI, or interaction effect, represent the potential environmental impact and so at test for no inter-

action is equivalent to a test for no environmental impact.

12.4.1 Analysis

A table of means and standard deviations can be computed in the usual ways. In *R*, we use the *summaryBy()* function in the *doBy* package.

```
report <- ddply(crabs, c("Period","SiteClass"), function(x){
   res <- sf.simple.summary(x, "Density", crd=TRUE)
   return(res)
})
cat("\n\n Summary report \n")
report
```

giving:

```


 Summary report
  Period SiteClass n nmiss  mean       sd       se      lcl      ucl
1 Before   Control 6     0 30.50 3.391165 1.384437 26.94119 34.05881
2 Before    Impact 4     0 26.25 2.986079 1.493039 21.49848 31.00152
3  After   Control 4     0 27.50 3.109126 1.554563 22.55269 32.44731
4  After    Impact 5     0 20.20 3.633180 1.624808 15.68881 24.71119
```

The sample standard deviations are approximately equal indicating that the assumption of equal population standard deviations is tenable. The quadrats were sampled far enough apart that they can be considered to be independent. Separate quadrats were sampled in the two time periods.

This is a simple two-factor CRD experiment as seen in previous chapters because no quadrat is repeatedly measured over time, and all observations are independent of each other. The model is

$$Density = SiteClass \;\; Period \;\; SiteClass*Period$$

where *SiteClass* represents the effect of the Control vs. Impact sites; *Period* represents the before vs. after contrast; and the *SiteClass*Period* interaction represents the BACI effect where as noted earlier the non-parallelism (the interaction) represent evidence of an environmental impact.

The *lm()* function can be used to fit this model. The *anova()* function is applied to the result for the effect tests. The *lm()* function can be used for balanced and unbalanced data, but GREAT care is needed in getting the correct ANOVA tables with unbalanced data – the *anova()* function does NOT give the correct sums-of-squares (it gives what are known as incremental or Type I sums of squares). Refer to `http://r-eco-evo.blogspot.com/2007/10/infamous-type-iii-ss-before-i-started_20.html` for more details.

The Type III tests from the *Anova()* function from the *car* package requires that you need to set the treatment contrasts to *sum-to-zero* rather than the default *treatment* BEFORE fitting the *lm()* model! [It this is all greek to you, please contact me for more information.] This is done using the *options()* functions as noted in the code. Refer to `http://r.789695.n4.nabble.com/Type-I-v-s-Type-III-Sum-Of-Square` `html` for further information.

So use the *lm()* with the default contrasts for everything EXCEPT the ANOVA table! Note that if you specify the interaction term as the LAST term in the model, then all the various types of tests (I, II, III) will give identical results for the BACI contrast which is really the only term of interest.

The model is fit:

```
cat("The sum of squares and F-tests from the anova() below are INCORRECT in unbalance
cat("because they are sequential and only adjust for effect\n")
cat("that enter the model prior to the term in question.")
result.lm <- lm( Density ~ SiteClass + Period + SiteClass:Period, data=crabs)
cat("\n\n Analysis of variance -- this is NOT CORRECT because design is unbalanced \n
anova(result.lm)

cat("\n\nUse the Type III tests from the Anova() function from the car package")
cat(  "\nbut you need to set the treatment contrasts to sum rather than treatment")
cat("  \nSee http://r.789695.n4.nabble.com/Type-I-v-s-Type-III-Sum-Of-Squares-in-ANOV
library(car)
result.lm2 <- lm( Density ~ SiteClass + Period + SiteClass:Period, data=crabs,
                  contrasts=list(SiteClass="contr.sum", Period='contr.sum'))
Anova(result.lm2,type=3)
```

The order of the terms in the model is not important.

The output from most packages is voluminous and only part of it will be explored here.

First, the *Effects* tests:

```
The sum of squares and F-tests from the anova() below are INCORRECT in unbalanced dat
because they are sequential and only adjust for effect
that enter the model prior to the term in question.

 Analysis of variance -- this is NOT CORRECT because design is unbalanced
Analysis of Variance Table

Response: Density
                Df  Sum Sq Mean Sq F value    Pr(>F)
SiteClass        1 194.695 194.695 17.5877 0.0007827 ***
Period           1  92.205  92.205  8.3293 0.0113100 *
SiteClass:Period  1  10.734  10.734  0.9696 0.3403935
Residuals       15 166.050  11.070
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Use the Type III tests from the Anova() function from the car package
but you need to set the treatment contrasts to sum rather than treatment
See http://r.789695.n4.nabble.com/Type-I-v-s-Type-III-Sum-Of-Squares-in-ANOVA-td15736

Response: Density
                Sum Sq Df   F value    Pr(>F)
(Intercept)     12588.2  1 1137.1485 1.474e-15 ***
SiteClass         153.9  1   13.9048  0.002016 **
```

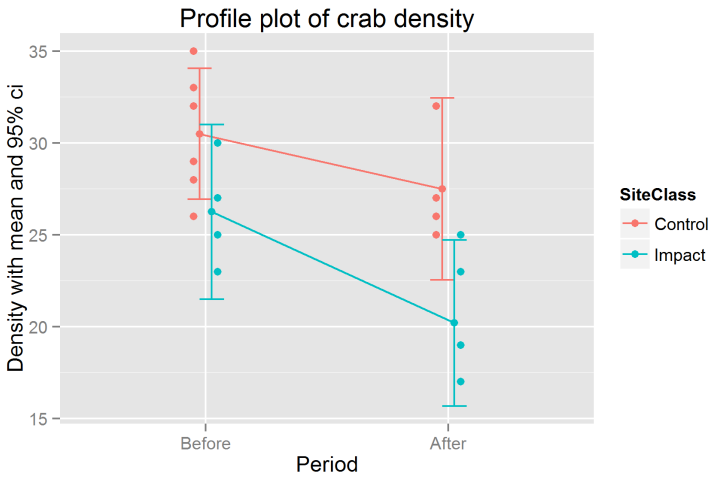
```
Period               94.5  1    8.5368  0.010519 *
SiteClass:Period     10.7  1    0.9696  0.340393
Residuals           166.1 15
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

indicate significant effects of sites (the mean density varied between the two sites regardless of the potential impact), of time (the mean density changed over time due to temporal effects unrelated to the impact), but there was no evidence of an environmental impact as the test for an interaction effect was not statistically significant ($p = 0.34$).

Because the simple BACI is a two-factor CRD, it is relatively straight forward to find the profile plot and add confidence limits for each of the 4 means:

```
profileplot <- ggplot(data=report, aes(x=Period, y=mean, group=SiteClass, color=SiteC
  ggtitle("Profile plot of crab density")+
  ylab("Density with mean and 95% ci")+
  geom_point(position=position_dodge(w=0.1))+
  geom_line(position=position_dodge(w=0.1))+
  geom_errorbar(aes(ymax=ucl, ymin=lcl), width=0.2, position=position_dodge(w=0.1))+
  geom_point(data=crabs, aes(y=Density), position=position_dodge(w=0.2))
profileplot
```

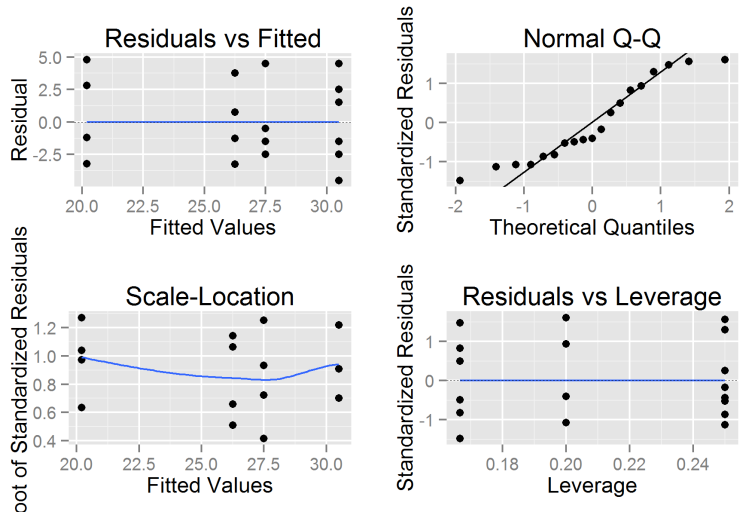
giving:



The profile plot shows two approximately parallel lines.

The residual and other diagnostic plots:

```
diagplot <- sf.autoplot.lm(result.lm)
diagplot
```



don't show any problems.

What is of primary interest, is the difference of the difference, i.e. how much did the change differ between the control and impact sites. This called the BACI contrast and is the estimated environmental impact. Look first at the estimates of the *Least Square Means*. In simple designs, the least-square means are equal to the simple raw means, but in more complex designs they will differ. he least-square means will adjust for missing data and unbalance in the experimental design.

The estimates of the marginal means (and standard errors) are requested using the *popMeans()* function from the *doBy* package.

```
result.lsmo.SP <- lsmeans::lsmeans(result.lm, ~SiteClass:Period)
cat("\n\n Estimated marginal means \n\n")
summary(result.lsmo.SP)
```

The LSMeans estimates are equal to the raw sample means – this is will be true ONLY in balanced data. In the case of unbalanced data (see later), the LSMEANS seem like a sensible way to estimate marginal means.

```


 Estimated marginal means

 SiteClass Period lsmean       SE df lower.CL upper.CL
 Control   Before  30.50 1.358308 15 27.60484 33.39516
 Impact    Before  26.25 1.663580 15 22.70416 29.79584
 Control   After   27.50 1.663580 15 23.95416 31.04584
 Impact    After   20.20 1.487952 15 17.02851 23.37149

Confidence level used: 0.95
```

The estimated mean density in the control site dropped by 3; the mean density in the impact site dropped by 6.05 crabs. The mean difference-in-the-difference is $3.05 = 6.05 - 3$ crabs. But we would like to get an standard error for this estimate. This is obtained by specifying a contrast among the means.

The BACI contrast is:

$$\mu_{CA} - \mu_{CB} - (\mu_{TA} - \mu_{TB}) = \mu_{CA} - \mu_{CB} - \mu_{TA} + \mu_{TB}$$

The estimate of the BACI effect is:

The estimates of the BACI contrast is available form the *summary()* function applied to the linear model results and looking for the term in the table corresponding to the interaction effect:

```
contrast(result.lsmo.SP, list(baci=c(1,-1,-1,1)))
confint(contrast(result.lsmo.SP, list(baci=c(1,-1,-1,1))))
```

```
 contrast estimate       SE df t.ratio p.value
 baci        -3.05 3.097418 15  -0.985  0.3404

 contrast estimate       SE df  lower.CL upper.CL
 baci        -3.05 3.097418 15 -9.651991 3.551991

Confidence level used: 0.95
```

There could a sign difference in the estimate depending on the order in which *R* sorts the levels.

The *p*-value ($p = .34$) is the same as the test for interaction (as it must). It is pretty clear that the 95% confidence interval for the difference-in-the-difference will include 0, indicating no evidence of an effect.

The key advantage of the estimate and 95% confidence interval over the *p*-value is that it enables the analyst to see if the test for interaction was not statistically significant because the effect was small (good news) or because the experiment has so much noise that the standard error for the effect is enormous (bad news).

There is no statistical rule of thumb to distinguish between these two cases. The analyst must rely on biological knowledge. In this case, is a difference-in-the-difference of 3 crabs of biological importance when the base densities are around 20-30 crabs per quadrat?

12.5 Simple BACI design - limitations

The simple BACI design is less than satisfactory for most impact studies for several reasons.

First, the experimental and observation unit are different sizes. The impact is applied at the site level, but the measurements are taken within each unit at the quadrat level, i.e., are pseudo-replicates as outlined by Hurlbert (1984). This means that the scope of inference is limited – if you do detect an effect, you really only can conclude that this impact site behaved different than this particular control site. This leaves the study open to the criticism that the control site was accidentally badly chosen and so the effect detected was an artefact of the study.

Second, the measurements are taken at a single point in time before and after the impact. Again, one could argue that the times chosen were (accidentally) poorly chosen and the results were specific to these time points. Multiple time points before and after the impact could be measured.

Third, there is only one control and one impact site. While impacts are normally applied to a single site, there is no reason why a single control site must be measured. Again, if only a single control site was chosen, one could argue that this single control site was poorly chosen and that the results are again specific to this single control site. Multiple control sites could be measured to ensure that the results are generalizable to more than this specific control site.

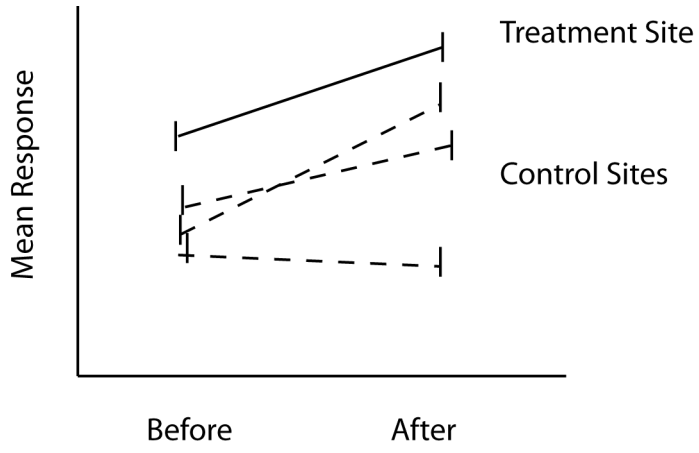
Fourth, if separate quadrats are measured at each site-time combination, additional variability is introduced making it harder to detect effects. Following the principles of experimental designs, blocking, where the same unit is measured over time improves the design.

12.6 BACI with Multiple sites; One year before/after

As noted in the previous section, a pointed critique of the simple BACI design is the use of a single control site. The main criticism is that with a single control site, the replicated measurements within each site are pseudo-replicates. The problem is that the treatment (control or impact) operates at the site level, but measurements are taken at the quadrat within the site level. This implies that inference is limited to these specific impact and control sites. While limiting inference to the single impact site seem reasonable (e.g. only one dam is built), limiting inference to the single control site seems too narrow.

Consequently, it a almost always advisable to have multiple control sites in the impact-assesment design. The theory presented below is general enough to allow replicated impact sites as well.

A conceptual diagram of the BACI design with multiple control sites is:



We will assume that all the sites will be measured both before and after the impact, but only one year of monitoring is done before or after impact. In later sections, we will extend this approach to allow for multiple years of monitoring as well.

All sites are usually measured in both years, i.e. both before and after impact occurs. The analysis below can also deal with sites that are measured only before the impact or after the impact (i.e. is missing some data). However, for these cases, it is important that good software be used (e.g. *SAS*, *JMP*, *R*) because sophisticated algorithms are required. These incomplete sites do have information that should be integrated into the analysis. In particular, sites that are measured only before or only after impact provide information about the variance components which are needed when the formal tests of hypotheses are done.

The use of multiple controls avoids the criticism that the results of the BACI experiment are solely due to a poor choice of control sites. For example, in the above diagram, if only a single control site was monitored, one could conclude that the impact was positive, neutral, or negative depending on which control site was selected.

There are several, mathematically equivalent, ways to analyze these types of designs. The choice among the analysis is often dictated by the availability of software. However, if you want to use this assessment to investigate alternatives, such as increasing the subsampling, increasing the number of years of monitoring, or increasing the number of site, then the "full" analysis that provides estimates of the individual variance components is needed.

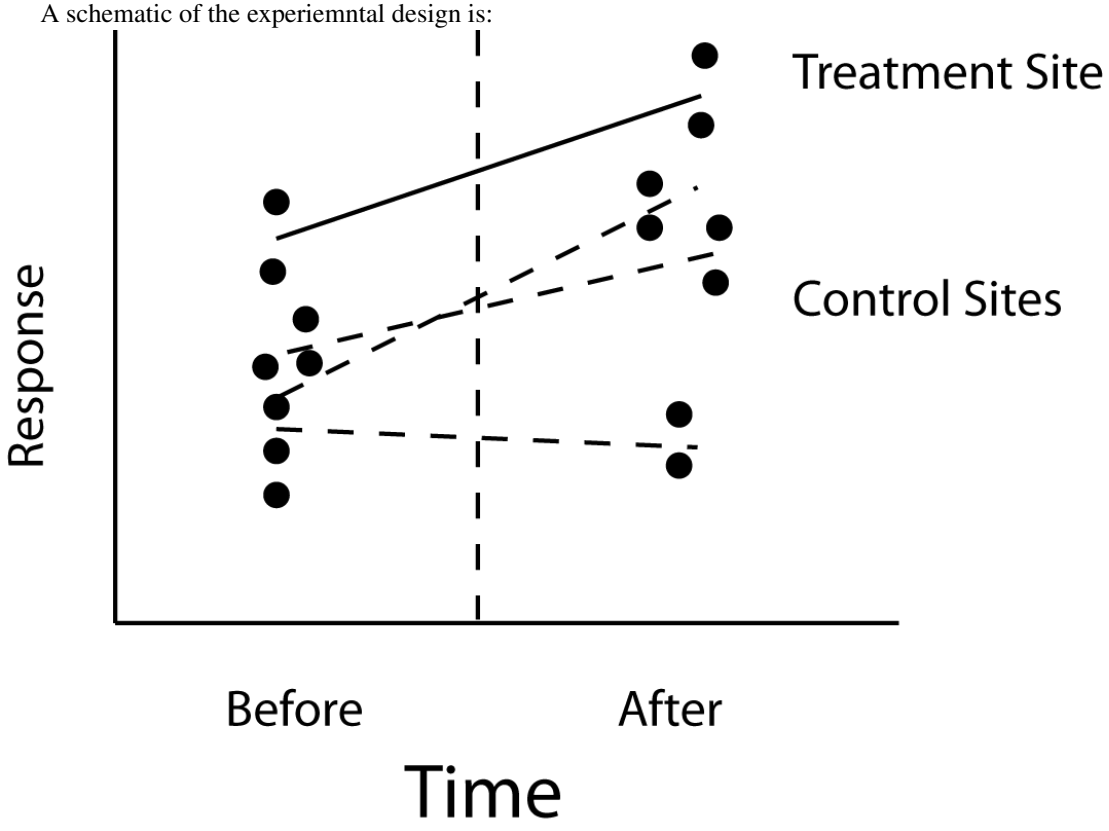
We will proceed by example.

12.7 Example: Density of crabs - BACI with Multiple sites; One year before/after

Let us extend the density of crabs example seen in an earlier section.

The beach near the outlet of the cooling water was sampled using several quadrats before and after the plant started operation. Two control beaches at other locations in the inlet were also sampled.

A schematic of the experiemntal design is:



The raw data is available in the *baci-crabs-mod.csv* file available at the Sample Program Library at `http://www.stat.sfu.ca/~cschwarz/Stat-650/Notes/MyPrograms`. The data are

imported into *R* in the usual way:

```
cat(" BACI design measuring crab Density with multiple sites but one year before/afte

crabs <- read.csv("baci-crabs-mod.csv", header=TRUE, as.is=TRUE, strip.white=TRUE)
crabs$trt <- interaction(crabs$SiteClass, crabs$Site, crabs$Period)
crabs$Site      <- factor(crabs$Site)
crabs$SiteClass <- factor(crabs$SiteClass)
crabs$Period    <- factor(crabs$Period, levels=c("Before","After"), ordered=TRUE)
crabs$trt       <- factor(crabs$trt)
cat("Listing of part of the raw data \n")
crabs[1:10,]
```

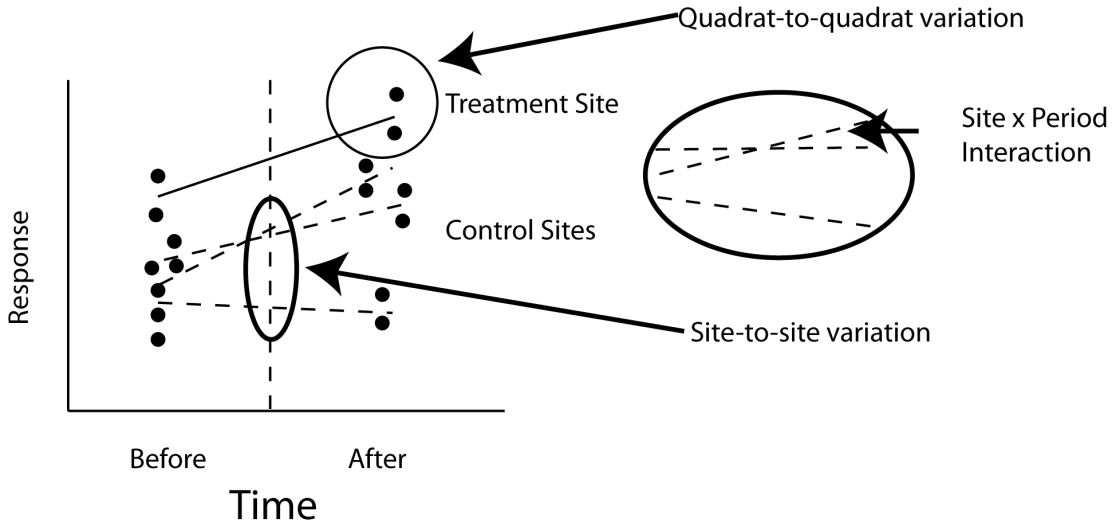
Part of the raw data are shown below:

```
 BACI design measuring crab Density with multiple sites but one year before/after

Listing of part of the raw data
   SiteClass Site Period Quadrat Density               trt
1     Impact   I1 Before      Q1      23  Impact.I1.Before
2     Impact   I1 Before      Q2      30  Impact.I1.Before
3     Impact   I1 Before      Q3      27  Impact.I1.Before
4     Impact   I1 Before      Q4      25  Impact.I1.Before
5     Impact   I1  After      Q5      17   Impact.I1.After
6     Impact   I1  After      Q6      19   Impact.I1.After
7     Impact   I1  After      Q7      23   Impact.I1.After
8     Impact   I1  After      Q8      25   Impact.I1.After
9     Impact   I1  After      Q9      17   Impact.I1.After
10   Control   C1 Before     Q10      32 Control.C1.Before
```

There was one impact site monitored (Site I1 before and after the impact) and two control beachs (C1 and C2 both measured before and after impact). In each site-year combination, between four and six quadrats were measured. It is not necessary that the same number of quadrats be measured in each site-year combination and most good software and deal with this automatically. Notice that each site has a unique label as does each quadrat. This is a good data management practise as it makes it clearer that different sites are used and that all of the quadrats were located and measured independenly.

We can again partition the variation in the data to that due to Period (before vs. after), SiteClass (treatment vs. control), and the BACI contrast (the differential change). In addition, with multiple sites, there are now additional sources of random variation that must be accounted for:



The quadrat-to-quadrat variation (within a beach) represent micro-habitat effects that affect the very localized density of the crabs. The site-to-site variation within each site class (i.e. the multiple sites within the control group) represents the effect of site-specific factors that cause the means in the individual sites (all within the control class) to vary. Finally, the Site-Period or Site-Year interaction represents the INconsistency in the response of the sites to the before vs. after period. For example, not all of the control sites have exactly the same increase in the mean between the before and after periods.

Note that if only one quadrat is measured per site in each year, then the quadrat-to-quadrat variation and the site-year variation are confounded together and cannot be separated.

The analyses presented below start with the analysis of a highly "distilled" dataset and progressively move to the analysis of the raw data.

12.7.1 Converting to an analysis of differences

The measurements taken in each site-year combination are pseudo-replicates. Different quadrats were measured in each year, so there is no "linking" among the quadrats. This idea is reinforced in the dataset by using separate quadrat labels for each quadrat.

We first find the average count over the four to six quadrats in each site-year combination. If the number of quadrats varied among sites, this will give an "approximate" analysis because the design is unbalanced, but is likely "good-enough".

Use the *summaryBy()* function from the *doBy* packaged to compute the mean density for each combination of SiteClass, Site, and Period:

```
# The averages are available in the report dataframe
crabs.avg <- report[,c("Site","SiteClass","Period","mean")]
crabs.avg
```

This gives the following means:

```
  Site SiteClass Period  mean
```

```
1   C1   Control Before 30.50
2   C2   Control Before 38.00
3   I1    Impact Before 26.25
4   C1   Control  After 27.50
5   C2   Control  After 31.20
6   I1    Impact  After 20.20
```

Next, you need to compute the difference in the sample means between the *before* and *after* period within each site. Use the *reshape()* function to bring the two means for the same site on the same record and then compute the difference between the two periods.

```
# Reshape the file to get the Before and After measurements on the same line
#
crabs.site <- dcast(crabs.avg, Site+SiteClass ~ Period, value.var="mean" )
crabs.site$diff <- crabs.site$After - crabs.site$Before
crabs.site
```

This gives the following final file

```
  Site SiteClass Before After  diff
1   C1   Control  30.50  27.5 -3.00
2   C2   Control  38.00  31.2 -6.80
3   I1    Impact  26.25  20.2 -6.05
```

Finally, we now test if the mean difference is the same for the *control* and *impact* groups. This is a simple two-sample *t*-test. Use the *t.test()* function to test if the mean difference is the same in the before and after Periods. Note that because of the very small effective sample size, only the equal variance *t*-test can be performed.

