Factorial Designs & Interaction Effects

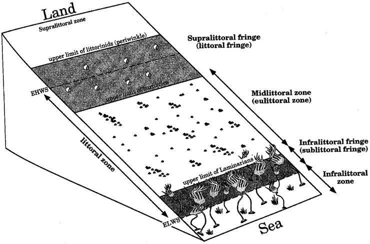
Is the world additive?

- Until now, we have assumed factors combine additively
- BUT - what if the effect of one factor depends on another?
- This is an **INTERACTION** and is quite common
- Yet, challenging to think about, and visualize

Intertidal Grazing!

Do grazers reduce algal cover in the intertidal?

Experiment Replicated on Two Ends of a gradient



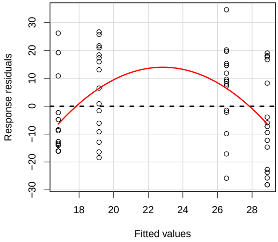
What happens if you fit this data using * instead of + in the linear model?

Humdrum Linear Model

```
graze_linear <- lm(sqrtarea ~ height + herbivores, data=algae)
Anova(graze_linear)

# Anova Table (Type II tests)
#
# Response: sqrtarea
#             Sum Sq Df F value Pr(>F)
# height          89  1    0.32  0.573
# herbivores    1512  1    5.46  0.023
# Residuals    16887 61
```


Pattern in Fitted v. Residuals



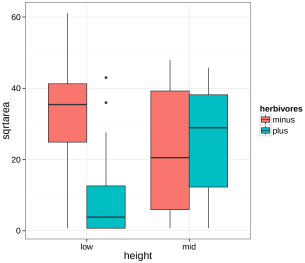
Nonlinearity!

```
residualPlots(graze_linear, plot=F)

#            Test stat Pr(>|t|)
# height            NA       NA
# herbivores        NA       NA
# Tukey test    -3.317    0.001
```

(Note: This test is typically used when there is no replication within blocks)

Problem: Categorical Predictors are Not Additive!



You can only see this if you have replication of treatments (grazing) within blocks (tide height)

Fitting and evaluating interaction effects

The General Linear Model

$$\boldsymbol{Y} = \boldsymbol{\beta}\boldsymbol{X} + \boldsymbol{\epsilon}$$

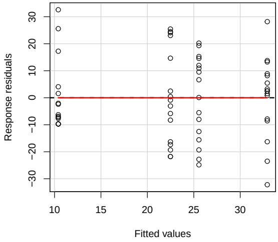
- $\boldsymbol{X}$ can have Nonlinear predictors
- e.g., It can encompass A, B, and A*B

How do you Fit an Interaction Effect?

```
graze_int <- lm(sqrtarea ~ height + herbivores + herbivores:height,
                data=algae)
```

```
#Or, more compact syntax
graze_int <- lm(sqrtarea ~ height*herbivores, data=algae)
```


No More Pattern in Fitted v. Residuals



F-Tests for Interactions

$SS_{Total} = SS_A + SS_B + SS_{AB} + SS_{Error}$

$$SS_{AB} = n\sum_i \sum_j (\bar{Y_{ij}} - \bar{Y_i} - \bar{Y_j} - \bar{Y})^2$$, df=(i-1)(j-1)

MS = SS/DF, e.g, $MS_W = \frac{SS_W}{n-k}$

$F = \frac{MS_{AB}}{MS_{Error}}$ with DF=(j-1)(k-1),n - 1 - (i-1) - (j-1) - (i-1)(j-1)

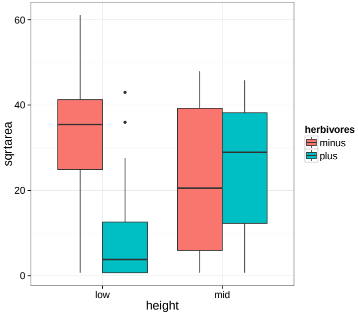
ANOVA shows an Interaction Effect

```
# Anova Table (Type II tests)
#
# Response: sqrtarea
#                    Sum Sq Df F value Pr(>F)
# height                 89  1    0.37 0.5431
# herbivores           1512  1    6.36 0.0144
# height:herbivores    2617  1   11.00 0.0015
# Residuals           14271 60
```

