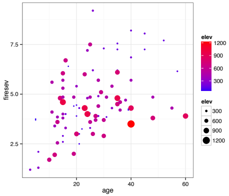
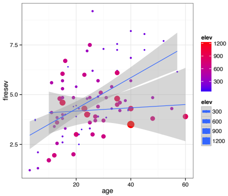
Interaction Effects with Continuous Variables

Problem: What if Continuous Predictors are Not Additive?



Problem: What if Continuous Predictors are Not Additive?



Five year study of wildfires & recovery in Southern California shurblands in 1993. 90 plots (20 x 50m)
(data from Jon Keeley et al.)

Exercise: Fire!

- Fit and evaluate a model that shows stand age and elevation interacting to impact fire severity
- Use ggplot2 to plot the data

ANOVA with an Interaction

```
# Anova Table (Type II tests)
#
# Response: firesev
#            Sum Sq Df F value   Pr(>F)
# age          53.0  1   27.71    1e-06
# elev          6.3  1    3.27 0.07399
# age:elev     22.3  1   11.67 0.00097
# Residuals   164.4 86
```

Type I, II, and III Sums of Squares

	Type I	Type II	Type III
Test for A	A v. 1	A + B v. B	A + B + A:B v. B + A:B
Test for B	A + B v. A	A + B v. A	A + B + A:B v. A + A:B
Test for A:B	A + B + A:B v. A + B	–	–

- What do type III models mean?
- Interactions the same for all, and if A:B is real, main effects not important
- Type III has lower power for main effects

Type II v. Type III Sums of Squares

```
Anova(keeley_lm)

# Anova Table (Type II tests)
#
# Response: firesev
#           Sum Sq Df F value   Pr(>F)
# age         53.0  1   27.71    1e-06
# elev         6.3  1    3.27 0.07399
# age:elev    22.3  1   11.67 0.00097
# Residuals  164.4 86

Anova(keeley_lm, type="III")

# Anova Table (Type III tests)
#
# Response: firesev
#             Sum Sq Df F value  Pr(>F)
# (Intercept)   16.6  1    8.68 0.00415
# age           63.9  1   33.43 1.2e-07
# elev          10.2  1    5.36 0.02302
# age:elev      22.3  1   11.67 0.00097
# Residuals    164.4 86
```

What does the Interaction Coefficient Mean?

```
#                Estimate Std. Error t value  Pr(>|t|)
# (Intercept)  1.8132153  0.6156070   2.945 4.148e-03
# age          0.1206292  0.0208618   5.782 1.161e-07
# elev         0.0030852  0.0013329   2.315 2.302e-02
# age:elev    -0.0001472  0.0000431  -3.416 9.722e-04

# [1] 0.3235
```

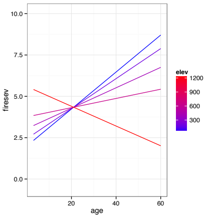
Construct a Data Frame of Lines over Relevant Range

```
pred.df <- expand.grid(age = quantile(keeley$age),
                       elev = quantile(keeley$elev))
pred.df <- cbind(pred.df,
                 predict(keeley_lm, pred.df, interval="confidence"))
#
pred.df$firesev <- pred.df$fit
```

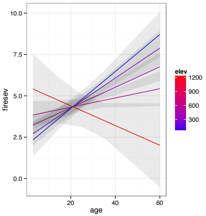
Construct a Data Frame of Lines over Relevant Range

```
keeley_fit <- ggplot(data=pred.df, aes(x=age, y=firesev,
                                        ymin=lwr, ymax=upr,
                                        group=elev)) +
              geom_line(mapping=aes(color=elev)) +
              scale_color_continuous(low="blue", high="red") +
              theme_bw()
#
keeley_fit
```

Construct a Data Frame of Lines over Relevant Range



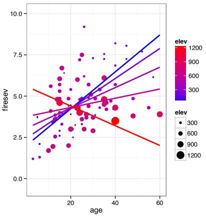
Construct a Data Frame of Lines over Relevant Range



Match Lines with Data Overlay

```
k_plot2 <- k_plot+geom_line(data=pred.df, aes(x=age, y=firesev,
                                              ymin=lwr, ymax=upr,
                                              group=elev), size=1)
k_plot2
```

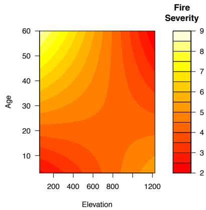
Match Lines with Data Overlay



Surfaces and Other 3d Objects

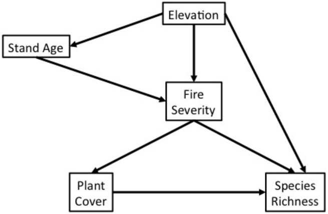
```
kelev <- seq(min(keeley$elev), max(keeley$elev), 1)
kage <- seq(min(keeley$age), max(keeley$age), .1)
#
firesevMat <- outer(kelev, kage,
                    function(x,y) predict(keeley_lm,
                                          data.frame(elev=x, age=y)))
#
filled.contour(kelev, kage, firesevMat,
               color.palette=heat.colors,
               xlab="Elevation", ylab="Age",
               key.title=title(main="Fire\nSeverity"))
```