```
result <- t.test(diff ~ SiteClass, data=crabs.site, var.equal=TRUE)
result$diff.in.means <- sum(result$estimate*c(1,-1))
names(result$diff.in.means)<- "diff.in.means"
result$se.diff <- result$statistic / abs(result$diff.in.means)
names(result$se.diff) <- 'SE.diff'
result
result$diff.in.means
result$se.diff
```

giving:

```

        Two Sample t-test

data:  diff by SiteClass
t = 0.34945, df = 1, p-value = 0.786
alternative hypothesis: true difference in means is not equal to 0
```

```
95 percent confidence interval:
 -40.66481  42.96481
sample estimates:
mean in group Control  mean in group Impact
               -4.90                -6.05

diff.in.means
         1.15
  SE.diff
0.3038686
```

The p-value for the test of the hypothesis of no difference in the mean difference between *Control* and *Impact* sites is 0.79 indicating no evidence of a difference in the means, i.e. no evidence of an impact. The output also includes the estimated difference in the mean differences along with a 95% confidence interval. Not surprising, the 95% confidence interval contains the value of 0.

At first glance, it looks as if we have thrown away much information – the multiple quadrats were averaged for each site-year combination, and only the difference between the averages before and after impact were used. In fact, all of the relevant information has been captured. The multiple quadrats within each site-year combination contain information only on the quadrat-to-quadrat variation within each site-year combination. The mean of these values will automatically capture the fact that 5 or 50 quadrats were measured because if a large number of quadrats were measured, the variation of the sample mean will be reduced which implies that there will be a reduction in the variation of the differences, which implies that the power of the test to detect differences in the mean difference will be improved.

Similarly, a BACI design finds evidence of an impact if the change between the before and after period is the different for the control and treatment groups, i.e. if the mean difference between the before and after periods is the same for the treatment and control groups which is exactly what is tested by the two-sample t-test above.

This was a very small experiment so it is not surprising that no effect was detected. Later on, the power of this design will be examined.

12.7.2 Using ANOVA on the averages

Rather than going though the contortions of splitting a column and finding differences, it is possible to do an ANOVA on the averages directly.

This design is an example of a **Split-Plot** design which is covered in more detail elsewhere in my course notes on my web pages.

In split-plot designs, there are two levels to the experiment. The first level is at the *site* level. Here sites are chosen (at random) from each of the treatment and control areas. Each site is then "split" by taking measurements at two time periods (before and after).

Split-plot designs are examples of two-factor designs (see my notes on my web page). They differ from two-factor completely randomized designs where there is only a single level in the experiment. The model used to analyze data of this form must include terms for both the two factors (and their interaction which is the BACI effect of interest) but also for the two levels in the experiment. The experimental unit at the lowest level (in this case the time period within each site) is usually not specified in the model and

is implicit[1].

The model to be fit is:

There are two functions in *R*, for fitting linear models with additional random effects. The older function *lme()* and the newer function *lmer()*. The two functions have slightly different syntax but are very similar. The *lme()* function is awkward to use when the random effects are crosses (see later in this section), but the *lmer()* function don't report the standard *p*-values as we might expect to see. We will use the *lme()* function in this section of the notes.

Notice in the code below how the function distinguishes from the fixed effects and the random effects. Be sure that the random effects are **factors** rather than continuous numbers or you will be a very unusual (and WRONG) model fit.

```
result.lmer <- lmerTest::lmer(mean ~ SiteClass+Period+SiteClass:Period +
                     (1|Site), data=crabs.avg)
anova(result.lmer, ddf="Kenward-Roger")
```

Many packages will use a technique called REML (Restricted Maximum Likelihood estimation) to obtain the test statistics and estimates. REML is a variant of maximum likelihood estimation and extract the maximal information from the data. With balanced data, the F-tests are identical to an older method called the Expected Mean Squares (EMS) method. Modern statistical theory prefers REML over the EMS methods.

The REML method comes with two variants. In one variant (the default with *JMP*), estimated variance components are allowed to go negative. This might seem to be problematic as variances must be non-negative! However, the same problem can occur with the EMS method, and usually only occurs when sample sizes are small and variance components are close to 0. It turns out that the F-tests are "unbiased" when the unbounded variance component option is chosen even if some of the variance estimates are negative. As noted in the *JMP* help files,

> "If you remain uncomfortable about negative estimates of variances, please consider that the random effects model is statistically equivalent to the model where the variance components are really covariances across errors within a whole plot. It is not hard to think of situations in which the covariance estimate can be negative, either by random happenstance, or by a real process in which deviations in some observations in one direction would lead to deviations in the other direction in other observations. When random effects are modeled this way, the covariance structure is called compound symmetry.
>
> So, consider negative variance estimates as useful information. If the negative value is small, it can be considered happenstance in the case of a small true variance. If the negative value is larger (the variance ratio can get as big as 0.5), it is a troubleshooting sign that the rows are not as independent as you had assumed, and some process worth investigating is happening within blocks.

So if a variance component is negative and has a large absolute value, this is an indication that your model may not be appropriate for the data at hand. Please contact me for assistance. The default for *SAS* is to constrain the variance components to be non-negative and so the F-tests may vary slightly from those of *JMP*.

Part of the output (the *Fixed Effect Tests*) is presented below:

[1] This corresponds to the *Error* line in the ANOVA table.

The *anova(result.lme)* function applied to the result from the fit, gives the effect tests. Note that when using the *lme()* function, you no longer have to worry about Type I, II, or III tests.

```
Analysis of Variance Table of type III  with  Kenward-Roger
approximation for degrees of freedom
                 Sum Sq Mean Sq NumDF DenDF F.value Pr(>F)
SiteClass        11.286  11.286     1     1  3.1263 0.3277
Period           39.968  39.968     1     1 11.0713 0.1859
SiteClass:Period  0.441   0.441     1     1  0.1221 0.7860
```

The test for no interaction between the *Time* and *Treatment* effects is the test of no BACI effect, i.e. no impact. The *p*-value of 0.786 is identical to that found in the previous analyses. The other F-tests represent the tests for the main effects of *Treatment* or *Time* and are usually not of interest.

The estimates of the variance components provide some information on the relative sizes of variation found in this experiment:

The *VarCorr()* extractor function applied to the result from the fit gives the variance components:

```
# Extract the variance components
vc <- VarCorr(result.lmer)
vc
```

```
 Groups   Name        Std.Dev.
 Site     (Intercept) 3.7249
 Residual             1.9000
```

The estimated variance of site-to-site effects is 13.9, while the variation of the average within a site is 3.61. Note that the latter must be interpreted carefully as it represents the variance on the AVERAGE of the replicated quadrats and not the quadrat-to-quadrat variation. It also include the site-time interaction. This variance component represent the potential of different changes between before and after depending on site. For example, in some of the sites, the effect of time may be to increase the response, while in other sites, the effect of time is to decrease the response (this is after accounting for random error). If this occurs, then you have serious problems in detecting an environmental impact as it not even clear how to properly define an impact when the direction of the impact is not consistent.

A naked *p*-value should never be reported. Regardless if the test of interaction (the BACI effect) is statistically significant or not, an estimate of the effect should should also be reported.

An estimate of the BACI contrast that is independent of the internal parameterization used can be obtained by looking at the *Effect Details*, in particular at the *Time*Treatment* estimates. The BACI effect is the difference in the differences, i.e.

$$BACI = \mu_{CA} - \mu_{CB} - (\mu_{TA} - \mu_{TB}) = \mu_{CA} - \mu_{CB} - \mu_{TA} + \mu_{TB}$$

The estimate is found by substituting in the estimated means (called the *Least Square Means*) or

$$\widehat{BACI} = 29.35 - 34.25 - 20.2 + 26.25 = 1.15$$

The estimates of the BACI contrast is available form the *summary()* function applied to the linear model results and looking for the term in the table corresponding to the interaction effect:

```
# Estimate the BACI contrast along with a se
contrast(result.lmer.lsmo.SP, list(baci=c(1,-1,-1,1)))
confint(contrast(result.lmer.lsmo.SP, list(baci=c(1,-1,-1,1))))
```

```
 contrast estimate       SE df t.ratio p.value
 baci        -1.15 3.290897  1  -0.349  0.7860

 contrast estimate       SE df lower.CL upper.CL
 baci        -1.15 3.290897  1 -42.9648  40.6648

Confidence level used: 0.95
```

Note that the *p*-values for the test of the interaction (the BACI effect) are all identical (as they must be) regardless of the parameterization used.

It is not straightforward to estimate the marginal means for the four treatment combinations of *SiteClass* and *Period*, and required a bit of *R* magic.

```
result.lmer.lsmo.SP <- lsmeans::lsmeans(result.lmer, ~SiteClass:Period)
cat("\n\n Estimated marginal means for SiteClass:Period \n\n")
summary(result.lmer.lsmo.SP)
```

```


 Estimated marginal means for SiteClass:Period

 SiteClass Period lsmean       SE   df   lower.CL upper.CL
 Control   Before  34.25 2.956772 1.23   9.714860 58.78514
 Impact    Before  26.25 4.181507 1.23  -8.447928 60.94793
 Control   After   29.35 2.956772 1.23   4.814860 53.88514
 Impact    After   20.20 4.181507 1.23 -14.497928 54.89793

Confidence level used: 0.95
```

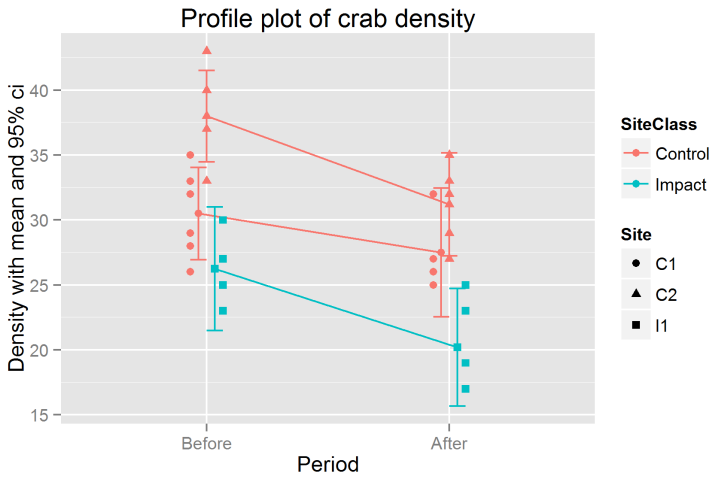
These terms were used to estimate the BACI contrast seen earlier.

We can also get the standard profile plot:

```
profileplot <- ggplot(data=report, aes(x=Period, y=mean,
                    group=Site, color=SiteClass, shape=Site))+
  ggtitle("Profile plot of crab density")+
  ylab("Density with mean and 95% ci")+
  geom_point(position=position_dodge(w=0.1))+
  geom_line(position=position_dodge(w=0.1))+
```

```
  geom_errorbar(aes(ymax=ucl, ymin=lcl), width=0.2, position=position_dodge(w=0.1))+
  geom_point(data=crabs, aes(y=Density), position=position_dodge(w=0.2))
profileplot
```

and then get the profile plot:



This shows that the response is almost parallel – it is no wonder no BACI effect was detected.

12.7.3 Using ANOVA on the raw data

Finally, it is possible to do an ANOVA on the raw data. This will correctly account for the unbalance in the number of quadrats sampled at each site-year combination (and so the *p*-values will be slightly different than the above analyses based on averages). The advantage of this analysis method is that it also produces estimates of all of the variance components that can then be used for a power analysis.

The analysis of the raw data is a variant of a split-plot design with sub-sampling (the quadrats). The statistical model will include terms for the site level (the random effects of site) and also a term for the possible site-time interaction that was subsumed in to the residual error in the previous analysis.

The model is We hope that this variance component will be negligible – otherwise, it indicates that the effect of before/after varies among the sites as well as between the control and impact classifications. This is generally not good news.

We again use the *lme()* function. Notice the *R* magic to get the *Site(R)* and the *Period*Site(R)* terms!

```
result.all.lmer <- lmer(Density ~ SiteClass+Period+SiteClass:Period +
                  (1|Site) + (1|Site:Period), data=crabs)
anova(result.all.lmer, ddf="Kenward-Roger")
```

The *anova(result.lme)* function applied to the result from the fit, gives the effect tests. Note that when using the *lme()* function, you no longer have to worry about Type I, II, or III tests.

```
Analysis of Variance Table of type III  with  Kenward-Roger
approximation for degrees of freedom
                  Sum Sq Mean Sq NumDF  DenDF F.value Pr(>F)
SiteClass         31.897  31.897     1 1.0144  2.9191 0.3345
Period           109.060 109.060     1 1.1140  9.9808 0.1743
SiteClass:Period   1.099   1.099     1 1.1140  0.1006 0.8002
```

However, *R* does not have the Kenward-Rogers approximations to the degrees of freedom for unbalance data as does *SAS* and *JMP*. Consequent, *R* uses a different approximation to the degrees of freedom and the results will be (slightly) different from those of *SAS* or *JMP*.

Because the number of quadrats was not identical in all site-year combinations, the degrees of freedom can be fractional (except in *R* which does have the same methods of computing the df as *SAS* or *JMP*. Also notice that the *p*-value is slightly different from the previous results (but very similar). As long as the imbalance in the number of quadrats is small, the analyses on the averages as presented earlier, should not be affected much. The *p*-value for the test of interaction is 0.80 and there is no evidence of a BACI effect.

Again, regardless of the *p*-value, an estimate of the BACI effect should be obtained. The reported parameter estimates or the indicator parameterization estimates should not be used directly (unless you are very sure of what they mean!), and again construct the BACI contrast based on the *Effect Details*:

It is not straightforward to estimate the marginal means for the four treatment combinations of *SiteClass* and *Period*, and required a bit of *R* magic.

```
result.all.lmer.lsmo.SP <- lsmeans::lsmeans(result.all.lmer, ~SiteClass:Period)
cat("\n\n Estimated marginal means for SiteClass:Period \n\n")
summary(result.all.lmer.lsmo.SP)
```

```


 Estimated marginal means for SiteClass:Period

 SiteClass Period   lsmean       SE   df   lower.CL upper.CL
 Control   Before 34.25000 3.015372 1.19   7.646548 60.85345
 Impact    Before 26.25000 4.369842 1.31  -6.138049 58.63805
 Control   After  29.30381 3.070464 1.27   5.344424 53.26320
 Impact    After  20.20000 4.306875 1.24 -15.107054 55.50705

Confidence level used: 0.95
```

These terms were used to estimate the BACI contrast next.

The BACI contrast is:

$$BACI = \mu_{CA} - \mu_{CB} - (\mu_{TA} - \mu_{TB}) = \mu_{CA} - \mu_{CB} - \mu_{TA} + \mu_{TB}$$

and

$$\widehat{BACI} = 29.30 - 34.25 - 20.2 + 26.25 = 1.10$$

The estimates of the BACI contrast is available form the *summary()* function applied to the linear model results and looking for the term in the table corresponding to the interaction effect:

```
# Estimate the BACI contrast along with a se
contrast(result.lmer.lsmo.SP, list(baci=c(1,-1,-1,1)))
confint(contrast(result.lmer.lsmo.SP, list(baci=c(1,-1,-1,1))))
```

```
 contrast estimate       SE df t.ratio p.value
 baci        -1.15 3.290897  1  -0.349  0.7860

 contrast estimate       SE df lower.CL upper.CL
 baci        -1.15 3.290897  1 -42.9648  40.6648

Confidence level used: 0.95
```

The estimated variance components are:

The *VarCorr()* extractor function applied to the result from the fit gives the variance components:

```
# Extract the variance components
vc <- VarCorr(result.all.lmer)
vc
```

```
 Groups      Name        Std.Dev.
 Site:Period (Intercept) 1.2963
 Site        (Intercept) 3.8319
 Residual                3.3056
```

Notice that the term labelled *Period* is the *Period*Site(R)* variance component.

The estimated site-to-site variation is 14.68; the estimated interaction of sites and time is 1.68. This is small which is good, because it indicates that the response of the sites over time is fairly consistent. If this value were large, it would throw into doubt the whole BACI paradigm of this experiment. Finally, the residual variance component of 10.92 measures the quadrat-to-quadrat variation within each site-year combination.

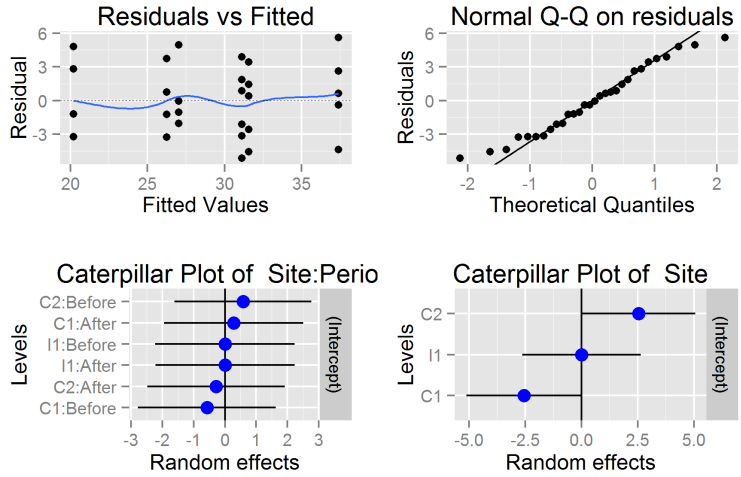
12.7.4 Model assessment

Regardless of the model chosen, you should carefully examine the data for outliers and unusual points. If there are outliers, and the data was incorrectly entered, then, of course, the data should be corrected. If the outlier is "real", then you may wish to run the analysis with and without the outlier present to see what effect it has. If the final conclusions are the same, then there isn't a problem. However, if inclusion

of the outlier greatly distorts the results, then you have a problem, but it is not something that statistical theory can help with. If the data point is real, it should be included!

You should also look a the residual plots (look under the red triangles to plot residuals). Check to see that the residuals appear to be randomly scattered around zero.

Here is the model assessment plots from the final model on the raw data:



If you have a large number of sites, it is possible to estimate the site-residuals and plots of these may also be informative. In many cases, the number of sites is small so these are less than informative.

The assumption of Normality is often made. Unfortunately, with small sample sizes, it is very difficult to assess this assumption. Fortunately, ANOVA is fairly robust to non-Normality if the standard deviations are roughly equal in each site-year combination. A transformation (e.g. the $\ln()$ transformation) may be required.

12.8 BACI with Multiple sites; Multiple years before/after

The previous section looked at designs with only a single year of measurement before and after the impact. This design can be improved by also monitoring for several years after and before the impact. While the term "year" is used generally in these notes, the same methods can be used when sampling takes place more frequently than years. For example, samples could be taken every season which would add additional structure to the data.

Once additional years of monitoring are used, there is now a concern about the potential type of responses that can occur following impact. For example, the impact could cause a permanent step change in the impact site(s) that persists for many years. Or, the impact could cause a temporary change in the response variable that slowly returns to pre-impact levels.

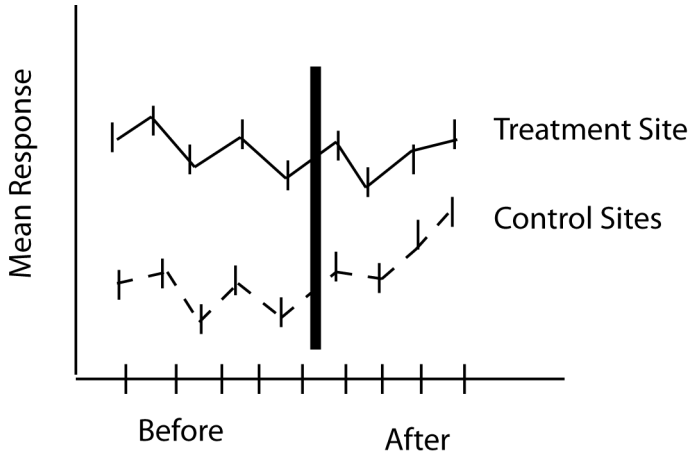
The simplest case is that of a permanent step change after impact that persists for several years. It is implicitly assumed that the relationship between the control and treatment areas is consistent in the pre-impact years.

For these types of studies, the most sensible design is the BACI-paired design that was outlined

earlier. Here, permanent monitoring sites are established in each of the control and treatment areas. All sites are measured each year. [The analyzes presented below can deal with cases where not all sites are measured in all years as long as the missing data is missing completely at random (MCAR). This implies that the missingness of measurement at a site is not related to the response or any other covariate in the study.] The year effects are assumed to operate in parallel across all sites. [This assumptions can be relaxed if you have multiple sites and every site is measured in every year.]

We will also allow for sub-sampling at each year-site combination and will assume that all sub-samples are independent both within and across years. A case where this is not true, would be the use of permanent sub-sampling sites within each permanent site. Please contact me for more details on this design.

Conceptually this is diagrammed as:



Note that the BACI contrast looks at changes in the mean response between the before and after periods.

As before, you should enter the data carefully using unique labels for every distinct experimental or observational unit. For example, if separate quadrats are measured within each site-time combination, then each quadrat should have a different label. So if five quadrats are measured at site I1 (the treatment site) and at the two control sites (C1 and C2), for two years both before and after the impact, there there are $5 \times 3 \times 2 \times 4 = 120$ quadrats and they should be labelled Q1, Q2, ... Q120, rather than Q1 to Q5 within each site-year combination. This latter (but common) labelling is confusing because the data appears to be coming from a design with permanent sub-sampling quadrats. Even more confusing, how can quadrat Q1 be sampled from two different sites?

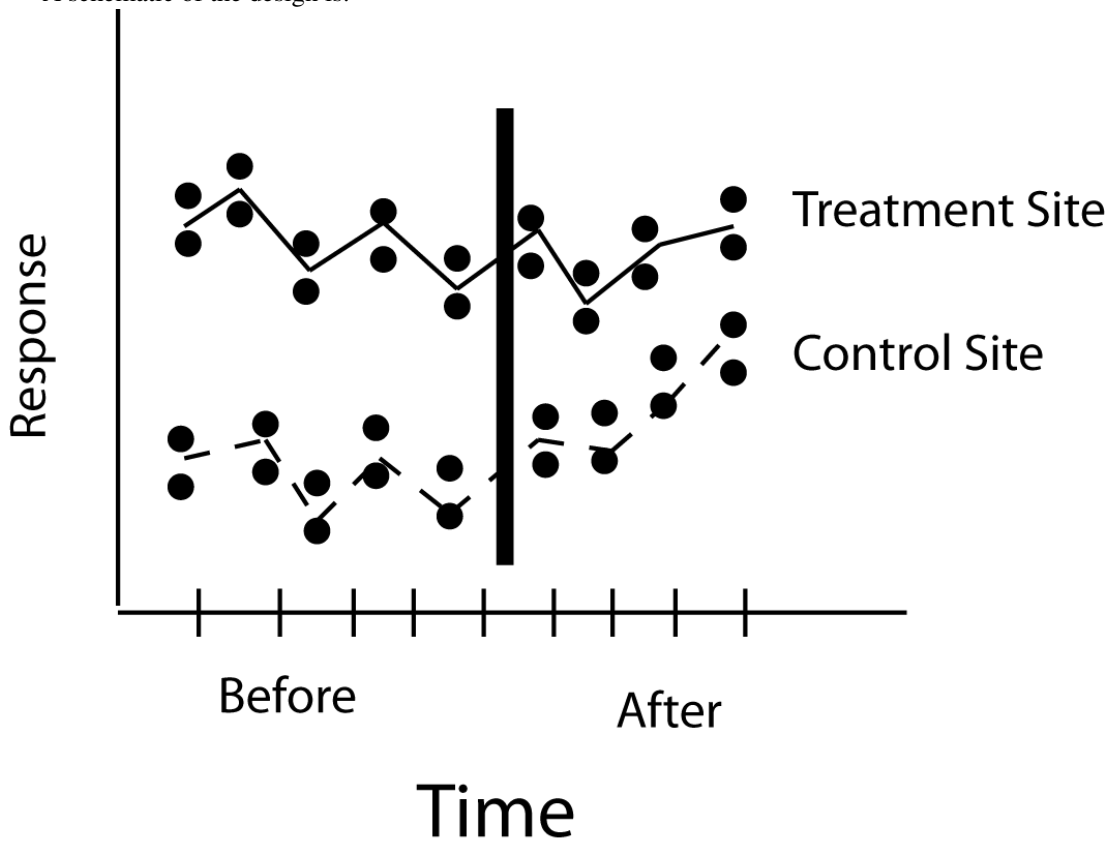
The key to the analysis of these experiment is to recognize the different sizes of experimental and observational units in the impact designs. At the top level, the multiple sites within the control and treatment areas are again the starting of a split-plot type of design. The effect of year is what is known in statistical parlance as a strip as it "operates" across all sites regardless if it is a treatment or control site. The quadrats measured in each site-year combination are again treated as pseudo-replicates.

We will proceed by example.

12.9 Example: Counting fish - Multiple years before/after; One site impact; one site control

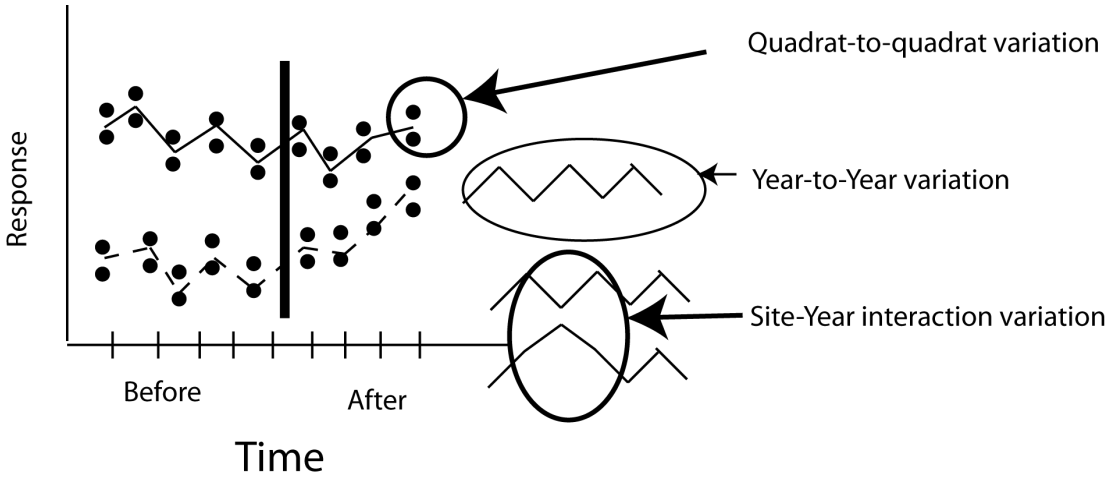
This is the example that appears in Smith (2002, Table 6). Samples of fish were taken for a period of 12 months before and 13 months after a nuclear power plant began operations. The power plant is cooled by water that is drawn from a river. When the water exits the plant, its temperature is elevated. The concern is that the warmed water will adversely affect the abundance and composition of fish below the plant. [It wasn't clear from Smith (2002) what the response measure is, so I'll arbitrarily assume that it is counts.]

A schematic of the design is:



except that for this example, there is only one reading per site per year rather than multiple readings.

As before, we can identify the Period (before vs. after), SiteClass (impact vs. control) and the BACI effect (the interaction between Period and SiteClass). There are also several sources of variation:



The quadrat-to-quadrat variation represents localized effects within each site. The Year-to-Year variation represent year-specific factor that cause the response to move higher or lower at both site classes. For example, a very cool year, could cause productivity of a population to be lower than normal in both the impact and control sites. Again, the site-year interaction variance component represents the non-parallel behavior of sites over time. We expect that year-specific factors to shift the means in the sites up and down in parallel, but not all sites respond in the same way.

Again note that if only a single measurement is taken in each site-year combination then the quadrat-to-quadrat and the site-year interaction variance components are completely confounded together and cannot be separated.

The raw data is available in the *baci-fish.csv* file available at the Sample Program Library at `http://www.stat.sfu.ca/~cschwarz/Stat-650/Notes/MyPrograms`. The data are imported into *R* in the usual way:

```
cat(" BACI design measuring fish counts with multiple (paired) yearly measurements be

fish <- read.csv("baci-fish.csv", header=TRUE, as.is=TRUE, strip.white=TRUE)
fish$trt <- interaction(fish$SiteClass, fish$Site, fish$Period)
fish$Period        <- factor(fish$Period)
fish$SiteClass     <- factor(fish$SiteClass)
fish$Site          <- factor(fish$Site)
fish$SamplingTimeF <- factor(fish$SamplingTime)
fish$trt           <- factor(fish$trt)
cat("Listing of part of the raw data \n")
fish[1:10,]
```

Part of the raw data are shown below:

```
 BACI design measuring fish counts with multiple (paired) yearly measurements before/

Listing of part of the raw data
   SamplingTime Year Period SiteClass Site Count               trt SamplingTimeF
1             1   75 Before   Control   C1     1 Control.C1.Before            1
2             1   75 Before    Impact   I1     1  Impact.I1.Before            1
3             2   75 Before   Control   C1     0 Control.C1.Before            2
4             2   75 Before    Impact   I1     1  Impact.I1.Before            2
```

```
5              3   75 Before   Control   C1     0 Control.C1.Before              3
6              3   75 Before    Impact   I1     0  Impact.I1.Before              3
7              4   75 Before   Control   C1     2 Control.C1.Before              4
8              4   75 Before    Impact   I1     6  Impact.I1.Before              4
9              5   75 Before   Control   C1    47 Control.C1.Before              5
10             5   75 Before    Impact   I1   103  Impact.I1.Before              5
```

In this study there was one (permanent) site measured in each of the control and impact areas and the site is labelled as C1 and I1 respectively. At each site, monthly (paired) readings were taken at sampling times $1, \ldots, 25$ denoted by the sampling time variable. Notice that in the pre-impact year, all months were measured, while in the post-impact years sampling spans two years. The actual months of sampling were not given by Smith (2002), so without further assumptions it is not possible to use the calendar months as another covariate (e.g. counts in January are expected to be different than those in July in some systematic fashion).

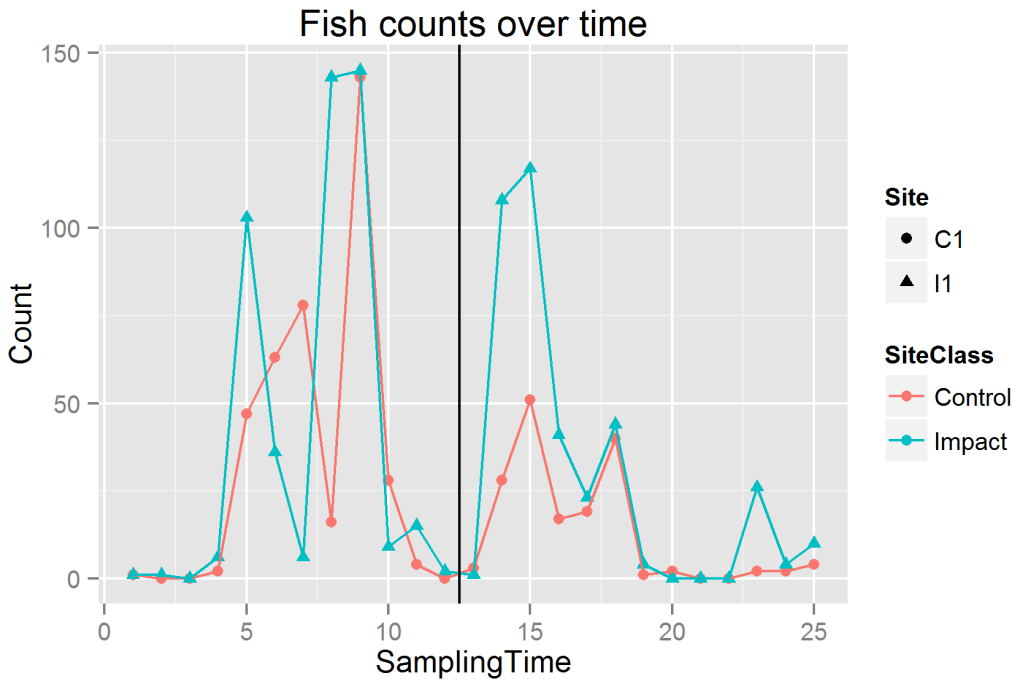
No sub-sampling was conducted so it is not necessary to average over the sub-samples.

As only one site is measured for each of the control and impact areas, this experiment is an example of pseudo-replication and so conclusions will be limited to inference about these two sites only.

We being by a simple plot of the data over time. We plot of the two series over time:

```
prelimplot <- ggplot(data=fish, aes(x=SamplingTime, y=Count,
                          group=Site, color=SiteClass, shape=Site))+
  ggtitle("Fish counts over time")+
  geom_point()+
  geom_line()+
  geom_vline(xintercept=-0.5+min(fish$SamplingTime[as.character(fish$Period) == "Afte
prelimplot
```

giving:



Notice that the two curve track each other over time indicating that the pairing by month will be effective in accounting for the natural variation over the seasons.

As in the previous section, there are several (equivalent) ways to analyze this data.

12.9.1 Analysis of the differences

This is an example of simple paired-design where the sampling times induce the pairing in the observations across the sites. We start by reshaping the data to create two columns - one for the impact and one for the control site.

Use the *reshape()* function to bring the two values taken in the same month on the same record and then compute the difference between the two periods.

```
# Reshape the file to get the Control and Impact measurements on the same line
fish.month <- reshape(fish, v.names=c("Count"), timevar=c("SiteClass"),
   idvar=c("SamplingTime","Period"), drop=c("Site","trt"), direction="wide" )
fish.month$diff <- fish.month$Count.Impact - fish.month$Count.Control
fish.month[1:10,]
```

This gives the following final file

```
   SamplingTime Year Period SamplingTimeF Count.Control Count.Impact diff
1             1   75 Before             1             1            1    0
3             2   75 Before             2             0            1    1
5             3   75 Before             3             0            0    0
```

```
7              4   75 Before              4              2              6   4
9              5   75 Before              5             47            103  56
11             6   75 Before              6             63             36 -27
13             7   75 Before              7             78              6 -72
15             8   75 Before              8             16            143 127
17             9   75 Before              9            143            145   2
19            10   75 Before             10             28              9 -19
```

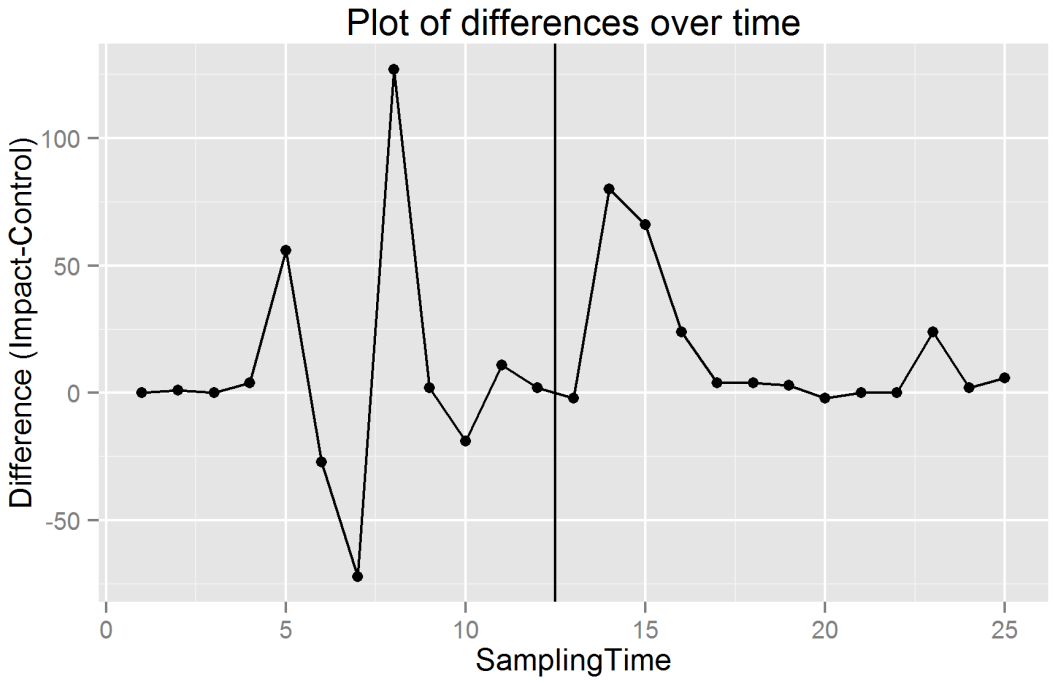
We start with a plot of the differences over the sampling times:

We plot the difference over time:

```
plotdiff <- ggplot(data=fish.month, aes(x=SamplingTime, y=diff))+
  ggtitle("Plot of differences over time")+
  ylab("Difference (Impact-Control)")+
  geom_point()+
  geom_line()+
  geom_vline(xintercept=-0.5+min(fish$SamplingTime[as.character(fish$Period) == "Afte

plotdiff
```

giving:



We notice that there is very large swing in the difference in the before period. This is "unfortunate" as this will give rise to a tremendous residual variation and make it difficult to detect effects.

A two-sample t-test is used to compare the difference between the two *Periods*. Use the *t.test()*

function to test if the mean difference is the same in the before and after Periods. Note that because of the very small effective sample size, only the equal variance t-test can be performed.

```
result <- t.test(diff ~ Period, data=fish.month, var.equal=TRUE)
result$diff.in.means <- sum(result$estimate*c(1,-1))
names(result$diff.in.means)<- "diff.in.means"
result$se.diff <- result$statistic / abs(result$diff.in.means)
names(result$se.diff) <- 'SE.diff'
result
result$diff.in.means
result$se.diff
```

giving:

```

        Welch Two Sample t-test

data:  diff by Period
t = 0.5741, df = 17.019, p-value = 0.5734
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -24.05346  42.04064
sample estimates:
 mean in group After mean in group Before
           16.076923             7.083333

diff.in.means
      8.99359
   SE.diff
0.06383726
```

Notice that we used two-sample t-test allowing for unequal variances in the two time periods. If the pooled variance t-test is used it replicates the results from ANOVA in the next section exactly.

The p-value is 0.57 which provides no evidence of a difference in the differences between the two periods (the BACI contrast). The estimate of the BACI contrast (with *se*) is also provided.

The large differences is worry some – this indicates perhaps a lack of fit of this model. If a non-parametric test is done (click on the red triangle), it also fails to detect a difference:

Use the *wilcox.test()* function for a non-parametric test (Wilcoxon Rank Sum Test) to test if the median difference is the same in the before and after Periods.

```
result <- wilcox.test(diff ~ Period, data=fish.month, conf.int=TRUE)
result
```

giving:

```
        Wilcoxon rank sum test with continuity correction

data:  diff by Period
W = 97, p-value = 0.3124
alternative hypothesis: true location shift is not equal to 0
95 percent confidence interval:
 -3.000054 26.999971
sample estimates:
difference in location
              3.409536
```

12.9.2 ANOVA on the raw data

The model for the raw data is

$$Count = SiteClass \quad Period \quad SiteClass*Period \quad SampleTime(R)$$

where the term *SiteClass* represents the control vs. impact effect; the term *Period* represents the before vs. after effect; and the term *SiteClass*Period* represents the BACI effect, i.e. the non-parallel response. Because there is only one site measured in each of the control or treatment areas, the site-effects are completely confounded with the treatment effects (as happens with pseudo-replicated data). Consequently, it is not necessary to enter the *Site* effect into the model.[2] We do have multiple times measured before and after the impact. If the *Sampling Times* use unique labels, we can once again enter a random effect for *SamplingTimes&Random* into the model. Alternatively you could enter *SamplingTimes (Time)&Random* as the term in the model if the sampling times were not uniquely labelled.

We again use the *lme()* function. Notice the *R* magic to get the *SamplingTime(R)* term! **You must insure that the numeric variable** *SamplingTime* **is treated as a factor and not as a continuous covariate.**

```
# Because there is ONLY one measurement per year, the SamplingTime*Site and
# residual variance are total confounded and cannot be separated. This is
# the residual term.
result.lmer <- lmer(Count ~ SiteClass+Period+SiteClass:Period + (1|SamplingTimeF),
                    data=fish)
anova(result.lmer, ddf="Kenward-Roger")
```

The *anova(result.lme)* function applied to the result from the fit, gives the effect tests. Note that when using the *lme()* function, you no longer have to worry about Type I, II, or III tests. However, *R* does not have the Kenward-Rogers approximations to the degrees of freedom for unbalance data as does *SAS* and *JMP*. Consequent, *R* uses a different approximation to the degrees of freedom and the results may be (slightly) different from those of *SAS* or *JMP*.

The tests for the fixed effects are:

```
Analysis of Variance Table of type 3  with  Kenward-Roger
approximation for degrees of freedom
```

[2] If Site is entered as a random effect, it will be displayed with 0 degrees of freedom.

```
                  Sum Sq Mean Sq NumDF DenDF F.value Pr(>F)
SiteClass        1728.72 1728.72     1    23 2.28304 0.1444
Period            685.01  685.01     1    23 0.93448 0.3438
SiteClass:Period  252.36  252.36     1    23 0.34427 0.5631
```

The interaction term is the BACI contrast and the large p-value presents no evidence of an impact.

Never report a naked *p*-value. Use the *LSMeans* for the *Time*Treatment* interaction to estimate the BACI contrast as in previous sections:

$$BACI = \mu_{CA} - \mu_{CB} - (\mu_{TA} - \mu_{TB}) = \mu_{CA} - \mu_{CB} - \mu_{TA} + \mu_{TB}$$

It is not straightforward to estimate the marginal means for the four treatment combinations of *SiteClass* and *Period*, and required a bit of *R* magic.

```
result.lmer.lsmo.SP <- lsmeans::lsmeans(result.lmer, ~SiteClass:Period)
cat("\n\n Estimated marginal means \n\n")
summary(result.lmer.lsmo.SP)
```

```


 Estimated marginal means

 SiteClass Period   lsmean       SE    df   lower.CL upper.CL
 Control   After  13.00000 11.56583 34.47 -10.492878 36.49288
 Impact    After  29.07692 11.56583 34.47   5.584045 52.56980
 Control   Before 31.83333 12.03810 34.47   7.381171 56.28550
 Impact    Before 38.91667 12.03810 34.47  14.464504 63.36883

Confidence level used: 0.95
```

These values are (indirectly) used to estimate the BACI contrast. The estimates of the BACI contrast is available form the *summary()* function applied to the linear model results and looking for the term in the table corresponding to the interaction effect:

```
# Estimate the BACI contrast along with a se
contrast(result.lmer.lsmo.SP, list(baci=c(1,-1,-1,1)))
confint(contrast(result.lmer.lsmo.SP, list(baci=c(1,-1,-1,1))))
```

```
 contrast estimate       SE df t.ratio p.value
 baci     -8.99359 15.32804 23  -0.587  0.5631

 contrast estimate       SE df  lower.CL upper.CL
 baci     -8.99359 15.32804 23 -40.70205 22.71487

Confidence level used: 0.95
```

The estimates of the variance components are:

The *VarCorr()* extractor function applied to the result from the fit gives the variance components:

```
# Extract the variance components
vc <- VarCorr(result.lmer)
vc
```

```
 Groups        Name        Std.Dev.
 SamplingTimeF (Intercept) 31.717
 Residual                  27.075
```

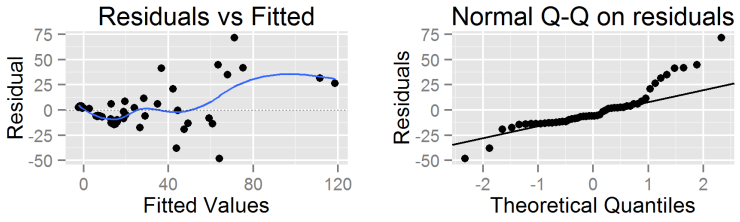
As seen by the plot, there is huge residual variation over and above the variation among sampling times. This large residual variation is what makes it difficult to detect any impact. From the previous analysis, much of this huge residual variation comes from a single sampling time where a very large difference was seen.

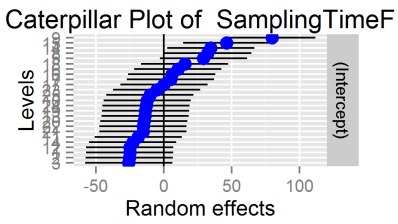
Notice that the sampling times do have an additional structure. For example, it appears that counts of fish vary seasonally. It is possible to fit a model that accounts for this seasonality, e.g. if the actual month of the year was known, but this simple paired analysis will remove the effects of this structure, so this more complicated analysis is not needed.

12.9.3 Model assessment

The very large differences seen in the first analysis is worry some. If you extract the residual and plot them, there impact of these large residual is clear. A Normal quantile plot also indicates evidence of a lack of Normality.

Here is the model assessment plots from the final model on the raw data:





There is some evidence of non-normality in the residuals which isn't too surprising given the huge swing in the differences seen in the earlier analysis.

12.10 Example: Counting chironomids - Paired BACI - Multiple-years B/A; One Site I/C

This example comes from Krebs (1999), Ecological Methodology, 2nd Edition, Box 10.3.

Estimates of chironomid (a type of invertebrate (an ohgochaste worm) present in aquatic sediments) abundance in sediments were taken at one station above and one below a pulp mill outflow pipe for 3 years before plant operation and six years after the plant started.

The raw data is available in the *baci-chironomic.csv* file available at the Sample Program Library at `http://www.stat.sfu.ca/~cschwarz/Stat-650/Notes/MyPrograms`. The data are imported into *R* in the usual way:

```
cat(" BACI design measuring chironomid counts with multiple (paired) yearly measureme
chironomid <- read.csv("baci-chironomid.csv", header=TRUE, as.is=TRUE, strip.white=TR
cat("Listing of part of the raw data \n")
head(chironomid)
```

Part of the raw data are shown below:

```
 BACI design measuring chironomid counts with multiple (paired) yearly measurements b

Listing of part of the raw data
  Year Control.Site Treatment.Site Period
1 1988           14             17 Before
2 1989           12             14 Before
3 1990           15             17 Before
4 1991           16             21  After
5 1992           13             24  After
6 1993           12             20  After
```

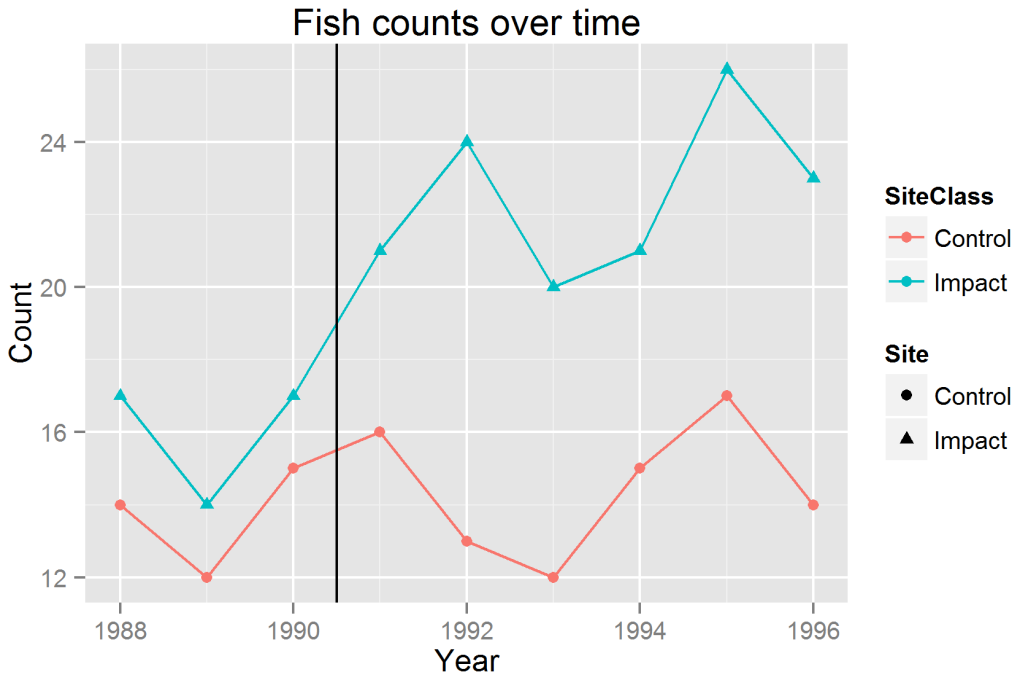
In this study there was one (permanent) site measured in each of the control and impact areas At each site, yearly (paired) readings were taken with three baseline years and five post-baseline years. No sub-sampling was conducted at each year so it is not necessary to average over the sub-samples.

As only one site is measured for each of the control and impact areas, this experiment is an example of pseudo-replication and so conclusions will be limited to inference about these two sites only.

We being by a simple plot of the data over time. We plot of the two series over time:

```
prelimplot <- ggplot(data=chironomid.long,
                     aes(x=Year, y=Count, group=Site, color=SiteClass, shape=Site))+
  ggtitle("Fish counts over time")+
  geom_point()+
  geom_line()+
  geom_vline(xintercept=-0.5+min(chironomid.long$Year[as.character(chironomid.long$Pe
prelimplot
```

giving:



Notice that the two curve track each other over time prior to the impact event, and then the shape are generally the same after the impact event indicating that the pairing by year will be effective in accounting for the natural variation over the years.

There are two (equivalent) ways to analyze this data.

12.10.1 Analysis of the differences

This is an example of simple paired-design where the sampling times (years) induce the pairing in the observations across the sites. We find the difference between the impact and control readings and then plot of the differences over time:

```
chironomid$diff <- chironomid$Treatment.Site - chironomid$Control.Site
head(chironomid)
```

This gives the following final file

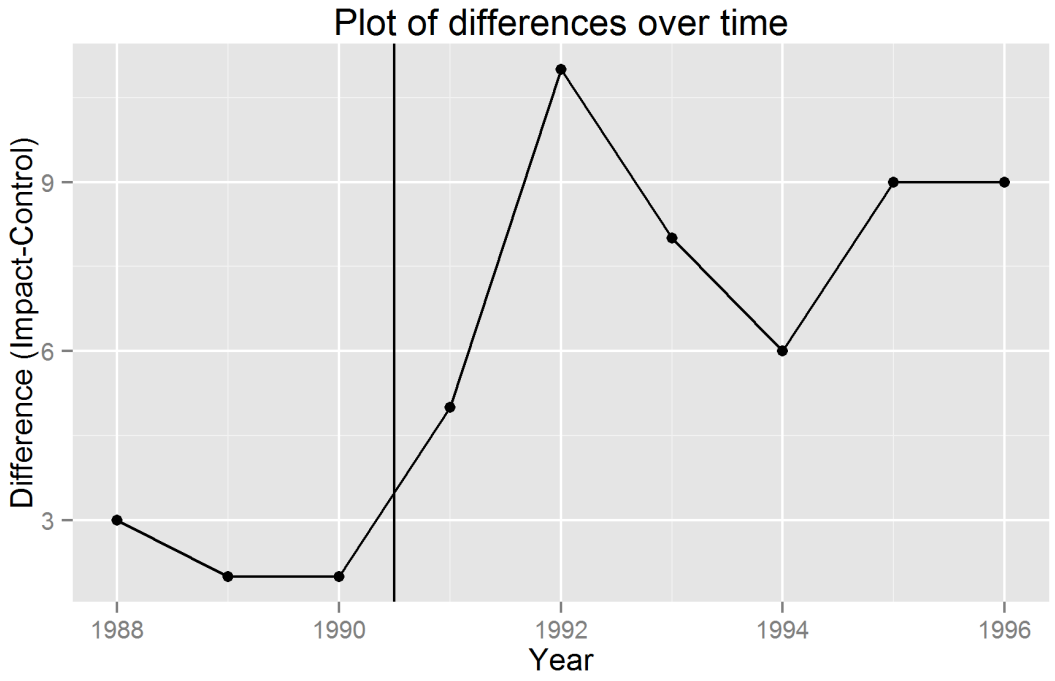
```
  Year Control.Site Treatment.Site Period diff
1 1988           14             17 Before    3
2 1989           12             14 Before    2
3 1990           15             17 Before    2
4 1991           16             21  After    5
```

```
5 1992           13             24  After   11
6 1993           12             20  After    8
```

We plot the difference over time:

```
plotdiff <- ggplot(data=chironomid, aes(x=Year, y=diff))+
  ggtitle("Plot of differences over time")+
  ylab("Difference (Impact-Control)")+
  geom_point()+
  geom_line()+
  geom_vline(xintercept=-0.5+min(chironomid$Year[as.character(chironomid$Period) == "
plotdiff
```

giving:



A two-sample t-test is used to compare the difference between the two *Time* periods. Use the *t.test()* function to test if the mean difference is the same in the before and after Periods. Note that because of the very small effective sample size, only the equal variance t-test can be performed.