What does the Interaction Coefficient Mean?

```
#                          Estimate Std. Error t value
# (Intercept)                 32.91      3.856   8.537
# heightmid                  -10.43      5.453  -1.913
# herbivoresplus             -22.51      5.453  -4.128
# heightmid:herbivoresplus    25.58      7.711   3.317
#                           Pr(>|t|)
# (Intercept)              5.980e-12
# heightmid                6.052e-02
# herbivoresplus           1.146e-04
# heightmid:herbivoresplus 1.549e-03
```


What does the Interaction Coefficient Mean?



Post-hoc Tests

Posthocs and Factorial Designs

- Must look at simple effects first
  - The effects of individual treatment combinations
- Main effects describe effects of one variable in the complete absence of the other
  - Useful only if one treatment CAN be absent

Posthoc Comparisons Within Blocks

```
library(contrast)

#compare plus in each height to minus in each height
contrast(graze_int,
         a=list(height=levels(algae$height), herbivores="plus"),
         b=list(height=levels(algae$height), herbivores="minus"))
```


Posthoc with Simple Effects Model

```
algae$int <- with(algae, interaction(height, herbivores))
graze_int2 <- lm(sqrtarea ~ int, data=algae)
#
library(multcomp)
summary(glht(graze_int2, linfct=mcp(int = "Tukey")))
```

Posthoc with Simple Effects Model

```
#
#   Simultaneous Tests for General Linear Hypotheses
#
# Multiple Comparisons of Means: Tukey Contrasts
#
#
# Fit: lm(formula = sqrtarea ~ int, data = algae)
#
# Linear Hypotheses:
#                            Estimate Std. Error t value
# mid.minus - low.minus == 0   -10.43       5.45   -1.91
# low.plus - low.minus == 0    -22.51       5.45   -4.13
# mid.plus - low.minus == 0     -7.36       5.45   -1.35
# low.plus - mid.minus == 0    -12.08       5.45   -2.22
# mid.plus - mid.minus == 0      3.07       5.45    0.56
# mid.plus - low.plus == 0      15.15       5.45    2.78
#                            Pr(>|t|)
# mid.minus - low.minus == 0    0.234
# low.plus - low.minus == 0    <0.001
# mid.plus - low.minus == 0     0.535
# low.plus - mid.minus == 0     0.131
# mid.plus - mid.minus == 0     0.943
# mid.plus - low.plus == 0      0.036
# (Adjusted p values reported -- single-step method)
```


Unbalanced Designs

Oh no! I lost a replicate (or two)

```
algae_unbalanced <- algae[-c(1:5), ]

graze_int_unbalanced <- lm(sqrtarea ~ height*herbivores,
                data=algae_unbalanced)
```


Type of Sums of Squares Matters

```
# Analysis of Variance Table
#
# Response: sqrtarea
#                   Df Sum Sq Mean Sq F value Pr(>F)
# height             1    152     152    0.64 0.4279
# herbivores         1   1384    1384    5.82 0.0192
# height:herbivores  1   2934    2934   12.33 0.0009
# Residuals         55  13089     238
# Anova Table (Type II tests)
#
# Response: sqrtarea
#                   Sum Sq Df F value Pr(>F)
# height                78  1    0.33 0.5696
# herbivores          1384  1    5.82 0.0192
# height:herbivores   2934  1   12.33 0.0009
# Residuals          13089 55
```

Enter Type III

```
Anova(graze_int_unbalanced, type="III")

# Anova Table (Type III tests)
#
# Response: sqrtarea
#                    Sum Sq Df F value  Pr(>F)
# (Intercept)         14189  1   59.62 2.5e-10
# height               1176  1    4.94   3e-02
# herbivores           4242  1   17.83 9.2e-05
# height:herbivores    2934  1   12.33   9e-04
# Residuals           13089 55
```


Type I, II, and III Sums of Squares

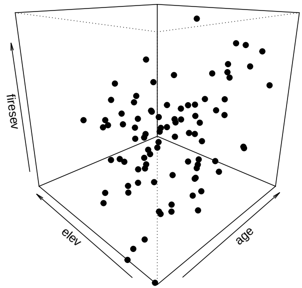
	Type I	Type II	Type III
Test for A	A v. 1	A + B v. B	A + B + A:B v. B + A:B
Test for B	A + B v. A	A + B v. A	A + B + A:B v. A + A:B
Test for A:B	A + B + A:B v. A + B	–	–

Which SS to Use?