Surfaces and Other 3d Objects



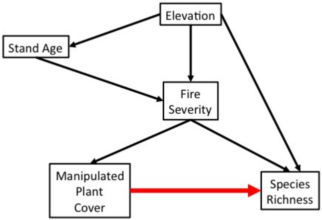
Experimental Design

Causal Diagram of the World

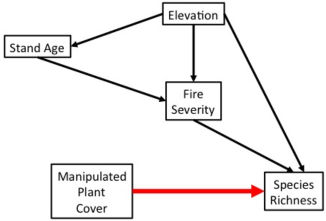


In an experiment, we want to isolate effects between pairs of variables.

Manipulation to Determine Causal Relationship

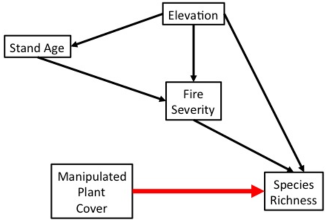


Manipulation to Determine Causal Relationship



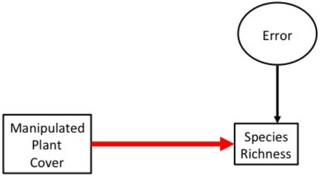
Experimental manipulation (done right) severs the link between a variable and its causes. We can now test the causal effect of changing one variable on another.

Other Sources of Variation are "Noise"



Properly designed experiments will have a distribution of other variables effecting our response variable. We want to reduce BIAS due to biological processes

Other Sources of Variation are "Noise"



AND - this term also includes observer error. We must minimize OBSERVER BIAS as well.

Ensuring that our Signal Comes from our Manipulation

CONTROL

- A treatment against which others are compared
- Separate out causal v. experimental effects
- Techniques to remove spurious effects of time, space, gradients, etc.

Ensuring our Signal is Real

REPLICATION

- How many points to fit a probability distribution?
- Ensure that your effect is not a fluke accident
- $\frac{p^{3/2}}{n}$ should approach 0 for Likleihood (Portnoy 1988 Annals of Statistics)
- i.e.,$\sim$10 samples per paramter (1 treatment = 1 parameter, but this is total # of samples)

Removing Bias and Confounding Effects

TABLE 1. Potential sources of confusion in an experiment and means for minimizing their effect.

Source of confusion	Features of an experimental design that reduce or eliminate confusion
1. Temporal change	Control treatments
2. Procedure effects	Control treatments
3. Experimenter bias	Randomized assignment of experimental units to treatments Randomization in conduct of other procedures "Blind" procedures*
4. Experimenter-generated variability (random error)	Replication of treatments
5. Initial or inherent variability among experimental units	Replication of treatments Interspersion of treatments Concomitant observations
6. Nondemonic intrusion†	Replication of treatments Interspersion of treatments
7. Demonic intrusion	Eternal vigilance, exorcism, human sacrifices, etc.

* Usually employed only where measurement involves a large subjective element.
† Nondemonic intrusion is defined as the impingement of chance events on an experiment in progress.

Randomization Can Come at a Cost

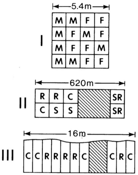


FIG. 2. Three experimental layouts exhibiting partial but inadequate interspersion of treatments. (I) test to compare predation rates on male (M) vs. female (F) floral parts placed on forest floor (Cox 1981, 1982); (II) test of effects on dispersal of removing from unfenced field plots one (S, R), both (SR), or neither (C) of two rodent species (Joule and Cameron 1975); (III) test to compare effects on algae, of removing grazers (R) vs. not doing so (Slocum 1980); shading represents unused portion of study areas.

RCBD (and factorial) and Latin Squares Designs

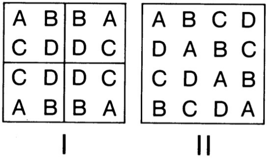
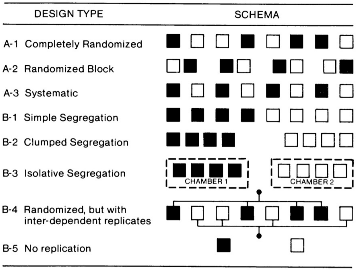
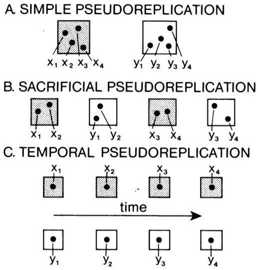


FIG. 3. Examples of segregated arrangements of four treatments, each replicated four times, that can result from use of restricted randomization procedures: (I) randomized block design, (II) Latin square design.

Pseudoreplication and Confounded Results



Types of Psuedoreplication Easily Missed



General Principles of Experimental Design

- Start with a Causal Model. What will you manipulate?
- Consider proper controls & replication
- Consider how treatments will be arrayed accross gradients (known and unknown)
- Rigorously investigate possible pseudoreplication

(Hurlbert 1984)