```
result <- t.test(diff ~ Period, data=chironomid, var.equal=TRUE)
result$diff.in.means <- sum(result$estimate*c(1,-1))
names(result$diff.in.means)<- "diff.in.means"
result$se.diff <- result$statistic / abs(result$diff.in.means)
names(result$se.diff) <- 'SE.diff'
result
result$diff.in.means
result$se.diff
```

giving:

```
        Welch Two Sample t-test

data:  diff by Period
t = 5.9367, df = 6.187, p-value = 0.0009123
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 3.348040 7.985293
sample estimates:
 mean in group After mean in group Before
            8.000000             2.333333

diff.in.means
     5.666667
 SE.diff
1.047645
```

Notice that we used two-sample t-test allowing for unequal variances in the two periods. If the pooled variance t-test is used it replicates the results from ANOVA in the next section exactly.

The p-value is 0.0009 which provides strong evidence of a difference in the differences between the two periods (the BACI contrast). The estimate of the BACI contrast is -5.67 (SE 0.95 from the unpooled variance analysis) and (SE .1.33) (from the pooled variance analysis).

12.10.2 ANOVA on the raw data

The model for the raw data is

$$Count = SiteClass \quad Period \quad SiteClass * Period \quad Year(R)$$

where the term *SiteClass* represents the control vs. impact effect; the term *Period* represents the before vs. after effect; and the term *SiteClass*Period* represents the BACI effect, i.e. the non-parallel response. Because there is only one site measured in each of the control or treatment areas, the site-effects are completely confounded with the treatment effects (as happens with pseudo-replicated data). Consequently, it is not necessary to enter the *Site* effect into the model.[3] We do have multiple years measured before and after the impact. If the *Year* variable uses unique labels, we can once again enter a random effect for *Year&Random* into the model. Alternatively you could enter *Year (Period)&Random* as the term in the model if the years were not uniquely labelled.

Use the *reshape()* function to stack the Control and Treatment values

```
# We  reshape the data from wide to long format
chironomid.long <- reshape(chironomid, varying=c("Control.Site","Treatment.Site"),
     v.names="Count", direction="long",
     timevar=c("SiteClass"),
     times=c("Control","Impact"),
     drop=c("diff"), idvar=c("Year"))
```

[3] If Site is entered as a random effect, it will be displayed with 0 degrees of freedom.

```
chironomid.long$SiteClass     <-  factor(chironomid.long$SiteClass)
chironomid.long$Site          <-  factor(chironomid.long$Site)
chironomid.long$YearF         <-  factor(chironomid.long$Year)
chironomid.long$Period        <-  factor(chironomid.long$Period)
head(chironomid.long)
```

This gives the following final file showing part of the stacked data:

```
  Year Period SiteClass Count YearF
1 1988 Before   Control    14  1988
2 1989 Before   Control    12  1989
3 1990 Before   Control    15  1990
4 1991  After   Control    16  1991
5 1992  After   Control    13  1992
6 1993  After   Control    12  1993
```

We again use the *lme()* function. Notice the *R* magic to get the *Year(R)* term! **You must insure that the numeric variable *Year* is treated as a factor and not as a continuous covariate.**

```
# Because there is ONLY one measurement per year, the SamplingTime*Site and
# residual variance are total confounded and cannot be separated. This is
# the residual term.
result.lmer <- lmer(Count ~ SiteClass+Period+SiteClass:Period + (1|YearF),
                data=chironomid.long)
anova(result.lmer, ddf="Kenward-Roger")
```

The *anova(result.lme)* function applied to the result from the fit, gives the effect tests. Note that when using the *lme()* function, you no longer have to worry about Type I, II, or III tests. However, *R* does not have the Kenward-Rogers approximations to the degrees of freedom for unbalance data as does *SAS* and *JMP*. Consequent, *R* uses a different approximation to the degrees of freedom and the results may be (slightly) different from those of *SAS* or *JMP*.

The tests for the fixed effects are:

```
Analysis of Variance Table of type 3  with  Kenward-Roger
approximation for degrees of freedom
                 Sum Sq Mean Sq NumDF DenDF F.value    Pr(>F)
SiteClass        168.056 168.056     1     7  60.604 0.0001084 ***
Period            16.047  16.047     1     7   9.108 0.0194448 *
SiteClass:Period  32.111  32.111     1     7  18.225 0.0037048 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The interaction term is the BACI contrast and the small p-value (0.0037) indicates strong evidence of an impact. Notice that the p-value from the test for no interaction from the ANOVA matches the p-value of the Period effect in the t-test assuming that the variances were equal in the two periods.

Never report a naked *p*-value. Use the *LSMeans* for the *Time*Treatment* interaction to estimate the BACI contrast as in previous sections:

$$BACI = \mu_{CA} - \mu_{CB} - (\mu_{TA} - \mu_{TB}) = \mu_{CA} - \mu_{CB} - \mu_{TA} + \mu_{TB}$$

It is not straightforward to estimate the marginal means for the four treatment combinations of *SiteClass* and *Period*, and required a bit of *R* magic.

```
result.lmer.lsmo.SP <- lsmeans::lsmeans(result.lmer, ~SiteClass:Period)
cat("\n\n Estimated marginal means \n\n")
summary(result.lmer.lsmo.SP)
```

```


 Estimated marginal means

 SiteClass Period   lsmean        SE    df lower.CL upper.CL
 Control   After  14.50000 0.7993053 10.84 12.73748 16.26252
 Impact    After  22.50000 0.7993053 10.84 20.73748 24.26252
 Control   Before 13.66667 1.1303883 10.84 11.17409 16.15924
 Impact    Before 16.00000 1.1303883 10.84 13.50743 18.49257

Confidence level used: 0.95
```

These values are (indirectly) used to estimate the BACI contrast. The estimates of the BACI contrast is available form the *summary()* function applied to the linear model results and looking for the term in the table corresponding to the interaction effect:

```
# Estimate the BACI contrast along with a se
contrast(result.lmer.lsmo.SP, list(baci=c(1,-1,-1,1)))
confint(contrast(result.lmer.lsmo.SP, list(baci=c(1,-1,-1,1))))
```

```
 contrast  estimate       SE df t.ratio p.value
 baci     -5.666667 1.327368  7  -4.269  0.0037

 contrast  estimate       SE df  lower.CL  upper.CL
 baci     -5.666667 1.327368  7 -8.805392 -2.527941

Confidence level used: 0.95
```

The estimated mean difference-in-the-difference is -5.67 with a standard error of 1.33 which is the same as found in the t-test earlier.

The estimates of the variance components are:

The *VarCorr()* extractor function applied to the result from the fit gives the variance components:

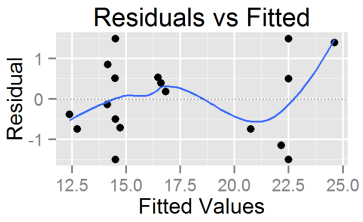
```
# Extract the variance components
vc <- VarCorr(result.lmer)
vc
```

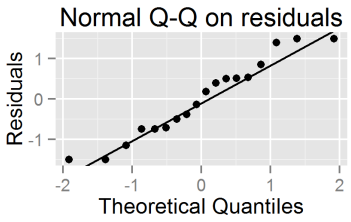
```
 Groups   Name        Std.Dev.
 YearF    (Intercept) 1.4392
 Residual             1.3274
```

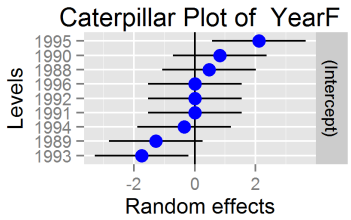
The year-to-year variance is comparable in size to the residual variation which implies that the year effects are about the same size as measurement error.

12.10.3 Model assessment

Here is the model assessment plots from the final model on the raw data:







There is no evidence of any problems in the diagnostic plots.

12.11 Example: Fry monitoring - BACI with Multiple sites; Multiple years before/after

This example is based (loosely) on a consulting project from an Independent Power Producer who was interested in monitoring the effects of an in-stream hydroelectric project. The response variable for the project was the minnow density at different locations in the stream.

The monitoring design has the river divided into six segments of which three are upstream of the diversion and three are downstream of the diversion. In each reach, several sites have been located where minnow fry congregate. In each of the sites, minnow traps are set for various lengths of time. At the end of the soaking period, the traps are removed and the number of minnows are counted and classified by species. The counts are standardized to a common period of time to adjust for the different soak-times. [This could be done directly in the analysis by using the soak-time as a covariate, but the soak-time data was not available.]

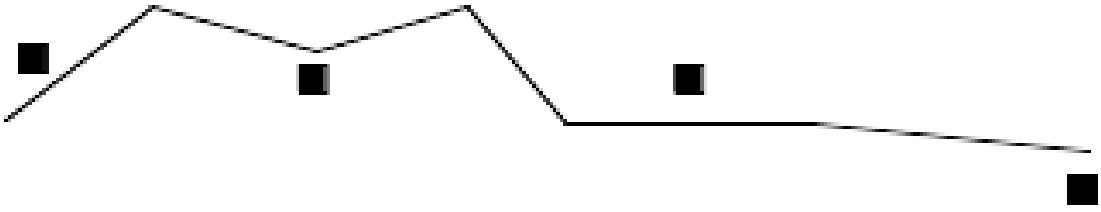
An initial pre-project monitoring study was run in 2000 to 2002. The project became operational in late 2002, and post-project monitoring continued in 2003 and 2004.

This design is a variant of a classic Before-After-Control-Impact (BACI) design where presumably the control sites will be located in the upper reaches and potentially impacted sites in the lower reaches.

12.11.1 A brief digression

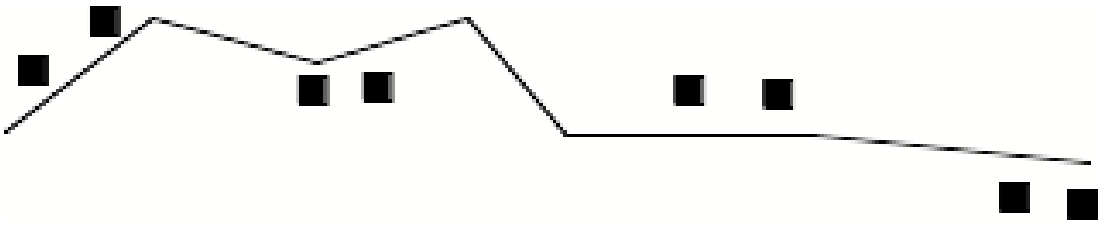
A major concern for this study is proper replication and the avoidance of pseudo-replication (Hurlburt, 1984). Replication provides information about the variability of the collected data under identical treatment conditions so that differences among treatments can be compared to variation within treatments. This is the fundamental principle of ANOVA.

Consider first a survey to investigate fry density at various locations on a river. A simple design may take a single sample (i.e. single fry trap) at each of 4 locations



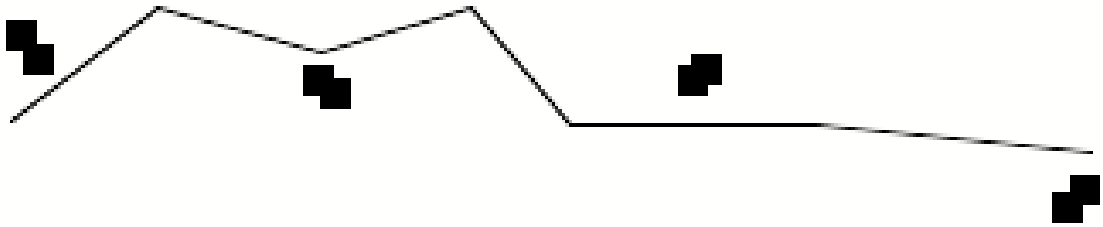
These four values are insufficient for any comparison of the fry density across the four locations because the natural variation present in readings at a particular location is not known.

In many ecological field studies, the concepts of experimental units and randomization of treatments to experimental units are not directly applicable making "replication" somewhat problematic. Replication is consequently defined as the taking of multiple INDEPENDENT samples from a particular location. The replicated samples should be located sufficiently far from the first location so that local influences that are site specific do not operate in common on the two samples. The exact distance between samples depends upon the biological process. For example if the locations are tens of kilometers apart, then spacing the samples hundreds of meters apart will likely do for most situations. This gives rise to the following design where two SITES are sampled at each location, each with a single trap operating:



Now a statistical comparison can be performed to investigate if the mean response is equal at four locations. The key point is that the samples should be independent but still representative of that particular location. Hence, taking two samples from the exact same location, or splitting the sample in two and doing two analyses on the split sample will not provide true replication.

Consequently, a design where duplicate samples (e.g. multiple traps) are taken from the exact same location and time would be an example of pseudo-replication



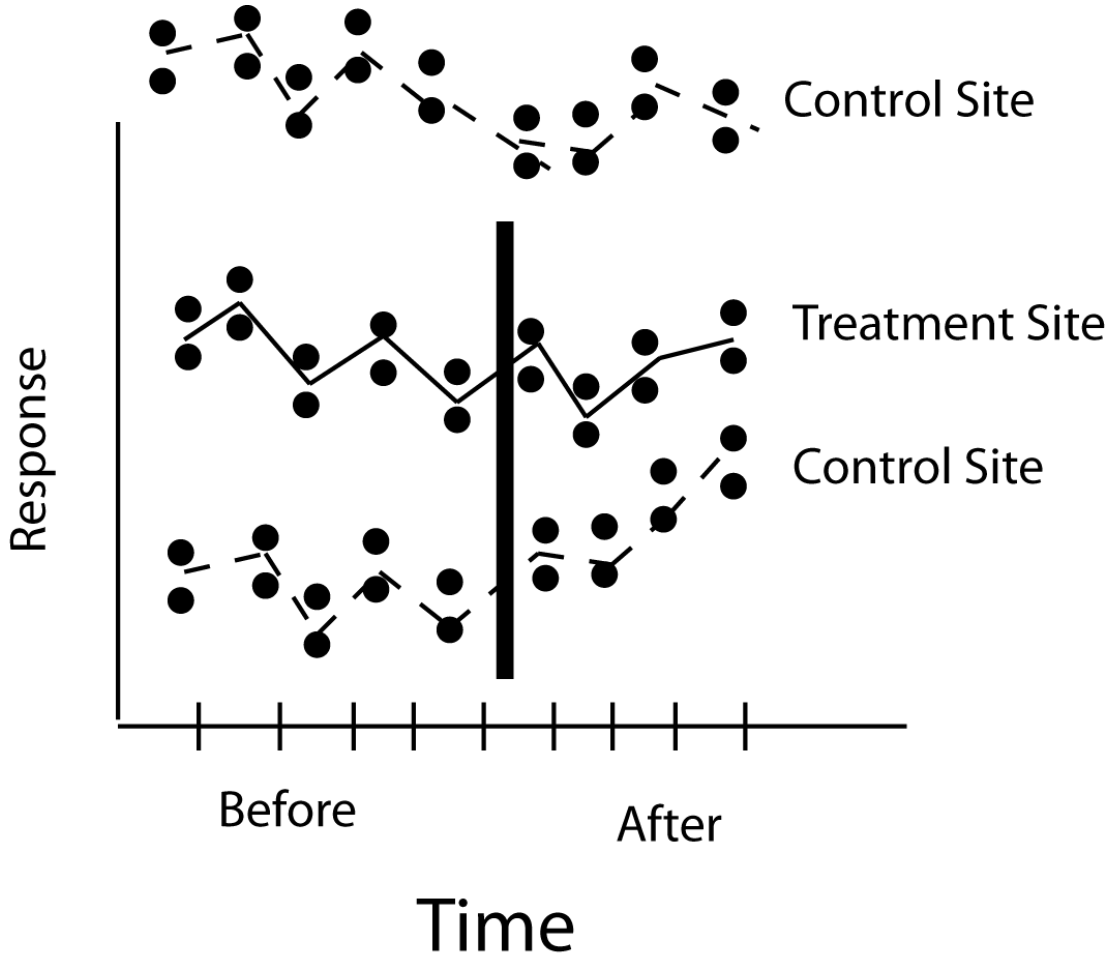
Note that the data from the last two designs looks "identical", i.e. pairs of "replicated" observations from four locations. Consequently, it would be very tempting to analyze both experiments using exactly the same statistical model. However, there is a major difference in interpretation of the results.

The second design with real replicates enables statements to be made about differences in the mean response among those four general locations. However, the third design with pseudo-replication only allows statements to be made about differences among those four particular sampling sites which may not truly reflect differences among the broader locations.

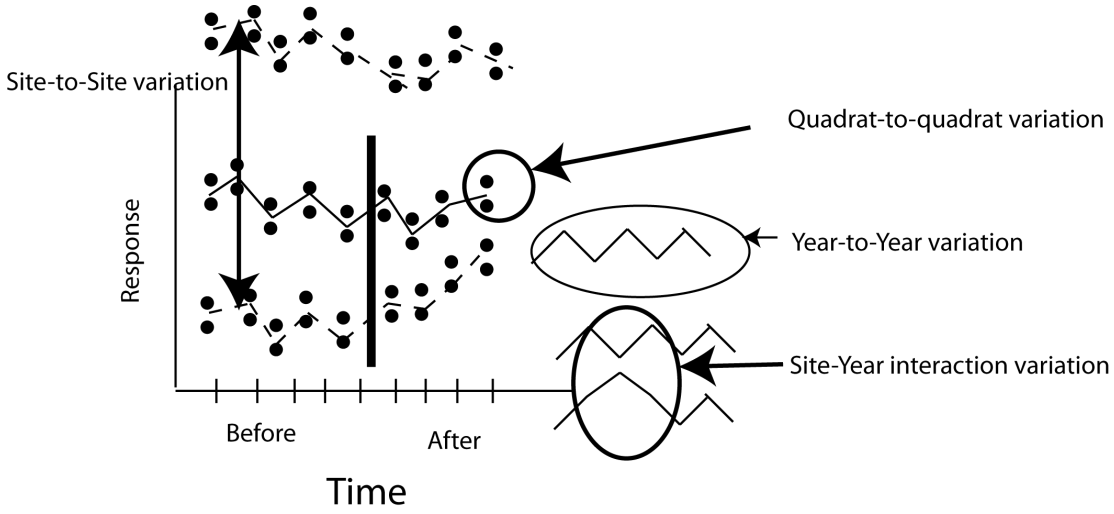
Obviously the line between real and pseudo-replication is somewhat ill-defined. Exactly how far apart do sampling sites have to be before they can be considered to be independent. There is no hard and fast rule and biological consideration and knowledge of the processes involved in the environmental impact must be used to make a judgment call.

In terms of the current monitoring design, we hope that the different sites are that are far enough apart to be considered independent of each other. The multiple minnow traps within each site would be considered pseudo-replicates and cannot be treated as independent across sites.

Here is a schematic of this general BACI design:



As before, we can indentify the Period (before vs. after), SiteClass (impact vs. control) and the BACI interaction. As well there are several sources of variation in the monitoring design that influence how much sampling and when sampling should take place.



At the lowest level is trap-to-trap variation within a single site-year combination. Not all traps will

catch the exact same number of minnows. The number of minnows captured is also a function of how long the traps soak with longer soak periods leading, in general, to more minnows being captured. However, by converting to a common soak-time, this extra variation has been reduced.

Next is site-to-site variation within the reaches of the river. This corresponds to natural variation in minnow density among sites due to habitat and other ecological variation.

Thirds is yearly variation. The number of minnows captured within a site varies over time (e.g. 2002 vs. 2003) because of uncontrollable ecological factors. For example, one year may have more precipitation or be warmer which may affect the fry density in all locations simultaneously.

Finally, there is the site-year interaction which represents the INconsistency in how the sites respond over time. For example, not all of the control sites respond exactly the same way across the multiple years prior to impact. Note that if only trap is measured each year, then the trap-to-trap and the site-year interaction variance components are completely confounded together and cannot be separated.

12.11.2 Some preliminary plots

The raw data is available in the *baci-fry.csv* file available at the Sample Program Library at `http://www.stat.sfu.ca/~cschwarz/Stat-650/Notes/MyPrograms`. The data are imported into *R* in the usual way:

```
sink('baci-fry-R-001.txt', split=TRUE)
fry <- read.csv('baci-fry.csv', header=TRUE, as.is=TRUE, strip.white=TRUE)
fry$logfry       <- log(fry$Fry)
fry$Period       <- factor(fry$Period)
fry$Site         <- factor(fry$Site)
fry$SiteClass    <- factor(fry$SiteClass)
fry$Period       <- factor(fry$Period)
fry$YearF        <- factor(fry$Year)
fry[1:12,]
```

Part of the raw data are shown below:

```
    SiteClass Site Sample Year Period Fry   logfry YearF
1  Downstream    D      1 2000 Before 185 5.220356  2000
2  Downstream    E      1 2000 Before 258 5.552960  2000
3  Downstream    E      2 2000 Before 630 6.445720  2000
4  Downstream    E      3 2000 Before 116 4.753590  2000
5  Downstream    F      1 2000 Before 208 5.337538  2000
6  Downstream    F      2 2000 Before 261 5.564520  2000
7  Downstream    D      1 2001 Before 321 5.771441  2001
8  Downstream    E      1 2001 Before  65 4.174387  2001
9  Downstream    E      2 2001 Before 755 6.626718  2001
10 Downstream    F      1 2001 Before 476 6.165418  2001
11 Downstream    F      2 2001 Before 521 6.255750  2001
12 Downstream    E      1 2002 Before 198 5.288267  2002
```

As you will see in a few moment, there is a good reason to analyze the $\log(fry)$ rather than the fry count directly.

It is always desirable to examine the data closely for unusual patterns, outliers, etc.

First, let us examine the pattern of when sites are measured over time and how many traps are used each time, etc. In *R*, we use the *summaryBy()* function in the *doBy* package.

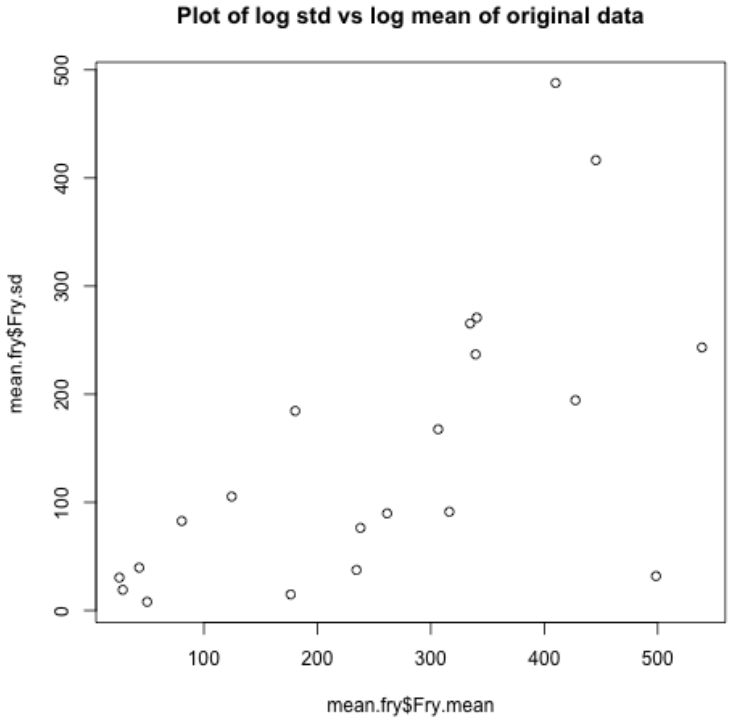
```
mean.fry <- ddply(fry, c("Year","YearF","Site","SiteClass","Period"), function(x){
  n <- nrow(x)
  nmiss <- sum(is.na(x$Fry))
  mean.fry <- mean(x$Fry, na.rm=TRUE)
  sd.fry   <- sd(  x$Fry, na.rm=TRUE)
  mean.logfry <- mean(x$logfry, na.rm=TRUE)
  sd.logfry   <- sd  (x$logfry, na.rm=TRUE)
  res<- data.frame(n, nmiss, mean.fry, sd.fry, mean.logfry, sd.logfry)
  })
mean.fry[1:5,]
```

giving:

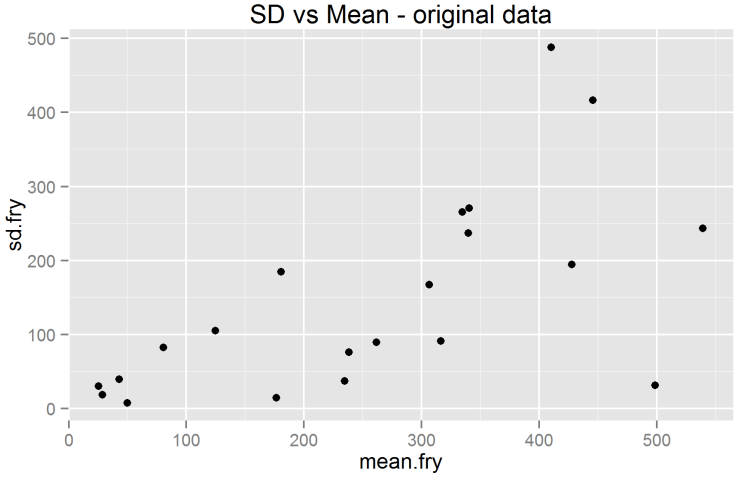
```
  Year YearF Site  SiteClass Period n nmiss mean.fry    sd.fry mean.logfry sd.logfry
1 2000  2000    A   Upstream Before 2     0  25.5000  30.40559    2.618221 1.7422073
2 2000  2000    B   Upstream Before 2     0 306.5000 167.58431    5.644266 0.5767497
3 2000  2000    C   Upstream Before 2     0 124.5000 105.35891    4.602664 0.9767137
4 2000  2000    D Downstream Before 1     0 185.0000        NA    5.220356        NA
5 2000  2000    E Downstream Before 3     0 334.6667 265.43800    5.584090 0.8464942
```

The design is unbalanced with some sites having from 1 to 3 samples taken and not all sites being measured in all years (e.g. site D is only measured in 2000, 2001, and 2003).

