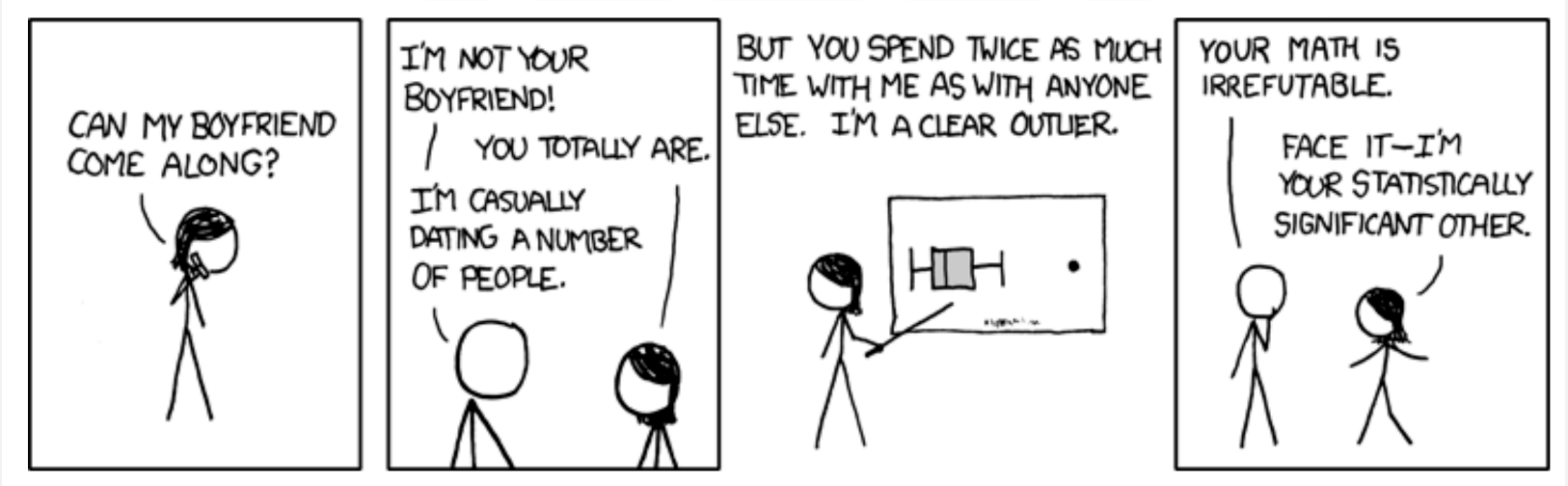
Random vs Fixed Effects
Outliers

Biology 683

Lecture 8



Heath Blackmon

Last week

1. Regression
2. General linear model – all fixed effects

Today

- Fixed effects
  - Two-factor ANOVA
- Random effects
- Mixed effects models

What is a fixed effect

These are the variables whose impact we wish to determine

- Characteristics of the media or habitat
- Alternate medical treatments
- Age groups
- Time points