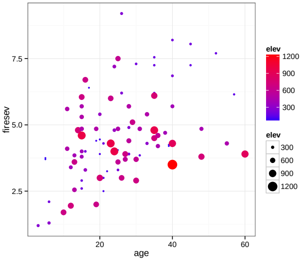
- Traditionally, urged to use Type III
- What do type III models mean?
  - A + B + A:B v. B + A:B
- Interactions the same for all, and if A:B is real, main effects not important
- Type III has lower power for main effects
- Type II produces more meaningful results if main effects are a concern - which they are!

Interactions with Continuous Variable Models

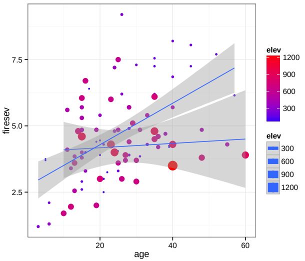
Problem: What if Continuous Predictors are Not Additive?



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
- Check Diagnostics
- Evaluate using type I, II, and III sums of squares

An Interaction Model with Continuous Predictors

```
keeley_lm <- lm(firesev ~ age*elev, data=keeley)
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
```

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
```


Interpreting Continuous Interactions

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
```

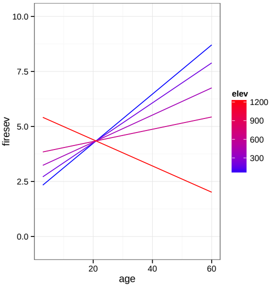

Construct a Data Frame of Lines over Relevant Range

```
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
```

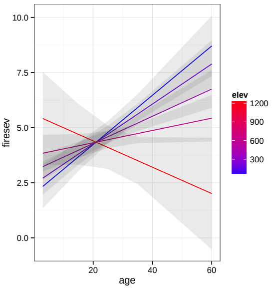
Construct a Data Frame of Lines over Relevant Range

```
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



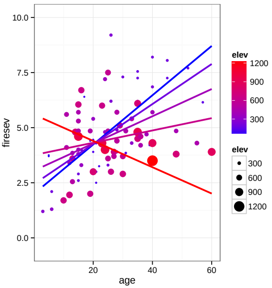
Construct a Data Frame of Lines over Relevant Range



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
```

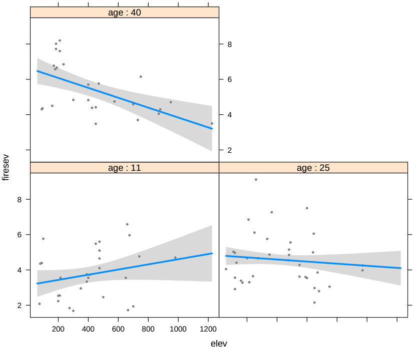
Match Lines with Data Overlay



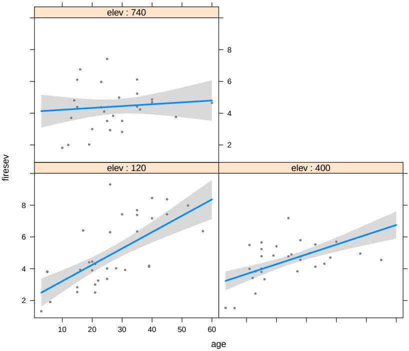
Split Each Predictor into Bins

```
library(visreg)
visreg(keeley_lm, "elev", by="age")
```

Split Each Predictor into Bins



Split Each Predictor into Bins



Surfaces and Other 3d Objects

```
source("./3dplotting.R")
abcSurf(keeley_lm, phi=20, theta=-65, col="lightblue") -> p
with(keeley, scatterPlot3d(age,elev,firesev,
                            add=T, background=p, col="black", alpha=0.4))
```


Surfaces and Other 3d Objects