The variability in the number of fry is very large with the standard deviation often being of the same magnitude as the mean number of fry. It is often quite useful to plot the $\log(std\ dev)$ vs. the $\log(mean)$ to see if there is a relationship between the two. The slope of the line of the relationship can be used to determine the appropriate transformation based on **Taylor's Power Law** which is covered elsewhere in these notes.



ANOVA assumes that variation in the data is roughly constant – here the differences in variation are too large for ANOVA.[4] Based on the relationship between the $\log(std\ dev)$ and the $\log(mean)$, the natural logarithmic transformation or $log(fry)$[5] is to be used. This was the reason why the $\log(fry)$ was computed earlier when the data wes imported by *R*. A plot of the standard deviation vs. the mean on the log-scale shows a dramatic improvement:



[4] A rough rule of thumb is that the standard deviations in various groups should not vary by a factor of more than about 5:1.

[5] In statistics, the $log()$ transformation is always the natural logarithm. If you want the common (base 10) logarithm, use the $log10()$ function.

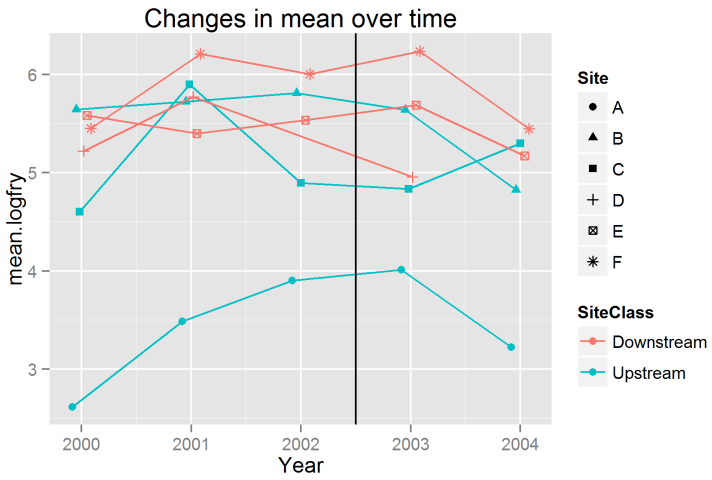
with the standard deviation not appears to be dependent upon the mean.

Note that the average of values on the log-scale corresponds to the logarithm of the geometric mean on the ordinary scale. The geometric mean is less sensitive to outliers and will always be less than the arithmetic mean. Differences on the log-scale correspond to RATIOS on the anti-log scale. For example, if the counts of fry are reduced by a factor of 2 (i.e. halved), this corresponds to a subtraction on the log-scale by a value of 0.69.

Next look at the plot of the mean log(fry) density over time.

The *plot()* function with the *apply()* function plot the site means over time:

```
meanplot <- ggplot(data=mean.fry, aes(x=Year, y=mean.logfry,
                   group=Site, color=SiteClass, shape=Site))+
  ggtitle("Changes in mean over time")+
  geom_point(position=position_dodge(w=0.2))+
  geom_line(position=position_dodge(w=0.2))+
  geom_vline(xintercept=-0.5+min(mean.fry$Year[as.character(mean.fry$Period) == "Afte
meanplot
```



There is no obvious evidence of an impact of the diversion. There is some evidence of a year-to-year effect with the curve generally moving together (e.g. the downward movement in 2004 in all sites).

12.11.3 Analysis of the averages

We start first with the analysis of the average log(fry) count. The number of traps monitored in each year varies from 1 to 3 so the results from this analysis will not match precisely those from the analysis of the raw data but the results should be "close enough."

We previously saw the following (partial) table of means:

```
  Year YearF Site  SiteClass Period n nmiss mean.fry    sd.fry mean.logfry sd.logfry
```

```
1 2000  2000    A   Upstream Before 2     0  25.5000  30.40559    2.618221 1.7422073
2 2000  2000    B   Upstream Before 2     0 306.5000 167.58431    5.644266 0.5767497
3 2000  2000    C   Upstream Before 2     0 124.5000 105.35891    4.602664 0.9767137
4 2000  2000    D Downstream Before 1     0 185.0000        NA    5.220356        NA
5 2000  2000    E Downstream Before 3     0 334.6667 265.43800    5.584090 0.8464942
```

The model to be fit is:

$$MeanLogFry = SiteClass\ \ Site(R)\ \ Period\ \ Year(R)\ \ SiteClass*Period$$

where *SiteClass* is the classification of sites as either Impact or Control; *Site(R)* is the random site effect within *SiteClass*[6] *Period* is the classification of time as before or after, *Year(R)* represents the multiple years within each period[7] and the *SiteClass*Period* term represent the BACI contrast of interest.

The response variable is the mean log(fry). The fixed effects are *SiteClass*, *Period* and their interaction (which is the BACI contrast of interest).

Next, the sites again serve as the larger experimental unit in a split-plot sense. If the sites are uniquely labelled, then the site-to-site variation can be entered as *Site&Random* as in previous examples.

The year-to-year variation is entered in a similar fashion, i.e. if the years are uniquely labelled, the simply enter *Year&Random*.

When ever there are more than one random effects in a model, the usual practice is to enter the "interaction" of the random effects as well (the *Site*Year* interaction). This term allows the site effect to vary among years. If there is a large site-by-year interaction, this is worrysome as it implies that the effects of years are not consistent over time and vice versa. In this case, because we are dealing with averages, we don't need to add such a term and it will be 'automatically' fit as the residual variance components. If you do add this term, you get exactly the same results.

We again use the *lme()* function. **You must insure that the numeric variable *Year* is treated as a factor and not as a continuous covariate.** The *lme()* function is not very easy to use when you have multiple, non-nested random factors and some *R* magic is needed as noted in the code.

```
# Notice that we use the YearF rather than Year in the random statements
# Because we are analyzing the mean, the Site:Year term is the residual error
result.mean.lmer <- lmer(mean.logfry ~ Period + SiteClass + Period:SiteClass+
                      (1|YearF)+(1|Site),
                   data=mean.fry)
anova(result.mean.lmer, ddf="Kenward-Roger")
```

The *anova(result.lme)* function applied to the result from the fit, gives the effect tests. Note that when using the *lme()* function, you no longer have to worry about Type I, II, or III tests. However, *R* does not have the Kenward-Rogers approximations to the degrees of freedom for unbalance data as does *SAS* and *JMP*. Consequent, *R* uses a different approximation to the degrees of freedom and the results may be (slightly) different from those of *SAS* or *JMP*.

[6] As noted in previous chapters, if you label each site with a unique label, it is not necessary to use the nesting notation of *Site(SiteClass)(R)*. The sites are considered a random effect because we wish to make inference about the impact over all sites in the study area.

[7] As noted in previous chapters, if you label each year with a unique label, it is not necessary to use the nesting notation of *Year(Period)(R)*. The years are considered a random effect because we wish to make inference about the impact over all year in the study area.

The *lmer()* function could also be used, but this function does not report the traditional p-values (don't ask!) and so we won't use it in these course notes.

REML is a variant of maximum likelihood estimation and extract the maximal information from the data. With balanced data, the F-tests are identical to an older method called the Expected Mean Squares (EMS) method. Modern statistical theory prefers REML over the EMS methods.

The REML method comes with two variants. In one variant (the default with *JMP*), estimated variance components are allowed to go negative. This might seem to be problematic as variances must be non-negative! However, the same problem can occur with the EMS method, and usually only occurs when sample sizes are small and variance components are close to 0. It turns out that the F-tests are "unbiased" when the unbounded variance component option is chosen even if some of the variance estimates are negative. As noted in the *JMP* help files,

> "If you remain uncomfortable about negative estimates of variances, please consider that the random effects model is statistically equivalent to the model where the variance components are really covariances across errors within a whole plot. It is not hard to think of situations in which the covariance estimate can be negative, either by random happenstance, or by a real process in which deviations in some observations in one direction would lead to deviations in the other direction in other observations. When random effects are modeled this way, the covariance structure is called compound symmetry.
>
> So, consider negative variance estimates as useful information. If the negative value is small, it can be considered happenstance in the case of a small true variance. If the negative value is larger (the variance ratio can get as big as 0.5), it is a troubleshooting sign that the rows are not as independent as you had assumed, and some process worth investigating is happening within blocks.

So if a variance component is negative and has a large absolute value, this is an indication that your model may not be appropriate for the data at hand. Please contact me for assistance. The default for *SAS* is to constrain the variance components to be non-negative and so the F-tests may vary slightly from those of *JMP*.

The Fixed Effect Test are:

```
Analysis of Variance Table of type 3  with  Kenward-Roger
approximation for degrees of freedom
                   Sum Sq  Mean Sq NumDF   DenDF F.value Pr(>F)
Period           0.041576 0.041576     1  2.9797 0.29868 0.6230
SiteClass        0.240604 0.240604     1  4.0194 1.59568 0.2748
Period:SiteClass 0.019793 0.019793     1 17.1098 0.13496 0.7179
```

Because *R* does NOT have the Kenward-Rogers correction for the df, the p-values won't match those from *SAS* or *JMP*.

There is no evidence of a BACI effect as the p-value is very large Never simply report naked p-values, so an estimate of the BACI contrast is needed. The BACI effect is the difference in the differences, i.e.

$$BACI = \mu_{CA} - \mu_{CB} - (\mu_{TA} - \mu_{TB}) = \mu_{CA} - \mu_{CB} - \mu_{TA} + \mu_{TB}$$

where C is Control, T is treatment site classification, B is before, and A is after the intervention. It is

not straightforward to estimate the marginal means for the four treatment combinations of *SiteClass* and *Period*, and required a bit of *R* magic.

```
result.mean.lmer.lsmo.SP <- lsmeans::lsmeans(result.mean.lmer, ~SiteClass:Period)
cat("\n\n Estimated marginal means \n\n")
summary(result.mean.lmer.lsmo.SP)
```

```


 Estimated marginal means

 SiteClass  Period   lsmean        SE   df lower.CL upper.CL
 Downstream After  5.421711 0.5192516 5.73 4.136593 6.706830
 Upstream   After  4.639403 0.5124357 5.47 3.355235 5.923572
 Downstream Before 5.624001 0.4973746 4.98 4.343763 6.904239
 Upstream   Before 4.732006 0.4943492 4.87 3.451018 6.012995

Confidence level used: 0.95
```

These values are (indirectly) used to estimate the BACI contrast. The estimates of the BACI contrast is available form the *summary()* function applied to the linear model results and looking for the term in the table corresponding to the interaction effect:

```
# Estimate the BACI contrast along with a se
contrast(result.mean.lmer.lsmo.SP, list(baci=c(1,-1,-1,1)))
confint(contrast(result.mean.lmer.lsmo.SP, list(baci=c(1,-1,-1,1))))
```

```
 contrast   estimate        SE    df t.ratio p.value
 baci     -0.1096867 0.2985781 17.11  -0.367  0.7179

 contrast   estimate        SE    df   lower.CL  upper.CL
 baci     -0.1096867 0.2985781 17.11 -0.7393237 0.5199502

Confidence level used: 0.95
```

The BACI estimate is small relative to the standard error which is not surprising given that the hypothesis of no BACI effect was not rejected.

Because we analyzed the data on the log-scale, the estimated effect on the anti-log scale is $exp(-.11) = .895$ which is interpreted to say that the change in fry density between before and after is .89 as large in the impact area as in the control area.

The estimated variance components can be obtained.
It is more difficult to extract the variance components when using the previous trick to cross the random factors of *Site* and *Year*:

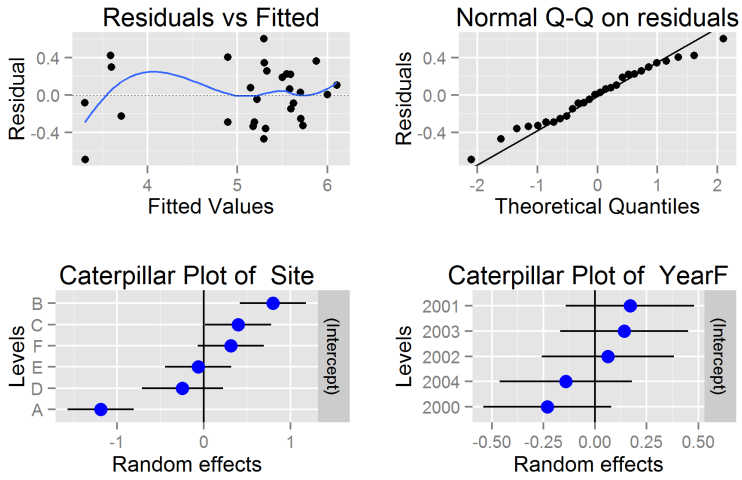
```
# Extract the variance components
```

```
vc <- VarCorr(result.mean.lmer)
vc
```

```
 Groups   Name        Std.Dev.
 Site     (Intercept) 0.78987
 YearF    (Intercept) 0.24617
 Residual             0.38205
```

The variance components show that site-to-site effects are about $10\times$ larger than year-to-year effects (see the earlier plot) and about $4\times$ that of the residual variation (of the averages). Because we are dealing with the averages, the residual variation is a composite of the site*year interaction and cage-to-cage variation (see next section).

As always, don't forget to look at the diagnostic plots: Here is the model assessment plots from the final model on the raw data:



12.11.4 Analysis of the raw data

The raw data can also be analyzed directly. The model to be fit is:

$$LogFry = SiteClass\ \ Site(R)\ \ Period\ \ Year(R)\ \ \ SiteClass*Period\ \ Site*Year(R)$$

where *SiteClass* is the classification of sites as either Impact or Control; *Site(R)* is the random site effect within SiteClass[8] *Period* is the classification of time as before or after, *Year(R)* represents the multiple years within each periodfootnoteAs noted in previous chapters, if you label each year with a unique label, it is not necessary to use the nesting notation of Year(Period)(R). The years are considered a random effect because we wish to make inference about the impact over all year in the study area. and the *SiteClass*Period* term represent the BACI contrast of interest. Because subsampling occurs at each combination of site and year, so need a "interaction" of the two random effects *Year*Site&Random*.

We again use the *lme()* function. **You must insure that the numeric variable *Year* is treated as a factor and not as a continuous covariate.** The *lme()* function is not very easy to use when you have

[8] As noted in previous chapters, if you label each site with a unique label, it is not necessary to use the nesting notation of Site(SiteClass)(R). The sites are considered a random effect because we wish to make inference about the impact over all sites in the study area.

multiple, non-nested random factors and some *R* magic is needed as noted in the code. Notice that we need three random effects and that the *Site*Year* is NOT an interaction between the two random effects, but a separate random effect for each year-site combination to allow for the multiple cages (subsamples) within the year-site combination.

```
# Notice that we use the YearF variable in the random effects
result.lmer <- lmer(logfry ~ Period + SiteClass + Period:SiteClass +
                     (1|YearF)+ (1|Site)+(1|YearF:Site), data=fry)
anova(result.lmer, ddf="Kenward-Roger")
```

The *anova(result.lme)* function applied to the result from the fit, gives the effect tests. Note that when using the *lme()* function, you no longer have to worry about Type I, II, or III tests. However, *R* does not have the Kenward-Rogers approximations to the degrees of freedom for unbalance data as does *SAS* and *JMP*. Consequent, *R* uses a different approximation to the degrees of freedom and the results may be (slightly) different from those of *SAS* or *JMP*.

The *lmer()* function could also be used, but this function does not report the traditional *p*-values (don't ask!) and so we won't use it in these course notes.

The Fixed Effect Tests again show no evidence of a BACI effect:

```
Analysis of Variance Table of type 3  with  Kenward-Roger
approximation for degrees of freedom
                 Sum Sq Mean Sq NumDF   DenDF F.value Pr(>F)
Period           0.09780 0.09780     1  3.0855 0.19241 0.6898
SiteClass        0.93625 0.93625     1  4.0254 1.73858 0.2573
Period:SiteClass 0.00678 0.00678     1 16.1019 0.01264 0.9119
```

Because *R* does NOT have the Kenward-Rogers correction for the df, the *p*-values won't match those from *SAS* or *JMP*.

The estimate of the BACI contrast:

$$BACI = \mu_{CA} - \mu_{CB} - (\mu_{TA} - \mu_{TB}) = \mu_{CA} - \mu_{CB} - \mu_{TA} + \mu_{TB}$$

where C is Control, T is treatment site classification, B is before, and A is after the intervention is found in the usual way

It is not straightforward to estimate the marginal means for the four treatment combinations of *SiteClass* and *Period*, and required a bit of *R* magic.

```
result.lmer.lsmo.SP <- lsmeans::lsmeans(result.lmer, ~SiteClass:Period)
cat("\n\n Estimated marginal means for SiteClass:Period \n\n")
summary(result.lmer.lsmo.SP)
```

```
 Estimated marginal means for SiteClass:Period

 SiteClass  Period   lsmean        SE   df lower.CL upper.CL
 Downstream After  5.475293 0.5372050 6.19 4.170457 6.780130
 Upstream   After  4.617077 0.5235770 5.74 3.321818 5.912336
 Downstream Before 5.625547 0.5058467 5.05 4.329121 6.921973
 Upstream   Before 4.719723 0.4926974 4.65 3.424177 6.015269

Confidence level used: 0.95
```

These values are (indirectly) used to estimate the BACI contrast. The estimates of the BACI contrast is available form the *summary()* function applied to the linear model results and looking for the term in the table corresponding to the interaction effect:

```
# Estimate the BACI contrast along with a se
contrast(result.lmer.lsmo.SP, list(baci=c(1,-1,-1,1)))
confint(contrast(result.lmer.lsmo.SP, list(baci=c(1,-1,-1,1))))
```

```
 contrast    estimate        SE   df t.ratio p.value
 baci     -0.04760761 0.4234021 16.1  -0.112  0.9119

 contrast    estimate        SE   df   lower.CL  upper.CL
 baci     -0.04760761 0.4234021 16.1 -0.9447187 0.8495035

Confidence level used: 0.95
```

Despite the value appears to be 'smaller' than that obtained by the analysis of the averages, the difference is just an artefact of the differing number of samples taken in some years.

The estimated variance components:

It is more difficult to extract the variance components when using the previous trick to cross the random factors of *Site* and *Year*:

```
# Extract the variance components
vc <- VarCorr(result.lmer)
vc
```

```
 Groups     Name        Std.Dev.
 YearF:Site (Intercept) 0.00000
 Site       (Intercept) 0.76697
 YearF      (Intercept) 0.21415
 Residual               0.72347
```

tell a similar story as happened in the analysis of the averages. Note that the estimate of the *Year*Site* variance component is very small (and actually is reported as being negative by *SAS* or *JMP* if the

variance components are left unbounded). Of course this is 'silly' as variances must be positive, but just indicates that the actual variance component is small (i.e. close to 0) and the negative value is an artefact of the data. You could re-run the model constraining the variance components to be non-negative and the results are similar.

Also note that because we have analyzed the individual measurements, we have separated the residual variance component (from the analysis of the averages in the previous section) into its site*year interaction and cage-to-cage variation components. Ignoring for now the problem of 'negative variance components', you see that the residual variance from the analysis of the averages (0.15 as reported by *SAS* and *JMP*) which was a composite of site*year and cage-to-cage (residual) variation is approximately equal to the current site*year component and the variance of the cage averages (based on an average of 1.85 cages per year) or $0.15 \approx -0.19 + 0.65/1.85 = 0.16$

12.11.5 Power analysis

A power analysis of these types of designs is vary difficult to do by hand (!) and still surprising tedious (but possible) to do in *JMP*. Here is where *SAS* and *R* are extremely useful as you construct a power analysis fairly simply as outlined in the papers:

> Stroup, W. W. (1999).
> Mixed model procedures to assess power, precision, and sample size in the design of experiments.
> Pages 15-24 in Proc. Biopharmaceutical Section.
> American Statistical Association, Baltimore, MD.
> Available at: `http://www.stat.sfu.ca/~cschwarz/Stat-650/Notes/MyPrograms/Power/stroup-1999-power.pdf`.

or

> Littell, R.C., Milliken, G.A., Stroup, W.W., Wolfinger, R.D., and Schabenberger, O. (2006),
> *SAS* for Mixed Models. 2nd Edition. Chapter 12.

Some examples of the use of this method are available at: `http://www.stat.sfu.ca/~cschwarz/Stat-650/Notes/MyPrograms/Power/`.

The basic idea of Stroup's method is to generate dummy data with NO variability and taking on only the means of the treatment applied to this observation. In the fry example, you would generate data corresponding to the 5 years of data, 6 sites, and multiple fry traps with the log(fry) values taking one of the 4 means corresponding to the treatment-before impact, treatment after-impact, control-before impact, and control-after impact. The BACI contrast among these values should represent a biologically interesting difference. This "mean" data is then analyzed using the BACI model using estimates of the unknown variance components you think are appropriate for your study. These values could be obtained from real life data (as above) or from expert opinion. Then a simple function of the F-statistic gives you the estimated power of your design.

An *R* function illustrating the method is available on the Sample Program Library (`http://www.stat.sfu.ca/~cschwarz/Stat-650/Notes/MyPrograms`). For this study, we fixed the estimated variance components as

- σ_{site} - STANDARD DEVIATION of the site-to-site effects = 0.75;
- σ_{year} - STANDARD DEVIATION of the year-to-year effects = 0.25;
- $\sigma_{site*year}$ - STANDARD DEVIATION of the site-by-year interaction = 0. 10;
- $\sigma_{residual}$ - STANDARD DEVIATION of the residual variation = 0.75.

Note that because the same sites are measured in all years (both before and after the impact starts), the sites serve as "blocks" and so the site-to-site variation does NOT affect the power. Similarly, because all years have all sites measured, the years also serve as "blocks" and so the year-to-year variation also does not affect power. Only the site-by-year variance and the residual variance impact the power of the design directly.

We also examined a number of scenarios for different number of years and different number of subsamples. Here are some results:

```
     alpha n_TA n_TB n_CA n_CB ny_B ny_A ns_T ns_C baci power
[1,] 0.05  3    3    3    3    3    2    3    3    -0.5 0.2996488
[2,] 0.05  6    6    6    6    3    2    3    3    -0.5 0.5068123
[3,] 0.05  9    9    9    9    3    3    3    3    -0.5 0.7619485
[4,] 0.05  6    6    6    6    3    4    3    3    -0.5 0.6722755
```

Given the very large sub-sampling STANDARD DEVIATIONS, it certainly is useful to increase the subsampling from 3 to 9 units per year-site combination to obtain power around the target of 80%. If the sampling is extended in the future, i.e. more years of sampling in the future, the power also increases.

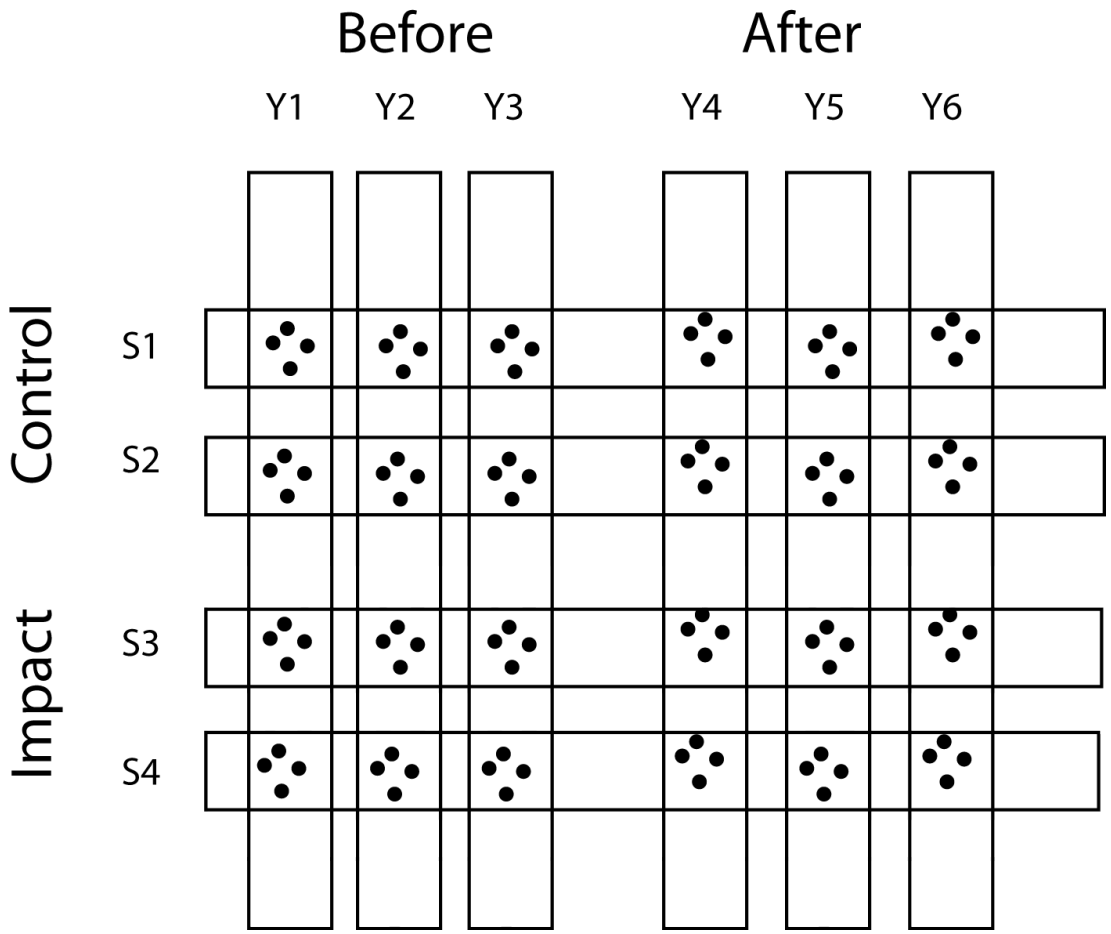
The power programs assume balanced data and no missing data, but can be modified to account for planned imbalance and missingness if needed. More details on the power computations are available later in this chapter.