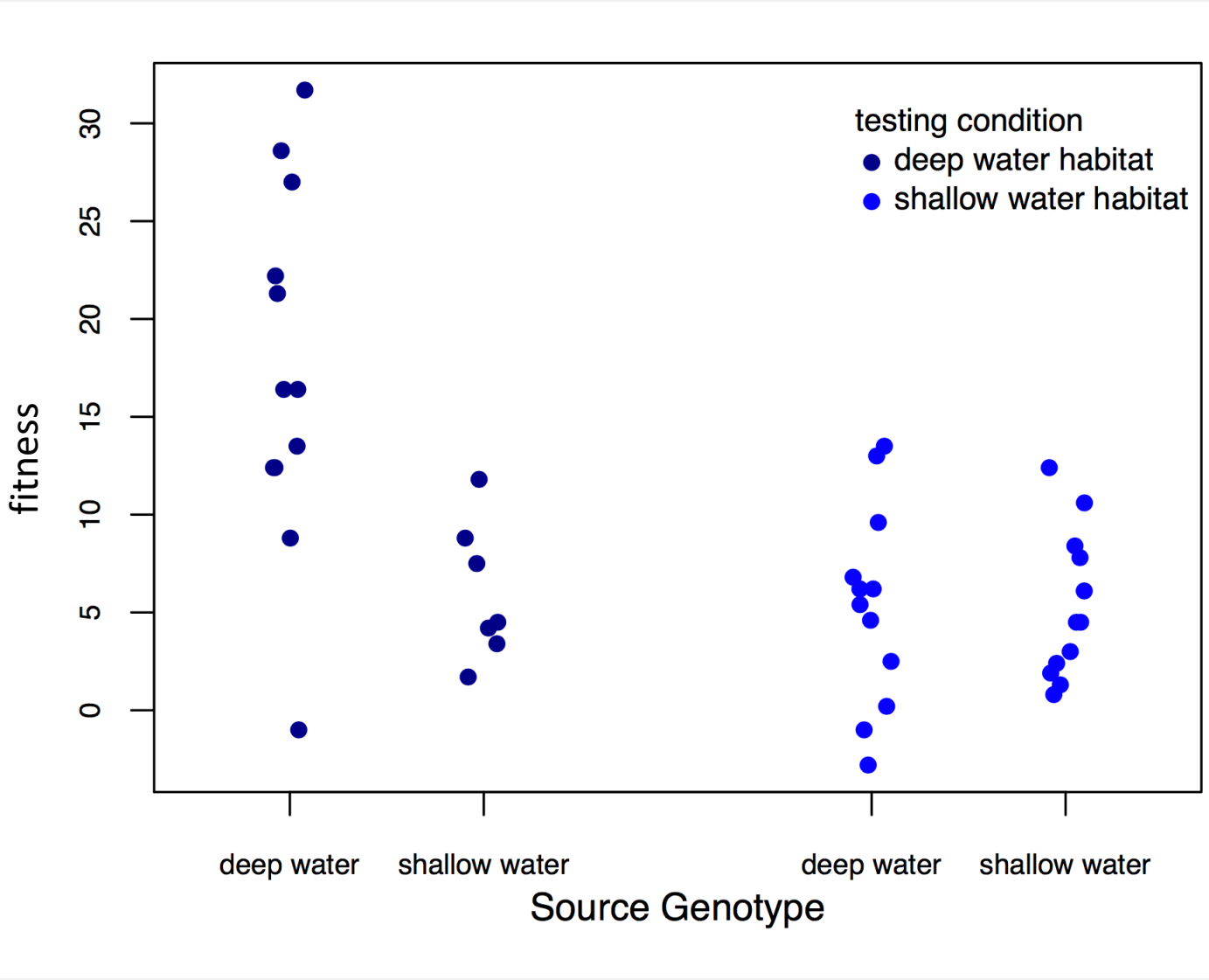
Conclusions that you reach are only applicable to the groups or treatments you include in the study

Example of fixed effects

Reciprocal relocation experiment to investigate how genotype and habitat interact to determine the fitness of stickleback fish (Rundle 2002).

		Source habitat	
		Shallow	Deep
Test habitat	Shallow	12 fish	11 fish
	Deep	7 fish	11 fish

Example of fixed effects (two factor ANOVA)




```
> anova(lm(fitness ~ genotype * test.habitat))
Analysis of Variance Table

Response: fitness
                       Df  Sum Sq Mean Sq F value    Pr(>F)
genotype                1  664.01  664.01 17.7278 0.0001452 ***
test.habitat            1  250.03  250.03  6.6753 0.0136378 *
genotype:test.habitat   1  333.58  333.58  8.9059 0.0048864 **
Residuals              39 1460.77   37.46
```

What is a random effect

These are randomly sampled categories of a variable that represent groups of individual measurements. Usually random effects are not repeatable.

- Study sites
- Environmental chambers
- Families made up of siblings
- Measurements within individuals

Conclusions that you reach are applicable to the population being studied – beyond the levels included in the experiment.