12.12 A statistical diversion

The analysis of BACI experiments can be complex because of the various features of the design. A complete discussion is available in

> Underwood, A. J. (1992).
> On beyond BACI: the detection of environmental impacts on populations in the real, but variable, world.
> Journal of Experimental Marine Biology, 161, 145-178.

There are numerous variants of BACI designs depending on the number of control/impact sites, the number of years sampling before or after the disturbance, and the number of samples taken in each site-year combination. A schematic of a general design is:

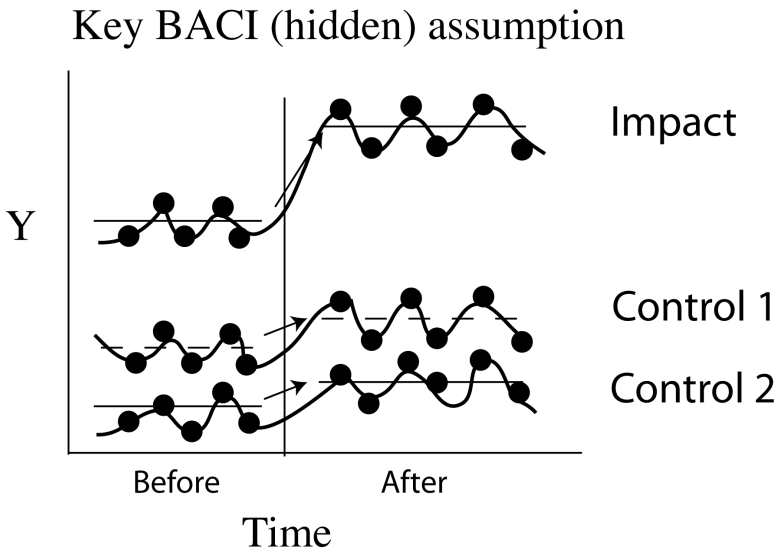


In this design, two sites in each of the control and impact areas will be sampled over time.[9]. Each site will be measured for three years before the disturbance occurs at the impacted sites, and three years after the disturbance occurs.[10] In each site-year combination, four measurements are taken (e.g. multiple quadrats; multiple fish density measurements, etc.) and these are assumed to be taken independently in each site-year combination.

Notice that the sites all have unique labels, rather than using the old-fashioned notation of site_1 (control) and site_1 (impact) where the numbering of the sites is no longer unique. Similarly, the years are also given unique labels for the same reason.

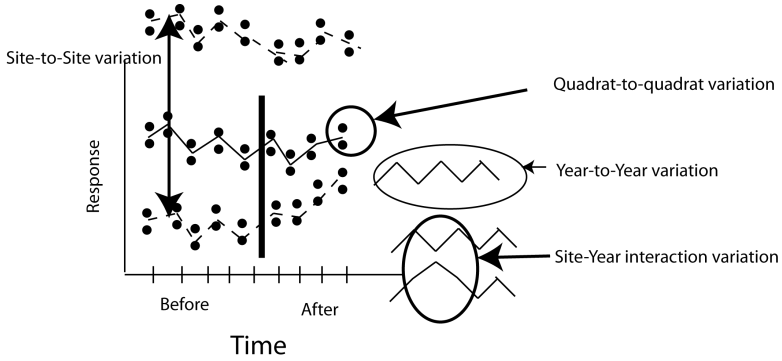
We are assuming that the response is a change in the mean that remains in effect after the environmental impact:

[9] It is not necessary that the number of sites be equal in the control and impacted areas.
[10] The number of years measured before and after disturbance also does not have to be equal.



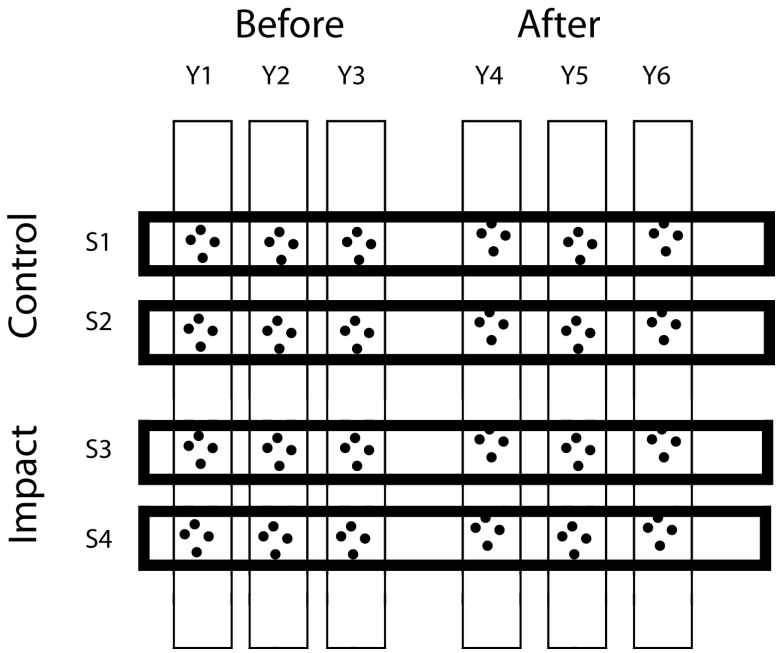
Here the effect of the disturbance is a step change that changes the mean response in the impacted site between the before and after periods. The control sites are allowed to also have a shift in the mean due to unexplained temporal effects (e.g. climate change), but the shift in the mean for impacted site is different than the shift in the mean for the control sites. This differential change is the BACI effect which is the interaction between temporal change and site class in the statistical model. Both Underwood (1992) and Underwood (1994) considered situations where the response is not a step change, but might be, for example, a pulse change of one year immediately after impact. These are not explored in this section of the notes. Notice that all sites are measured in every year before and after the disturbance.

There are several sources of variation:



which make the analysis more difficult than simple experiments.

This general BACI design actually combines several designs with different sizes of experimental units. First consider the difference between the mean at the control and impacted sites. This could be due to a number of factors. If the BACI design looks at the impact of a power plant on the productivity of a stream, the control sites could be on a different streams where the productivity is naturally different. We assume that the sites actually chosen are a random sample from all possible sites in the impacted or control areas – the discussion where we treat sites a fixed effects so that inference is restricted only to these actual sites measured is beyond these notes. This is an example of a strip-plot design, where the sites are "stripped" across the years:



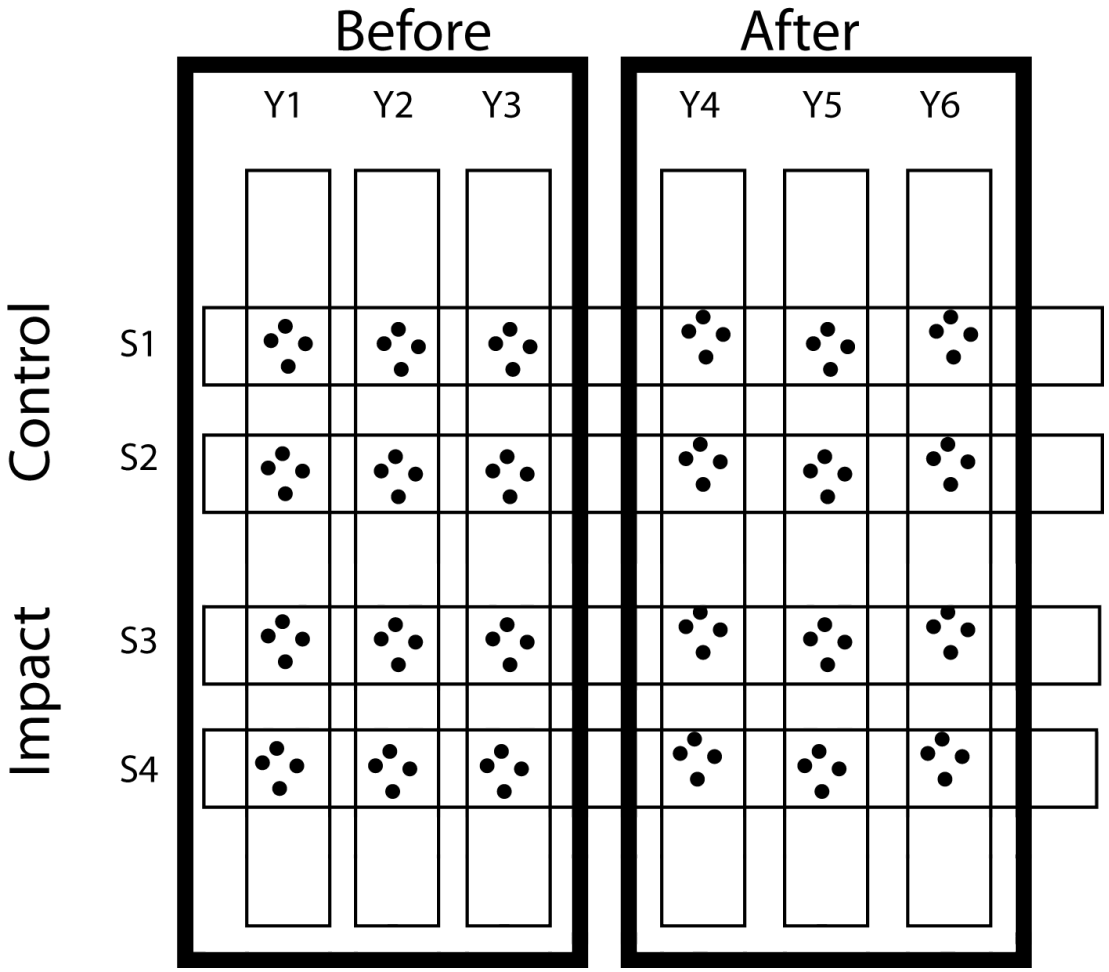
Some case is needed about choosing sites so that the same spatial variability is present among sites within the control and impacted areas. For example, a watershed may have two separate streams selected for the effect of an in-river hydroelectric system and two separate control streams. Presumably, the spatial variation among the streams in both site classes is now the same. However, a design may have a single stream where a power-plant will be built. Measuring two sites on this single stream and then using two separate streams for control will now mix spatial variations in the impact and control site classes. It would be better in this case to treat the two sites measured on the single impacted stream as pseudo-replicate measurements on that single stream. Underwood (1992, 1994) considered designs where the spatial scale varied among the control sites – please contact me for help with such designs.

The analysis of variance table for the effect of *SiteClass* (i.e. Control vs. Impact) is:

Source	df	F-ratio
SiteClass	1	MS(SiteClass)/MS(Site)
Site	2	

The df for *Sites* is found as the sum (of the number of sites in each site class -1) or in this case $2 = (2-1)+(2-1)$. We are not normally interested in difference in the mean between the impacted and control sites.

Consider the *Period* effects (i.e. before vs. after). Here is where we run into the first problem with BACI designs (and many related designs). The *Period* factor operates on a block of years and there is technically NO replicated experimental units for the period effect:



Technically, the years sampled within each *Period* are sub-samples (pseudo-replicates). Consequently, there is no formal statistical test available for the effect of *Period* unless you are willing to make some strong assumptions. For example, you could imagine long term cycles in the mean response that just happen to coincide with the stated before and after periods which would show up as a 'period" effect, but really are just natural variation over time. Fortunately, interest does not lie in the period effects, so the lack of a formal statistical test is not a concern. Regardless of the source of period effects, because every site is measured in every year, each year serves as a "block" and so impact vs. control comparisons are unaffected by the year/period effects.

This gives the following ANOVA table which is also a strip-plot design. Notice that the *Year* effects are subsamples within *Periods*. Its df are computed similarly as to those for sites, i.e. $4 = (3-1)+(3-1)$.

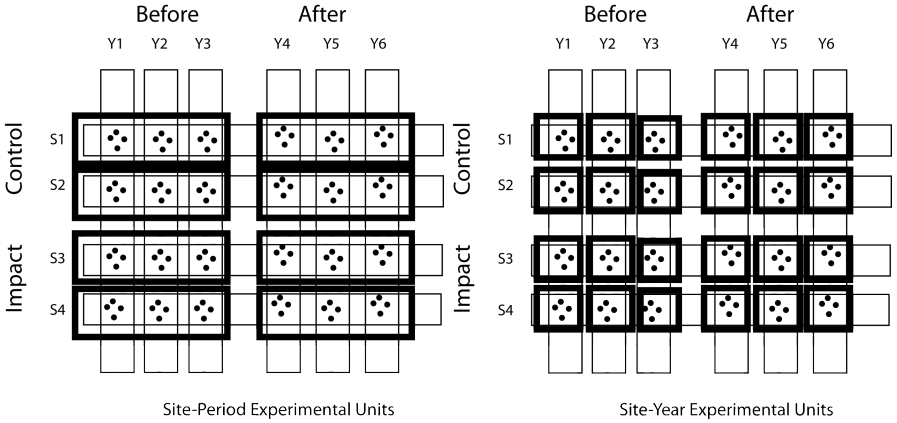
Source	df	F-ratio
Period	1	none
Period EU	0	
Years	4	

where *Period EU* in the ANOVA table means *experimental unit* for *Period* effects. Because there are no replicates of the *Period* experimental units, this has 0 df and so no test statistic can be computed.

These two design are applied orthogonally to each other, so the ANOVA table is extended by the "product" of the above tables. The discussion following the table explains how the F-ratios are created.

Source	*df*	F-ratio
BACI = SiteClass*Period	1	
SiteClass*Period EU	0	
SiteClass * Year	4	
Site*Period	2	
Site*Period EU	0	
Site*Years	8	

Whew... there are now two sizes of experimental units:



The *SiteClass*Year* and *Site*Year* terms in the ANOVA table both represent the same experimental unit and are typically pooled into a term representing the smallest experimental unit giving the reduced ANOVA table:

Source	*df*	F-ratio
BACI = SiteClass*Period	1	MS(BACI)/MS(Site*Period)
Site*Period	2	
Site*Period EU	0	
Site*Years + SiteClass*Year	14	

The BACI effect will be tested against the *Site*Period* experimental unit. Unfortunately, this F-ratio will only have few *df* (two *df* in this case) and likely has low power to detect effects. This test statistic "protects" against the possibility that the differences among the sites changes from the before to the after periods (i.e. that site effects are not consistent between the before and after periods). If you are will to ASSUME that differences among sites is consistent among periods, then it is also possible to pool this term with the final term in the ANOVA table. Dropping lines with 0 df gives the final (consolidated) ANOVA table:

Source	df	F-ratio
SiteClass	1	MS(SiteClass)/MS(Site)
Site	2	
Period	1	none
Years	4	
BACI = SiteClass*Period	1	MS(BACI)/MS(Site*Years + SiteClass*Year + Site*Period)
Site*Years + SiteClass*Year + Site*Period	14	

Underwood (1992) was also concerned about this type of pooling:

> Technically, this post-hoc removal of a component of variance involves making the assumption that the component is zero. This raises the possibility of Type II error (accepting that a value is zero when, in fact, the test is not capable of detecting it). This is the rationale behind post-hoc pooling procedures to reduce the probability of Type II errors (there are detailed discussions in Winer, 1971; Underwood, 1981). In this paper, the problem will be ignored on the grounds that Type II errors leading to inappropriate pooling will have the effect of increasing the probability of Type I errors in the tests for impacts. Thus, impacts may be detected when, in fact, they are not present. This seems a more desirable error than concluding that no impact is present when there is one – a likely risk in some of the tests described below which have few degrees of freedom and, probably, not very great power.

Finally, the multiple samples taken within each site-year combination are pseudo-replicates. The BACI analysis can be done on the average of these values. If the design is balanced, i.e. sample number of sub-samples in each site-year combination, then the analysis on the average values is exactly equivalent to the analysis on the individual values. The analysis on the individual values does provide additional information on the within-year-site variation and this would be useful in designing the experiment where the tradeoffs between more/less sites, more/less years, and more/less sampling within a site-year given the associated costs of new sites, new years, and new sub-samples can be explored.

The assumption of a step change between the before and after periods is crucial. In:

> Underwood, A. J. (1994).
> On beyond BACI: Sampling designs that might reliably detect environment disturbances.
> Ecological Applications, 4, 3-15.

is found a discussion for example, how should the data be analyzed if the expected response is pulse change that may occur only in the year after the impact as illustrated in the last two figures of Underwood (1994)'s Figure 2:

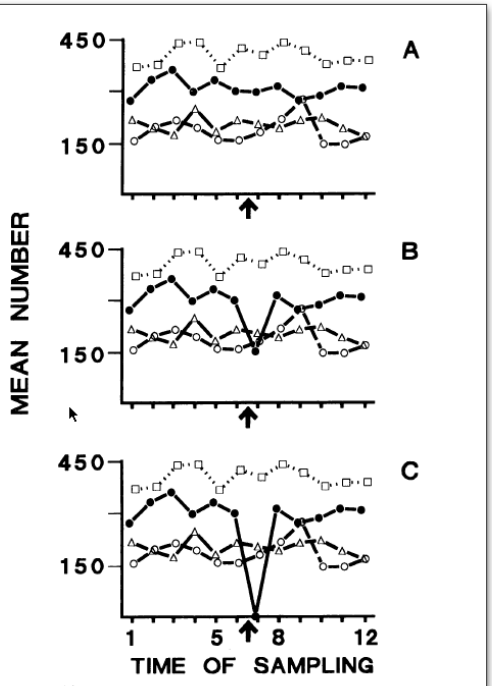


FIG. 2. Simulated environmental disturbances in one location (●——●), with three Controls, all sampled six times Before and After the disturbance (at the time indicated by the arrow). These data are analyzed in Table 3. (A) No effect of the disturbance; (B) a pulse reduction to 0.5 of the original mean; (C) a pulse reduction to 0.

12.13 Closing remarks about the analysis of BACI designs

As the examples above, the analysis of complex BACI designs takes some care. It is inadvisable to simply bash the numbers though a statistical package without a good understanding of exactly how the data were collected!

The key to a successful analysis of a complex experiment is the recognition of multiple levels (or sizes of experimental units) in the study. For example, with multiple control sites and multiple quadrats measured in each site-time combination, there are two sizes of experimental units – the site and quadrat level.

It is impossible to provide a "cookbook" to give examples of all the possible BACI analyses. Rather than relying on a "cookbook" approach, it is far better to use a "no-name" approach to model building.

The 'no-name' approach starts by adding terms to the model corresponding to the treatment effects:

$$Y = Treatment \;\; Time \;\; Treatment*time \;\; \ldots$$

where the term *Treatment* takes the values *impact* and *control*, the term *Time* represents the values *before* and *after*. These terms in the model represent the main effects (i.e. in the absence of an interaction these represents the site and temporal effects). The interaction term in the model again represent the potential environmental impact, i.e. is the change in the mean between before and after the same for the control and treatment sites?

To this model, terms must be included representing the various levels in the experiment, i.e. sites, years, etc. These terms can be specified in one of two ways. For example, the site term can be added as;

$$M1 : Y = \textit{Treatment Time Treatment*Time Site(Treatment)\&R} \ldots$$

$$M2 : Y = \textit{Treatment Time Treatment*Time Site\&R} \ldots$$

Both models are equivalent. The additional term $Site(Treatment)$ is the 'traditional' specification for the multiple sites within each treatment when the sites are NOT given unique labels, while the term $Site$ can be used (with modern software) when the site labels are unique.

The $\&R$ notation indicates that you need to specify these as *random* effects, i.e. you wish to generalize the results to more than the specific sites used in this experiment.

Similarly, if you have multiple years, add a term for the multiple years in a similar fashion.

Finally, if you have subsampling, you need to specify a term that indicates to *JMP* how the subsamples should be grouped. This is often a combination of year-and-site and so the *Year*Site&R* effect was added to the model.

Some textbooks recommend additional interactions terms be added to the model (e.g. the site*treatment interaction term). I find that unless you have a very large experiment, the data is often too sparse (especially if not all sites are measured in all years) to add these extra interaction terms. I haven't seen a real live experiment where they were useful. Refer back to the previous section on the statistical aspects of BACI designs.

After fitting the model, be sure to assess the fit of the model by looking at residual plots and the like.

12.14 BACI designs power analysis and sample size determination

12.14.1 Introduction

The basic concepts of power analysis are explained elsewhere in these course notes (e.g. in the first chapter on ANOVA). Briefly, power is the ability to detect a difference when that difference exists. The power of a design depends on several features:

- The α level. This is often set to 0.05 or 0.10. This represents the strength of evidence required before an effect is declared statistically significant.
- The variance components. You will need information on the variance components of all parts of your design. Depending upon the exact BACI design, there can be up to four variance components for which estimates will be needed: the site-to-site variance, the year-to-year variation, the site-year interaction variance, and the within site-year (typically quadrates or measurements) variance component. A preliminary study, or literature review can be very helpful in getting these components. As you will see later, some of the variance terms have no impact on the power analysis. For example, measuring the same sites or time, or pairing measurements across sites by year will allow you to "block" and remove these sources of variation from the analysis. This is the key advantage of "re-using" the sites in each year or taking paired measurements over time.

- The sample size. Here there are different levels to the samples. First is the number of quadrats measured in each site-year combination. This could vary, but for most power analyzes, it is commonly assumed that this is constant. Second, is the number of sites in the treatment or control areas. In many cases, there is only a single treatment site but the theory is completely general below to allow for multiple sites in the impacted areas. Lastly, how many years are measured before and after the impact?
- The effect size of interest. This is the most difficult part of the power analysis. You need to figure out how big of a BACI effect is important to detect. This seems rather obtuse, but can be obtained by drawing pictures of possible BACI outcomes (using hypothetical values) until people are sure that such an effect needs to be detected. This will be illustrated below.

What is a reasonable target for power? There are two "rules of thumb" commonly used. First, at $\alpha = 0.05$ a target power of $0.80 = 80\%$ is required; second at $\alpha = 0.10$ a target power of $0.90 = 90\%$ is required. These guidelines are somewhat arbitrary, but are widely used.

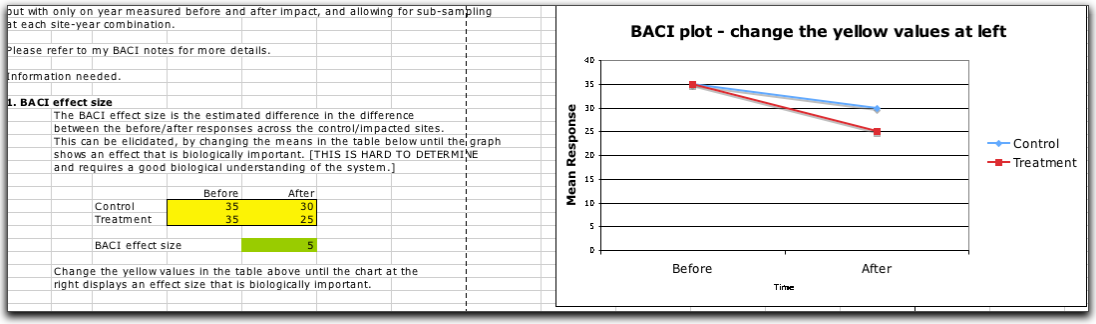
For many of the designs in this chapter, I have provided an Excel spreadsheet and *JMP*, *R*, or *SAS* code to help you determine power and sample size for various types of designs.

Let us consider the elements in more detail for general BACI designs, before considering power and sample size for a specific design.

Effect size of interest The effect size used in the power analysis is the BACI contrast. For many people this is difficult to ascertain. In the spreadsheet is an interactive graph that show the BACI plot for choices of the means before/after and control/treatment. Modify the values in the table until the graph shows a biologically significant difference that is important to detect.

The spreadsheet *baci-power-prep* has a section in the worksheet that can be used to determine the size of the BACI contrast that is of biological importance.

For example, in the crab example, we started with both sites having mean densities of about 35 crabs/quadrat. We suspect that both population may decline over over time, but if the difference in decline is more than about 5 crabs, we will be highly interested in detecting this. We varies the means in the table until the corresponding graph showed an effect that appears to be biologically important:



The BACI contrast value is obtained then in the usual way as the "difference in the differences":

$$BACI = (\mu_{CA} - \mu_{CB}) - (\mu_{TA} - \mu_{TB}) = \mu_{CA} - \mu_{CB} - \mu_{TA} + \mu_{TB} = 30 - 35 - 25 + 35 = 5$$

If a negative value is obtained for the BACI contrast, use the absolute value.

Variance components

- The site-to-site variance component measures how much different sites may vary among the mean because of local, site-specific effects (e.g. one site may have slightly better habit than a second site). Fortunately, a good estimates of this variance component is not crucial as it "cancels" because the same plot is repeated measured over time. A very rough rule of thumb is to take the range in means over sites and use the standard deviation as range/4.
- The year-to-year variance components measures how year-specific factors (e.g. a year may be wetter than normal which tends to increase productivity at all sites) affect the response. Again, good estimates of this variance component are not that crucial because if all sites are measured in all years (i.e. a blocked design), the year-effect is accounted for in the analysis and again "cancels" in the same way that blocking effects are removed.
- The site-year interaction variance component measures how inconsistent the response at sites are over time. Hopefully, it is small. Some "typical" values found in practice is that the site-time variance is about 20-50% of the site standard deviation. If is smaller than this, you are doing well! This typically is the limiting factor in the BACI study as nothing in the experiment can be modified to account for this variance component.
- The sub-sampling (or quadrat-to-quadrat or trap-to-trap) variance component asks how variable are the repeated measurements taken in the same site-year combination. A very rough rule of thumb is that for count data, the standard deviation is about the $\sqrt{average\ counts}$ under the assumption a Poisson distribution. It can be and often is higher - known as over-dispersion.

Estimates of these variance components can be obtained from a preliminary study, but notice that at least two years of preliminary data are required to estimate the year-to-year and year-site variance components.

Sample sizes There are two levels to the sample size specification.

At the lowest level is the number of quadrats (sub-samples) measured at each site-time combination. This is usually kept fixed for all sites and for all times and for treatment and control sites, but is allowed to vary. However, all sites that are measured in the treatment-time combination are assumed to have the same number of quadrats. If you wish different sites to have different number of sub-samples, please contact me.

At the upper level, you need to specify the number of sites measured in the treatment and control areas. In many cases, a single site is measured in the treatment area, but there are cases where multiple treatment sites can be measured. It is assumed that you will have at least 2 control sites measured for this design.

Similarly, at the upper level, you need to specify the number of years (before and after impact) for which monitoring will take place.

α level

Specify the α level (usually 0.05 or 0.10). In many cases, the choice of power and α level are given externally. For example, the Canadian Department of Environment often specifies that a power of 0.90 is the target at $\alpha = 0.10$.

Next we will examine how to determine the power/sample size for many of the BACI designs considered in this chapter.

12.14.2 Power: Before-After design

In Before-After studies, there are three attributes that are under the control of the experimenter:

- The number of years of monitoring before and after the impact;
- The number of locations monitored (the number of streams in the earlier example);
- The number of measurements taken at each year-location combination (the number of quadrats measured in the earlier example).

Each of these attributes attempt to control a particular source of variation. In the Before-After study, there are four source of variation:

- Year-to-Year variation. This is caused by year-specific factors such as increased rainfall in a year. A year-specific effect would impact all of the locations in the study.
- Location-to-Location variation (Stream-to-Stream variation in the earlier example). Not all locations have exactly the same mean response in a particular year. For example, some locations may be better habitat and generally have higher densities of organisms. We try and control for the the location-specific effects by measuring the same location in all years of the study. In this way, location is like a blocking factor, and so the effects of location can be "removed" from the study.
- Location-Year interaction variation (Stream-Year interaction in the earlier example). Even though there is a systematic difference among the streams, they may not all respond the same to year-specific factors. For example, locations with better habitat may be better able to weather a downturn in rainfall and so the change in density of organisms may not be as great as in locations with poorer habitat. A large Location-Year interaction variance would indicate that the results of the before-after study are not very reproducible in different locations.
- Residual variation (Quadrat-to-quadrat variation in the earlier example). Not all sampling sites at a location in a year have identical responses. For example, if the response are smallish counts, then the number of organisms in different sample units tends to follow a Poisson distribution.

In some cases, some of the variance components are confounded. For example, if only a single location (stream) is measured, then it is impossible to disentangle the year-to-year variation from the stream-year interaction variation and the combined variation is simply labelled as "year-to-year" variation.

In the following sections, we will show how to compute the power for both single and multiple location before-after studies.

Single Location studies

In these studies, only a single location is measured for multiple years before and after the impact, but at several sampling sites (quadrats) in each year. The full statistical model is:

$$Y_{ijk} = \mu_i + \epsilon_{ij}^{year} + \epsilon_{ijk}^{quadrat}$$

where y_{ijk} is the measured response in period i (either *before* or *after* the impact), in year j (of that period) and in quadrat k; μ_i is the mean response in period i; ϵ_{ij}^{year} is the year-specific (random) effect with variance σ^2_{year}; and finally $\epsilon_{ijk}^{quadrat}$ is the residual random error with variance σ^2.

For simplicity, assume that the same number of quadrats (n_q) is measured in each year. Then, as seen in the analysis of subsampling designs, the analysis of the averages is identical to that of the original data. The statistical model for the average response is then:

$$\overline{Y}_{ij\cdot} = \mu_i + \epsilon_{ij}^{year} + \overline{\epsilon}_{ij\cdot}^{quadrat}$$

where now $\overline{\epsilon}_{ij\cdot}^{quadrat}$ is the residual random error with variance σ^2/k. The variance of the yearly average is

$$V[\overline{Y}_{ij\cdot}] = \sigma^2_{year} + \frac{\sigma^2}{n_q}$$

We are now in the situation of a simple two-sample t-test whose power depends on the above variance and the number of years before and after the impact. Standard power programs for a two-sample t-test can be used. From the previous example, we found that the estimates of the combined year-to-year and stream-year interaction variance (labelled as the year variance) was 33.79 and the quadrat-to-quadrat variation (the residual variation) was 59.17. We had 5 years of before studies and want to know the power for various number of years after impact, assuming that a difference of 10 in the mean density was important to detect. We will assume that we can measure 10 quadrats each year.

Then the variance of the yearly average is

$$V[\overline{Y}_{ij\cdot}] = \sigma^2_{year} + \frac{\sigma^2}{n_q} = 33.79 + \frac{59.17}{10} = 39.707$$

and the standard deviation of the year average is $\sqrt{39.707} = 6.3$.

A sample program called *invert-power.r* is available in the Sample Program Library at `http://www.stat.sfu.ca/~cschwarz/Stat-650/Notes/MyPrograms`. The power is then computed using the *pwr.t2n.test()* function in the*pwr* package as:

```
# Note that d=effect size = difference to detect / standard deviation
power <- ldply(seq(5,20,1), function(n2){
  power1 <- pwr.t2n.test(n1=5, n2=n2, d=10/stddev,
                         alternative="two.sided")$power
  power2 <-before.after.power.stroup(Shift=10,
                           X.before=rep(1:5, each=10), X.after=5+rep(1:n2, each=10),
                           Process.SD=sqrt(33.79), Sampling.SD=sqrt(59.17) )

  res <- c(n2=n2, power.pwr=power1, power.stroup=power2["power"])
  res
})
cat("\n\nPower for detecting change in 10 in one stream with 5 years before and \n")
cat(    "different number of years after; 10 quadrats/year\n\n")
power
```

which gives the following power computations:

```
Power for detecting change in 10 in one stream with 5 years before and
different number of years after; 10 quadrats/year

   n2 power.pwr power.stroup.power
1   5 0.5961744          0.5961744
2   6 0.6464987          0.6464987
3   7 0.6860027          0.6860027
4   8 0.7174737          0.7174737
5   9 0.7429184          0.7429184
6  10 0.7637814          0.7637814
7  11 0.7811114          0.7811114
8  12 0.7956783          0.7956783
9  13 0.8080553          0.8080553
10 14 0.8186748          0.8186748
11 15 0.8278672          0.8278672
12 16 0.8358884          0.8358884
13 17 0.8429388          0.8429388
14 18 0.8491770          0.8491770
15 19 0.8547303          0.8547303
16 20 0.8597012          0.8597012
```

I have also written a specialized function that does the power computations once you specify the process and sampling standard deviations and the effect size to be detected. It also allows for multiple samples in each year in both balanced and unbalanced designs. Please consult the *R* code for more details.

It turn out that you would need at least 13 years measured after impact to have an 80% power to detect a difference of 10(!).

Note that the limiting factor for the power of this design is the combined year-to-year and year-stream interaction variance. Even if a very large number of quadrats (i.e. n_q) was measured each year, the repeated quadrats only drive down the final variance contribution and have no effect on the influence of year-specific and year-stream interaction effects.

Multiple Location studies

To improve the power, the same multiple locations can be measured in all year. As you will see in a minute, the multiple locations serve as blocks and so the location effects are "removed" from the analysis. Furthermore, the multiple locations (streams) influence the location-year (stream-year) interaction contribution. But the year-to-year variation is still limiting unless you move to a full BACI design where years now serve a blocks and the year-effect can be "removed".

The full statistical model is:

$$Y_{ijkl} = \mu_i + \epsilon_{ij}^{year} + \epsilon_{k}^{stream} + \epsilon_{ijk}^{stream-year} + \epsilon_{ijkl}^{quadrat}$$

where y_{ijkl} is the measured response in period i (either *before* or *after* the impact), in year j (of that period), stream k, and in quadrat l; μ_i is the mean response in period i; ϵ_{ij}^{year} is the year-specific (random) effect with variance σ^2_{year}; ϵ_k^{stream} is the stream-specific effect with variance σ^2_{stream}; $\epsilon_{ijk}^{stream-year}$ is the stream-year interaction effect with variance $\sigma^2_{stream-year}$; and finally $\epsilon_{ijkl}^{quadrat}$ is the residual random error with variance σ^2.

From the previous analysis of the raw data, the estimated variance components are:

- $\sigma^2_{year} = 36.64$;
- $\sigma^2_{stream} = 38.312$;
- $\sigma^2_{year-stream-interaction} = 0$;
- $\sigma^2_{quadrats} = 66.35$;

The method of Stroup can be used to estimate the power to detect a specified difference using *SAS* for any general design with varying number of locations measured each year, changing number of quadrats measured at each year-location combination, etc. The code is too complicated to present here – check with the *invert-power.sas* program in the Sample Program Library. For the special case of 4 streams and 5 quadrats measured per stream per year, the resulting output is:

Obs	n years before	n years after	n streams	n quadrats	Year var	Stream var	Stream year var	Quad var	Effect	Power
1	5	5	4	5	36.64	38.312	0	66.35	period	0.59
2	5	6	4	5	36.64	38.312	0	66.35	period	0.64
3	5	7	4	5	36.64	38.312	0	66.35	period	0.68
4	5	8	4	5	36.64	38.312	0	66.35	period	0.71
5	5	9	4	5	36.64	38.312	0	66.35	period	0.74
6	5	10	4	5	36.64	38.312	0	66.35	period	0.76
7	5	11	4	5	36.64	38.312	0	66.35	period	0.78
8	5	12	4	5	36.64	38.312	0	66.35	period	0.79
9	5	13	4	5	36.64	38.312	0	66.35	period	0.81
10	5	14	4	5	36.64	38.312	0	66.35	period	0.82
11	5	15	4	5	36.64	38.312	0	66.35	period	0.83
12	5	16	4	5	36.64	38.312	0	66.35	period	0.83
13	5	17	4	5	36.64	38.312	0	66.35	period	0.84
14	5	18	4	5	36.64	38.312	0	66.35	period	0.85
15	5	19	4	5	36.64	38.312	0	66.35	period	0.85
16	5	20	4	5	36.64	38.312	0	66.35	period	0.86

There is no simple way to use the Stroup method in *R* or *JMP*. However, in the special case when all locations are measured in all years and there are n_s locations (streams) and the same number of quadrats (n_q) is measured in each year in each location, then rather simple computations can be done and the power-analysis from a two-sample t-test can be used.

Because all locations are measured in all years, they play the same role as blocks and so the location (stream) effect is "removed" from the analysis. The location (stream) variance component play no part in the power determination. In order to use the power functions from a two-sample t-test, compute an appropriate variance for the two-sample t-test power computation as:

$$s^2 = \sigma^2_{year} + \frac{\sigma^2_{location-year}}{n_s} + \frac{\sigma^2_{quadrat}}{n_s \times n_q}$$

Using the estimated variance components above gives

$$s^2 = 33.79 + \frac{0}{4} + \frac{66.35}{4 \times 5} = 6.32$$

This can now be used in the standard power programs. A sample program called *invert-power.r* is available in the Sample Program Library at `http://www.stat.sfu.ca/~cschwarz/Stat-650/Notes/MyPrograms`. The power is then computed using the *pwr.t2n.test()* function in the*pwr* package as:

```
year_var        <- 36.64
stream_var      <- 38.3  # not needed
year_stream_var <- 0
quad_var        <- 66.35

# Note that d=effect size = difference to detect / standard deviation
power                <- data.frame(n_after=5:20)
power$n_before        <- 5
power$year_var        <- year_var
power$year_stream_var <- year_stream_var
power$quad_var        <- quad_var
power$n_stream        <- 4
power$n_quad          <- 5
power$alpha           <- .05
power$diff            <- 10
powerres <- adply(power,1,function(x){
  # compute power for one row of the table
  stddev <- sqrt(x$year_var + x$year_stream_var/x$n_stream +
                 x$quad_var/x$n_stream/x$n_quad)
  power <- pwr.t2n.test(n1=x$n_before,
                        n2=x$n_after,
                        d =x$diff/stddev,
                       alternative="two.sided")$power
  names(power) <- "power"
  return(power)
})

cat("\n\nPower for detecting change in 10 in 4 streams with 5 years before and \n")
cat(    "different number of years after; 5 quadrats/year\n\n")
cat("\n\nNote that power is now for a larger scope of inference \n\n")
powerres
```

which gives the following power computations:

```
Power for detecting change in 10 in 4 streams with 5 years before and
different number of years after; 5 quadrats/year



Note that power is now for a larger scope of inference

   n_after n_before year_var year_stream_var quad_var n_stream n_quad alpha diff     p
1        5        5    36.64                0    66.35        4      5  0.05   10 0.59
2        6        5    36.64                0    66.35        4      5  0.05   10 0.64
3        7        5    36.64                0    66.35        4      5  0.05   10 0.68
4        8        5    36.64                0    66.35        4      5  0.05   10 0.71
5        9        5    36.64                0    66.35        4      5  0.05   10 0.74
6       10        5    36.64                0    66.35        4      5  0.05   10 0.76
7       11        5    36.64                0    66.35        4      5  0.05   10 0.77
8       12        5    36.64                0    66.35        4      5  0.05   10 0.79
9       13        5    36.64                0    66.35        4      5  0.05   10 0.80
10      14        5    36.64                0    66.35        4      5  0.05   10 0.81
11      15        5    36.64                0    66.35        4      5  0.05   10 0.82
12      16        5    36.64                0    66.35        4      5  0.05   10 0.83
13      17        5    36.64                0    66.35        4      5  0.05   10 0.84
14      18        5    36.64                0    66.35        4      5  0.05   10 0.84
15      19        5    36.64                0    66.35        4      5  0.05   10 0.85
16      20        5    36.64                0    66.35        4      5  0.05   10 0.85
```

Notice that the power from the simpler method matches that from the Stroup method in this simplified design.

12.14.3 Power: Simple BACI design - one site control/impact; one year before/after; independent samples

Let us now turn our mind to power/sample size determination for the simplest BACI design. As noted, the simple BACI design has one site measured at impact and control; one year measured before and after; and independent samples taken at each year-site combination. This design is equivalent to a simple two-factor CRD considered in previous chapters. Unlike in the previous chapters, we are now specifically interested in detecting the interaction – the BACI effect.

If you have a program to determine power/sample size directly for the interaction in a two-factor CRD design, it can be used given the means and standard deviations as noted below. In some cases, you may not have access to such a program, but the computations for power and sample size are relatively straight-forward.

The full statistical model for a general BACI design is:

$$Y_{ijkl} = \mu_{ik} + \epsilon_{ij}^{site} + \epsilon_{ky}^{year} + \epsilon_{ijky}^{site*year} + \epsilon_{ijkyl}^{quadrat}$$

where Y_{ijkyl} is the measured response in treatment i (either control or treatment), site j (in treatment i), at period k (either before or after), in year y (within period k), and finally in subsample l (within treatment-site-period-year combination $ijky$); μ_{ik} is the mean response at treatment i and period k (i.e.

the four means used to determine the BACI contrast being the means at treatment-before, treatment-after, control-before, and control-after); ϵ_{ij}^{site} is site-to-site variation with variance (the square of the standard deviation) of σ_{site}^2; ϵ_{ky}^{year} is the year-to-year variation within the same period that affects all sites simultaneously; $\epsilon_{ijky}^{site*year}$ is the site-year interaction with variance of $\sigma_{\textit{site-year}}^2$; and $\epsilon_{ijkl}^{quadrat}$ is the subsampling variation with variance of $\sigma_{quadrat}^2$.

In the simplest BACI experiment, there are only 1 site in each of the treatment or control SiteClasses (so the k subscript only takes the value of 1), and only year measured in each of the two period (Before/After) so the y subscript also only takes the value of 1.

The variance of an individual observation is (by statistical theory):

$$V(Y_{ijkyl}) = \sigma_{site}^2 + \sigma_{year}^2 + \sigma_{\textit{site-year}}^2 + \sigma_{quadrat}^2$$

We start by averaging over the n_{ijky} subsamples taken in each site. This gives the sets of means:

$$\overline{Y}_{ijky\cdot} = \mu_{ik} + \epsilon_{ij}^{site} + \epsilon_{ky}^{year} + \epsilon_{ijky}^{\textit{site*year}} + \overline{\epsilon}_{ijky\cdot}^{quadrat}$$

Notice that because the average is over the subsamples within each treatment-site-period-year combination, the only thing that varies is the quadrats, and so the average occurs only over the quadrat-to-quadrat error. The variance of these averages is:

$$V(\overline{Y}_{ijk\cdot}) = \sigma_{site}^2 + \sigma_{year}^2 + \sigma_{\textit{site-year}}^2 + \frac{\sigma_{quadrat}^2}{n_{ik}}$$

Only the final variance is affect by averaging.

Now for each site, we take the difference between the after and before (average) readings – recall that in the simple BACI design, there is only 1 year in each period, i.e. the y subscript is always 1. This gives:

$$\begin{aligned}
&\overline{Y}_{ijA1\cdot} - \overline{Y}_{ijB1\cdot} \\
&\quad = \mu_{iA} + \epsilon_{ij}^{site} + \epsilon_{A1}^{year} + \epsilon_{ijA1}^{site*year} + \overline{\epsilon}_{ijA1\cdot}^{quadrat} - (\mu_{iB} + \epsilon_{ij}^{site} + \epsilon_{B1}^{year} + \epsilon_{ijB1}^{site*time} + \overline{\epsilon}_{ijB1\cdot}^{quadrat}) \\
&\quad = \mu_{iA} - \mu_{iB} + \epsilon_{A1}^{year} + \epsilon_{ijA1}^{site*time} + \overline{\epsilon}_{ijA1\cdot}^{quadrat} - \epsilon_{B1}^{year} - \epsilon_{ijB1}^{site*time} - \overline{\epsilon}_{ijB1\cdot}^{quadrat}
\end{aligned}$$

Notice that the site-to-site variance terms "cancel out". This occurs because the same site is measured on both the before and after period and so accounts for the extra site variation much like in paired designs.

The variance of the site differences is:

$$\begin{aligned}
V(\overline{Y}_{ijA\cdot} - \overline{Y}_{ijB\cdot}) &= \sigma_{year}^2 + \sigma_{\textit{site-year}}^2 + \frac{\sigma_{quadrat}^2}{n_{iA}} + \sigma_{year}^2 + \sigma_{\textit{site-year}}^2 + \frac{\sigma_{quadrat}^2}{n_{iB}} \\
&= 2\sigma_{year}^2 + 2\sigma_{\textit{site-year}}^2 + \frac{\sigma_{quadrat}^2}{n_{iA}} + \frac{\sigma_{quadrat}^2}{n_{iB}}
\end{aligned}$$

Finally, the BACI contrast is found as the difference of the difference in the treatment and control sites:

$$\begin{aligned}
BACI =& \overline{Y}_{T1A1\cdot} - \overline{Y}_{T1B1\cdot} - (\overline{Y}_{C1A1\cdot} - \overline{Y}_{C1B1\cdot}) \\
=& \mu_{TA} - \mu_{TB} - (\mu_{CA} - \mu_{CB}) + \\
& \epsilon^{site}_{T1} + \epsilon^{year}_{A1} + \epsilon^{site*year}_{T1A1} + \overline{\epsilon}^{quadrat}_{T1A1\cdot} - \\
& (\epsilon^{site}_{T1} + \epsilon^{year}_{B1} + \epsilon^{site*time}_{T1B1} + \overline{\epsilon}^{quadrat}_{T1B1\cdot}) - \\
& (\epsilon^{site}_{C1} + \epsilon^{year}_{A1} + \epsilon^{site*year}_{C1A1} + \overline{\epsilon}^{quadrat}_{C1A1\cdot}) + \\
& (\epsilon^{site}_{C1} + \epsilon^{year}_{B1} + \epsilon^{site*time}_{C1B1} + \overline{\epsilon}^{quadrat}_{C1B1\cdot}) \\
=& \mu_{TA} - \mu_{TB} - \mu_{CA} + \mu_{CB} + \\
& \epsilon^{site*year}_{T1A1} - \epsilon^{site*time}_{T1B1} - \epsilon^{site*year}_{C1A1} + \epsilon^{site*time}_{C1B1} + \overline{\epsilon}^{quadrat}_{T1A1\cdot} - \overline{\epsilon}^{quadrat}_{T1B1\cdot} - \overline{\epsilon}^{quadrat}_{C1A1\cdot} + \overline{\epsilon}^{quadrat}_{C1B1\cdot}
\end{aligned}$$

Now notice that both the *year* and *site* effect both cancel out (as the must with any blocking variable) because both sites were both measured in the same year before and after impact. The *site-year* interaction variance component still remains and this is **problematic**! The problem is that the *site-year* random effects are completely confounded with the respective means in the simple BACI design. This is exactly the problem that occurs in pseudo-replication where it is impossible to distinguish treatment effects from random effects. Because you only measured one site in the impact and control site classes and one year in the before and after periods, you cannot tell if the observed interaction effect is due to the underlying BACI effect (the difference in the means) or due to random, unknowable effects specific to each site-year combination. For example, suppose that your control site tended to have lower than average counts in one year because of some hidden problem at that site at that year. Then your control site would be lower than normal for one year and you might detect an "interaction effect" for that experiment that is simple random chance.

Consequently, the best we can do in a simple BACI is define a pseudo-BACI effect:

$$\begin{aligned}
BACI^* = \mu^*_{TA} - \mu^*_{TB} - \mu^*_{CA} + \mu^*_{CB} + \\
+ \overline{\epsilon}^{quadrat}_{T1A1\cdot} - \overline{\epsilon}^{quadrat}_{T1B1\cdot} - \overline{\epsilon}^{quadrat}_{C1A1\cdot} + \overline{\epsilon}^{quadrat}_{C1B1\cdot}
\end{aligned}$$

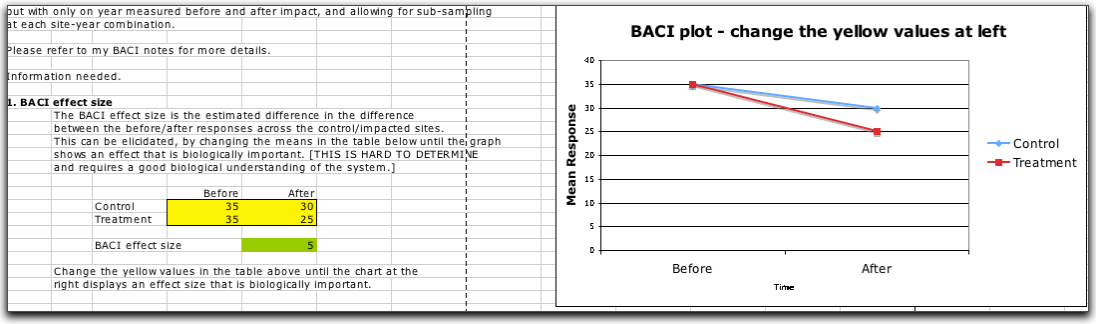
where the respective pseudo-means are specific to that choice of sites and years.

The final variance of the pseudo-BACI (where we must assume that the site-year variance is negligible and we are happy to restrict our findings to these particular years and particular sites) effect of the simple BACI design is found as:

$$V(BACI^*) = 4\frac{\sigma^2_{quadrat}}{n^{sites}_C} + \frac{\sigma^2_{quadrat}}{n^{sites}_C} + \frac{\sigma^2_{quadrat}}{n^{sites}_T} + \frac{\sigma^2_{quadrat}}{n^{sites}_T}$$

By taking multiple-subsamples, the contribution from the quadrat-to-quadrat variance can be reduced to a fairly small value, but no amount of sampling within that site-year combination will resolve the confounding of the site-year interaction terms with the means.

For example, in the crab example, we started with both sites having mean densities of about 35 crabs/quadrat. We suspect that both population may decline over over time, but if the difference in decline is more than about 5 crabs, we will be highly interested in detecting this. We varies the means in the table until the corresponding graph showed an effect that appears to be biologically important:



The BACI contrast value is obtained then in the usual way as the "difference of the differences":

$$BACI = (\mu_{CA} - \mu_{CB}) - (\mu_{TA} - \mu_{TB}) = \mu_{CA} - \mu_{CB} - \mu_{TA} + \mu_{TB} = 30 - 35 - 25 + 35 = 5$$

If a negative value is obtained for the BACI contrast, use the absolute value in the power program.

We need information on the STANDARD DEVIATIONS for the quadrat-to-quadrat variation in the experiment. A very rough rule of thumb is that for count data, the standard deviation is about the $\sqrt{average\ counts}$ assuming that the counts follow a poisson distribution but it can be considerably higher. If you have access to a preliminary survey, the average standard deviation among the quadrats measured in each site-year can be used, or if you have a preliminary ANOVA, the Root Mean Square Error (RMSE) also measures the "average" standard deviation. Based on the results from the analysis presented earlier, the estimated standard deviation is approximately 3.32.

I've created a general power-analysis function for BACI designs called *baci-power.r* available in the Sample Program Library at `http://www.stat.sfu.ca/~cschwarz/Stat-650/Notes/MyPrograms`.

This general method is based on the paper:

> Stroup, W. W. (1999)
> Mixed model procedures to assess power, precision, and sample size in the design of experiments.
> Pages 15-24 in Proc. Biopharmaceutical Section. Am. Stat. Assoc., Baltimore, MD

which is also discussed in Chapter 12 of

> Littell, R.C., Milliken, G.A., Stroup, W.W., Wolfinger, R.D., and Schabenberger, O.
> *SAS* for Mixed Models. 2nd Edition

Basically, the program generated "fake" data that is used to estimate the power of the design.

In the case of the simple BACI design, the *site-year* random effect is confounded with the the four treatment means and so we simply specify that its value is 0 in the function.

Here is some sample code to do the power analysis:

```
results <- NULL
results <- rbind(results,
            baci.power(n_TA=5, n_TB=5, n_CA=5, n_CB=5,
```

```
            ns_T=1, ns_C=1,
            ny_B=1, ny_A=1,
            mu_TA=30, mu_TB=35, mu_CA=35, mu_CB=35,
            sdYear=0, sdSite=0, sdSiteYear=0, sdResid=3.32))

results <- rbind(results,
            baci.power(n_TA=10, n_TB=10, n_CA=10, n_CB=10,
            ns_T=1, ns_C=1,
            ny_B=1, ny_A=1,
            mu_TA=30, mu_TB=35, mu_CA=35, mu_CB=35,
            sdYear=0, sdSite=0, sdSiteYear=0, sdResid=3.32))

results <- rbind(results,
            baci.power(n_TA=15, n_TB=15, n_CA=15, n_CB=15,
            ns_T=1, ns_C=1,
            ny_B=1, ny_A=1,
            mu_TA=30, mu_TB=35, mu_CA=35, mu_CB=35,
            sdYear=0, sdSite=0, sdSiteYear=0, sdResid=3.32))

results <- rbind(results,
            baci.power(n_TA=20, n_TB=20, n_CA=20, n_CB=20,
            ns_T=1, ns_C=1,
            ny_B=1, ny_A=1,
            mu_TA=30, mu_TB=35, mu_CA=35, mu_CB=35,
            sdYear=0, sdSite=0, sdSiteYear=0, sdResid=3.32))
```

giving the following power computations

```
     alpha n_TA n_TB n_CA n_CB baci power
[1,] 0.05  5    5    5    5    -5   0.3533207
[2,] 0.05  10   10   10   10   -5   0.6396126
[3,] 0.05  15   15   15   15   -5   0.8175554
[4,] 0.05  20   20   20   20   -5   0.9138353
```

For this example, if a BACI contrast of 5 is biologically important, then about 15 sub-samples are required in each site-year combination. Varying the number of subsamples has some, but little effect on the overall power compared to changes in the total sample size.

12.14.4 Power: Multiple sites in control/impact; one year before/after; independent samples

As you saw in the previous section, the simple BACI design with one site measured in one year at impact and control has a fundamental problem of pseudo-replication. We now extend the design to having multiple sites (with still a single year of measurements in each period) in this section.

We start again with the full statistical model for a general BACI design is:

$$Y_{ijkyl} = \mu_{ik} + \epsilon_{ij}^{site} + \epsilon_{ky}^{year} + \epsilon_{ijky}^{site*year} + \epsilon_{ijkyl}^{quadrat}$$

where Y_{ijkyl} is the measured response in treatment i (either control or treatment), site j (in treatment i), at period k (either before or after), in year y (within period k), and finally in subsample l (within treatment-site-period-year combination $ijky$); μ_{ik} is the mean response at treatment i and period k (i.e. the four means used to determine the BACI contrast being the means at treatment-before, treatment-after, control-before, and control-after); ϵ_{ij}^{site} is site-to-site variation with variance (the square of the standard deviation) of σ_{site}^2; ϵ_{ky}^{year} is the year-to-year variation within the same period that affects all sites simultaneously; $\epsilon_{ijky}^{site*year}$ is the site-year interaction with variance of $\sigma_{\textit{site-year}}^2$; and $\epsilon_{ijkl}^{quadrat}$ is the subsampling variation with variance of $\sigma_{quadrat}^2$.

In this BACI experiment, there are now (possibly) multiple sites in either or both of the treatment or control SiteClasses, but still only one year measured in each of the two period (Before/After) so the y subscript also only takes the value of 1.

The variance of an individual observation is (by statistical theory):

$$V(Y_{ijkyl}) = \sigma_{site}^2 + \sigma_{year}^2 + \sigma_{\textit{site-year}}^2 + \sigma_{quadrat}^2$$

We start by averaging over the n_{ijky} subsamples taken in each site. This gives the sets of means:

$$\overline{Y}_{ijky\cdot} = \mu_{ik} + \epsilon_{ij}^{site} + \epsilon_{ky}^{year} + \epsilon_{ijky}^{\textit{site*year}} + \overline{\epsilon}_{ijky\cdot}^{quadrat}$$

Notice that because the average is over the subsamples within each treatment-site-period-year combination, the only thing that varies is the quadrats, and so the average occurs only over the quadrat-to-quadrat error. The variance of these averages is:

$$V(\overline{Y}_{ijk\cdot}) = \sigma_{site}^2 + \sigma_{year}^2 + \sigma_{\textit{site-year}}^2 + \frac{\sigma_{quadrat}^2}{n_{ik}}$$

Only the final variance is affect by averaging.

Now for each site, we take the difference between the after and before (average) readings – recall that in this design BACI design, there is still only 1 year in each period, i.e. the y subscript is always 1. This gives:

$$\begin{aligned}
&\overline{Y}_{ijA1\cdot} - \overline{Y}_{ijB1\cdot} \\
&\quad = \mu_{iA} + \epsilon_{ij}^{site} + \epsilon_{A1}^{year} + \epsilon_{ijA1}^{site*year} + \overline{\epsilon}_{ijA1\cdot}^{quadrat} - (\mu_{iB} + \epsilon_{ij}^{site} + \epsilon_{B1}^{year} + \epsilon_{ijB1}^{site*time} + \overline{\epsilon}_{ijB1\cdot}^{quadrat}) \\
&\quad = \mu_{iA} - \mu_{iB} + \epsilon_{A1}^{year} + \epsilon_{ijA1}^{site*time} + \overline{\epsilon}_{ijA1\cdot}^{quadrat} - \epsilon_{B1}^{year} - \epsilon_{ijB1}^{site*time} - \overline{\epsilon}_{ijB1\cdot}^{quadrat}
\end{aligned}$$

Notice that the site-to-site variance terms "cancel out". This occurs because the same site is measured on both the before and after period and so accounts for the extra site variation much like in paired designs.

The variance of the site differences is:

$$V(\overline{Y}_{ijA\cdot} - \overline{Y}_{ijB\cdot}) = \sigma^2_{year} + \sigma^2_{\textit{site-year}} + \frac{\sigma^2_{quadrat}}{n_{iA}} + \sigma^2_{year} + \sigma^2_{\textit{site-year}} + \frac{\sigma^2_{quadrat}}{n_{iB}}$$
$$= 2\sigma^2_{year} + 2\sigma^2_{\textit{site-year}} + \frac{\sigma^2_{quadrat}}{n_{iA}} + \frac{\sigma^2_{quadrat}}{n_{iB}}$$

We now average over the (possible multiple) sites in the control areas:

$$\overline{Y}_{C\cdot A1\cdot} - \overline{Y}_{C\cdot B1\cdot} = \mu_{CA} - \mu_{CB} + \epsilon^{year}_{A1} + \overline{\epsilon}^{\textit{site*year}}_{C\cdot A1} + \overline{\epsilon}^{quadrat}_{C\cdot A1\cdot} - \epsilon^{year}_{B1} - \overline{\epsilon}^{\textit{site*year}}_{C\cdot B1} - \overline{\epsilon}^{quadrat}_{C\cdot B1\cdot}$$

which has variance of:

$$V(\overline{Y}_{C\cdot A\cdot} - \overline{Y}_{C\cdot B\cdot}) = 2\sigma^2_{year} + 2\frac{\sigma^2_{\textit{site-time}}}{n^{sites}_C} + \frac{\sigma^2_{quadrat}}{n_{CA}n^{sites}_C} + \frac{\sigma^2_{quadrat}}{n_{CB}n^{sites}_C}$$

A similar expression can be obtained for the average of (possibly) replicated treatment sites.

Finally, the BACI contrast is found as the difference of the difference in the treatment and control sites:

$$\begin{aligned}
BACI =& \overline{Y}_{T\cdot A1\cdot} - \overline{Y}_{T\cdot B1\cdot} - (\overline{Y}_{C\cdot A1\cdot} - \overline{Y}_{C\cdot B1\cdot}) \\
=& \mu_{TA} - \mu_{TB} - (\mu_{CA} - \mu_{CB}) + \\
& \overline{\epsilon}^{site}_{T\cdot} + \epsilon^{year}_{A1} + \overline{\epsilon}^{site*year}_{T\cdot A1} + \overline{\epsilon}^{quadrat}_{T1A1\cdot} - \\
& (\overline{\epsilon}^{site}_{T\cdot} + \epsilon^{year}_{B1} + \overline{\epsilon}^{site*year}_{T\cdot B1} + \overline{\epsilon}^{quadrat}_{T1B1\cdot}) - \\
& (\overline{\epsilon}^{site}_{C\cdot} + \epsilon^{year}_{A1} + \overline{\epsilon}^{site*year}_{C\cdot A1} + \overline{\epsilon}^{quadrat}_{C1A1\cdot}) + \\
& (\overline{\epsilon}^{site}_{C\cdot} + \epsilon^{year}_{B1} + \overline{\epsilon}^{site*year}_{C\cdot B1} + \overline{\epsilon}^{quadrat}_{C1B1\cdot}) \\
=& \mu_{TA} - \mu_{TB} - \mu_{CA} + \mu_{CB} + \\
& \overline{\epsilon}^{site*year}_{T\cdot A1} - \overline{\epsilon}^{site*time}_{T\cdot B1} - \overline{\epsilon}^{site*year}_{C\cdot A1} + \overline{\epsilon}^{site*time}_{C\cdot B1} + \overline{\epsilon}^{quadrat}_{T1A1\cdot} - \overline{\epsilon}^{quadrat}_{T1B1\cdot} - \overline{\epsilon}^{quadrat}_{C1A1\cdot} + \overline{\epsilon}^{quadrat}_{C1B1\cdot}
\end{aligned}$$

Now notice that both the *year* and *site* effect both cancel out (as the must with any blocking variable) because ALL sites were both measured in the same year before and after impact. The *site-year* interaction variance component still remains but now is no longer problematic as it is no longer confounded with the means. This is primary reason why multiple sites (at least multiple control sites) should be used with any BACI experiment.

The variance of the BACI contrast is found to be:

$$V(BACI) = 2\frac{\sigma^2_{\textit{site-year}}}{n^{sites}_C} + 2\frac{\sigma^2_{\textit{site-year}}}{n^{sites}_T} + \frac{\sigma^2_{quadrat}}{n_{CA}n^{sites}_C} + \frac{\sigma^2_{quadrat}}{n_{CB}n^{sites}_C} + \frac{\sigma^2_{quadrat}}{n_{TA}n^{sites}_T} + \frac{\sigma^2_{quadrat}}{n_{TB}n^{sites}_T}$$

The standard deviation of the BACI contrast is the square root of this value. Finally, the BACI value and the standard deviation of the BACI contrast can be used with any one-sample power analysis program to determine the power of the design.

The previous analysis of this data gave the following estimates of the variance components;

3.831	Site-to-site STANDARD DEVIATION
1.296	Site-time interaction STANDARD DEVIATION
3.305	Subsampling STANDARD DEVIATION

The BACI effect size is determined in the usual way. Again assume that a BACI contrast of 5 is of biological interest.

The baseline number of subsamples and sites is

Number of sub-samples at each site in the treatment-time combination.

	Subsamples Taken at sites	
	Before	After
Control	5	5
Treatment	5	5

Number of sites measured in the control or treatment areas. It is assumed that every site will be measured once before impact and once after impact has occurred with sub-samples as determined abov

	Number Sites
Control	2
Treatment	1

As in the previous section, we used the *baci-power.r* routine. Here is some sample code to do the power analysis based on the baseline values and several other scenarios:

```
results <- NULL
results <- rbind(results,
            baci.power(n_TA=5, n_TB=5, n_CA=5, n_CB=5,
            ns_T=1, ns_C=2,
            ny_B=1, ny_A=1,
            mu_TA=30, mu_TB=35, mu_CA=35, mu_CB=35,
            sdYear=0, sdSite=3.831, sdSiteYear=1.296, sdResid=3.30))

# More scenarios not listed
```

giving the following power computations

```
     alpha ns_T ns_C n_TA n_TB n_CA n_CB baci power
```

```
[1,] 0.05  1    2    5    5    5    5    -5   0.09574448
[2,] 0.05  1    2    40   40   40   40   -5   0.1296767
[3,] 0.05  1    4    5    5    5    5    -5   0.2074871
[4,] 0.05  1    4    20   20   20   20   -5   0.3162408
[5,] 0.05  1    4    20   20   5    5    -5   0.2842776
[6,] 0.05  1    10   5    5    5    5    -5   0.3356491
[7,] 0.05  2    10   5    5    5    5    -5   0.5551316
[8,] 0.05  2    10   40   40   40   40   -5   0.8368768
```

Notice that increasing the sub-sampling from 5 to 40 in each site-year combination has little effect on the power. The problem is that these are pseudo-replicates and not true replicates. Consequently, replication of the number of sites is much more effective. Now, because the number of treatment sites is limited (there is only 1), it may be sensible to increase sampling at the treatment site at the quadrat level as well. For this particular experiment, quite a few additional sites would be needed to have a reasonable chance of detecting this size of a biologically important difference.

12.14.5 Power: One sites in control/impact; multiple years before/after; no subsampling

The derivation for the power analysis for this design is relatively straightforward because there is no subsampling in each site.

We start again with the full statistical model for a general BACI design is:

$$Y_{ijkl} = \mu_{ik} + \epsilon_{ij}^{site} + \epsilon_{ky}^{year} + \epsilon_{ijky}^{site*year} + \epsilon_{ijkyl}^{quadrat}$$

where Y_{ijkyl} is the measured response in treatment i (either control or treatment), site j (in treatment i), at period k (either before or after), in year y (within period k), and finally in subsample l (within treatment-site-period-year combination $ijky$); μ_{ik} is the mean response at treatment i and period k (i.e. the four means used to determine the BACI contrast being the means at treatment-before, treatment-after, control-before, and control-after); ϵ_{ij}^{site} is site-to-site variation with variance (the square of the standard deviation) of σ^2_{site}; ϵ_{ky}^{year} is the year-to-year variation within the same period that affects all sites simultaneously; $\epsilon_{ijky}^{site*year}$ is the site-year interaction with variance of $\sigma^2_{\textit{site-year}}$; and $\epsilon_{ijkl}^{quadrat}$ is the subsampling variation with variance of $\sigma^2_{quadrat}$.

In this BACI experiment, there are now only one site in each of the treatment or control SiteClasses so the k subscript only takes the value of 1, but now multiple years measured in each of the two period (Before/After) so the y subscript no takes many values. There is only one measurement at each site-year combination, so the l subscript only takes the value 1 as well.

The variance of an individual observation is (by statistical theory):

$$V(Y_{ijkyl}) = \sigma^2_{site} + \sigma^2_{year} + \sigma^2_{\textit{site-year}} + \sigma^2_{quadrat}$$

Because both sites are measured every sampling event (a paired design), we first take the difference in response between the treatment and control sites at each time point:

$$\begin{aligned} d_{ky1} =& Y_{T1ky1} - Y_{C1ky1} = \mu_{Tk} + \epsilon^{site}_{T1} + \epsilon^{year}_{ky} + \epsilon^{site*year}_{T1ky} + \epsilon^{quadrat}_{T1kyl} - (\mu_{Ck} + \epsilon^{site}_{C1} + \epsilon^{year}_{ky} + \epsilon^{site*year}_{C1ky} + \epsilon^{quadrat}_{C1kyl}) \\ =& \mu_{Tk} + \epsilon^{site}_{T1} + \epsilon^{site*year}_{T1ky} + \epsilon^{quadrat}_{T1kyl} - (\mu_{Ck} + \epsilon^{site}_{C1} + \epsilon^{site*year}_{C1ky} + \epsilon^{quadrat}_{C1kyl}) \end{aligned}$$

Notice that the sampling-variation caused by different sampling times cancels – this is exactly what happens with paired designs.

The average differences in the before and after periods is then found:

$$\begin{aligned} \overline{d}_{B\cdot 1} =& \mu_{TB} + \epsilon^{site}_{T1} + \overline{\epsilon}^{site*year}_{T1B\cdot} + \overline{\epsilon}^{quadrat}_{T1B\cdot 1} - (\mu_{CB} + \epsilon^{site}_{C1} + \overline{\epsilon}^{site*year}_{C1B\cdot} + \overline{\epsilon}^{quadrat}_{C1B\cdot 1}) \\ \overline{d}_{A\cdot 1} =& \mu_{TA} + \epsilon^{site}_{T1} + \overline{\epsilon}^{site*year}_{T1A\cdot} + \overline{\epsilon}^{quadrat}_{T1A\cdot 1} - (\mu_{CA} + \epsilon^{site}_{C1} + \overline{\epsilon}^{site*year}_{C1A\cdot} + \overline{\epsilon}^{quadrat}_{C1A\cdot 1}) \end{aligned}$$

Finally the BACI contrast is the difference of these differences

$$\begin{aligned} BACI =& \overline{d}_{B\cdot 1} - \overline{d}_{A\cdot 1} = \mu_{TB} - \mu_{CB} - \mu_{TA} + \mu_{CA} + \\ & \overline{\epsilon}^{site*year}_{T1B\cdot} - \overline{\epsilon}^{site*year}_{C1B\cdot} - \overline{\epsilon}^{site*year}_{T1A\cdot} + \overline{\epsilon}^{site*year}_{C1A\cdot} + \\ & \overline{\epsilon}^{quadrat}_{T1B\cdot 1} - \overline{\epsilon}^{quadrat}_{C1B\cdot 1} - \overline{\epsilon}^{quadrat}_{T1A\cdot 1} + \overline{\epsilon}^{quadrat}_{C1A\cdot 1} \end{aligned}$$

Notice how the site-effects again cancel out because the same sites are measured in both periods (before and after).

The BACI contrast has variance:

$$V(BACI) = \frac{2\sigma^2_{\textit{site-year}}}{n_B} + \frac{2\sigma^2_{\textit{site-year}}}{n_A} + \frac{2\sigma^2}{n_B} + \frac{2\sigma^2}{n_A}$$

This can then be used with any power program for a single mean.

Note however, that with only one measurement per year, the *site-year* and residual (subsampling) variances cannot be disentangled and are completely confounded. Consequently, the estimated residual variation is a combination of the *site-year* and residual variation and is used as is. From the previous analysis of the fish data, we found that the year-to-year variance component (the *SamplingTime* variance component had a (standard deviation) value of 31.1; the residual variation (combination of the *site-year* and subsampling variation) had a (standard deviation) value of 27.07.

As in the previous section, we used the *baci-power.r* routine. Here is some sample code to do the power analysis based on the baseline values and several other scenarios. Note that we set the number of sub-samples to 1 and set the sub-sampling variance component to zero.

```
results <- NULL
# Note that because there is only 1 site, there is NO variance component
# associated with site. Use any value here. In general, the site variance
# component is not important because when you meaure the same sites over time
# the site effects cancel out (like blocking) and so site-to-site variability
# is not important. The site-year variation really is what drives the BACI power.
results <- rbind(results,
           baci.power(n_TA=1, n_TB=1, n_CA=1, n_CB=1,
```

```
            ns_T=1, ns_C=1,
            ny_B=12, ny_A=13,
            mu_TA=20, mu_TB=30, mu_CA=30, mu_CB=30,
            sdYear=31.71, sdSite=0, sdSiteYear=27.07, sdResid=0))
```

giving the following power computations

```
     alpha ny_B ny_A baci power
[1,] 0.05  12   13   -10  0.09591081
[2,] 0.05  24   26   -10  0.1476783
[3,] 0.05  12   13   40   0.7053772
[4,] 0.05  12   13   50   0.8777241
```

Note that the power is extremely small for all but very large differences even with monthly samples taken before and after impact. The key problem is the large variation in the differences among the different months. This may indicate that the design should be modified to incorporate sub-sampling to try and reduce some of the variation seen across months.

The power and sample size for the chironomid example is very similar and not repeated here.

12.14.6 Power: General BACI: Multiple sites in control/impact; multiple years before/after; subsampling

Whew! Now for a power/sample size analysis for the general BACI design with multiple sites in the SiteClasses (typically only the Control sites are replicated), multiple years in each period (before/after), and multiple subsamples.

The derivation of the BACI contrast and its variance follows the example shown in previous section and only the final results will be shown.

The BACI contrast is the difference of the differences after averaging over subsamples, multiple sites in each SiteClass, and multiple years in each period:

$$\begin{aligned} BACI =& \overline{d}_{B\cdot\cdot} - \overline{d}_{A\cdot\cdot} = \mu_{TB} - \mu_{CB} - \mu_{TA} + \mu_{CA} + \\ & \overline{\epsilon}^{site*year}_{T\cdot B\cdot} - \overline{\epsilon}^{site*year}_{C\cdot B\cdot} - \overline{\epsilon}^{site*year}_{T\cdot A\cdot} + \overline{\epsilon}^{site*year}_{C\cdot A\cdot} + \\ & \overline{\epsilon}^{quadrat}_{T\cdot B\cdot\cdot} - \overline{\epsilon}^{quadrat}_{C\cdot B\cdot\cdot} - \overline{\epsilon}^{quadrat}_{T\cdot A\cdot\cdot} + \overline{\epsilon}^{quadrat}_{C\cdot A\cdot\cdot} \end{aligned}$$

Notice how the site-effects and year-effects again cancel out because the same sites are measured in all years (before and after).

The general BACI contrast has variance:

$$V(BACI) = \frac{\sigma^2_{\text{site-year}}}{ny_B ns_T} + \frac{\sigma^2_{\text{site-year}}}{ny_B ns_C} + \frac{\sigma^2_{\text{site-year}}}{ny_A ns_T} + \frac{\sigma^2_{\text{site-year}}}{ny_A ns_C} + \frac{\sigma^2}{ny_B ns_T n_{TB}} + \frac{\sigma^2}{ny_B ns_C n_{CB}} + \frac{\sigma^2}{ny_A ns_T n_{TA}} + \frac{\sigma^2}{ny_A ns_C n_{CA}}$$

where ny_B is the number of years monitored before impact; ny_A is the number of years monitored after impact; ns_T is the number of sites measured in the Treatment SiteClass; ns_C is the number of sites measured in the Control SiteClass; n_{TB}, n_{TA}, n_{CB}, and n_{CA} is the number of subsamples taken in each site and year in the treatment-before combination etc. This can then be used with any power program for a single mean in the same fashion as before.

Consider the fry example. Here the variance components are:

- site-to-site standard deviation 0.75; This is irrelevant because all sites are measured in all years.
- year-to-year standard deviation 0.25; This is also irrelevant because all sites are measured in all years.
- site-year standard deviation 0.1;
- sub-sampling standard deviation 0.75.

A BACI contrast of 0.5 is of interest, so choose 4 means to give the relevant value. Finally, various combinations of the number of cages, number of sites, number of years are tried to obtain the following power.

As in the previous section, we used the *baci-power.r* routine. Here is some sample code to do the power analysis based on the baseline values and several other scenarios.

```
# Get the BACI power program
source("../baci-power.r")

# An illustration of how to find the power for one scenario
baci.power(n_TA=3, n_TB=3, n_CA=3, n_CB=3,
            ns_T=3, ns_C=3,
            ny_B=3, ny_A=2,
            mu_TA=5.0, mu_TB=5.5, mu_CA=4.5, mu_CB=4.5,
            sdYear=0.25, sdSite=0.75, sdSiteYear=0.1, sdResid=0.75)

#find the power under various scenarios
results <- NULL
results <- rbind(results,
            baci.power(n_TA=3, n_TB=3, n_CA=3, n_CB=3,
            ns_T=3, ns_C=3,
            ny_B=3, ny_A=2,
            mu_TA=5.0, mu_TB=5.5, mu_CA=4.5, mu_CB=4.5,
            sdYear=0.25, sdSite=0.75, sdSiteYear=0.1, sdResid=0.75))
```

giving the following power computations

```
     alpha n_TA n_TB n_CA n_CB ny_B ny_A ns_T ns_C baci power
[1,] 0.05  3    3    3    3    3    2    3    3    -0.5 0.2996488
[2,] 0.05  6    6    6    6    3    2    3    3    -0.5 0.5068123
[3,] 0.05  9    9    9    9    3    3    3    3    -0.5 0.7619485
[4,] 0.05  6    6    6    6    3    4    3    3    -0.5 0.6722755
```

You see that about 9 cages need to employed in each site-year combination to get reasonable power.