What is a random effect

Sometimes random effects are a nuisance

- Field sites
- Environmental chambers
- Field plots
- Repeated measures

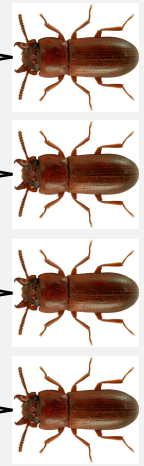
Occasionally random effects are of great interest

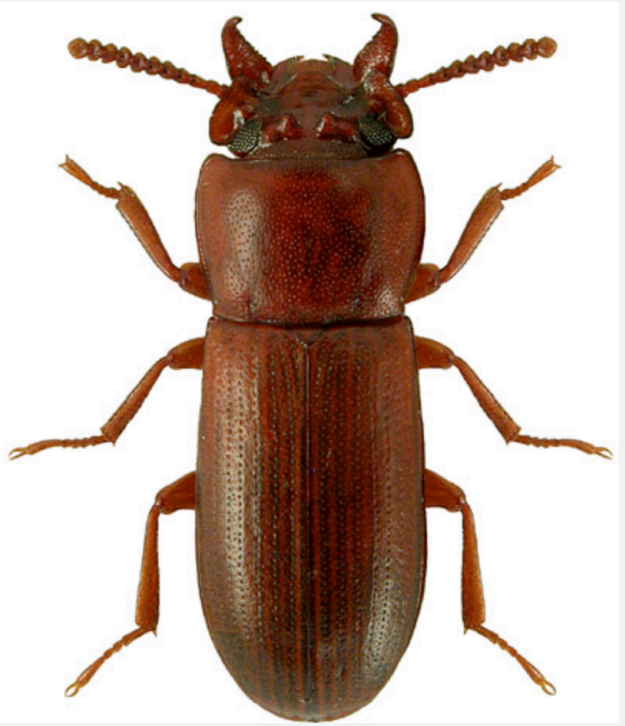
- Families - Heritability
- Individuals – Breeding value

Example of random effects

Impact of selective regime on horn size.
Measure both the left and right horn in 25 beetles from two different selective regimes.

Horn size	Beetle	Selective regime
256	1	High
276	1	High
321	2	High
321	2	High
423	3	Low
401	3	Low
381	4	Low
409	4	Low

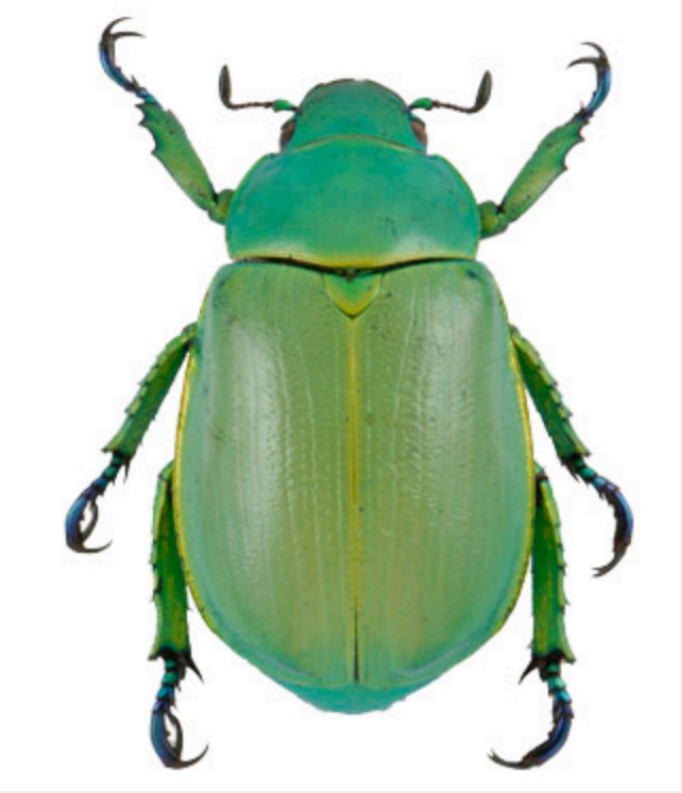




Example of random effects

Identifying the predictors for the presence or absence of Chrysina beetles.

Number collected	oak	juniper	site	date	trip
8	1	0	21	210	A
2	1	1	13	210	A
1	0	1	31	211	A
5	0	1	15	212	A
4	1	1	21	242	B
6	1	0	13	242	B
0	1	1	31	245	B
7	1	1	15	245	B



Implementing a mixed effects model

Mixed effect models can be fit using the LME function from the package nlme.

```
library(nlme)
fit <- lme(sqrt(beyeri) ~ oaks + jun + elev,
           random = list(~1|site,~1|trip),
           data=dat)
summary(fit)
```

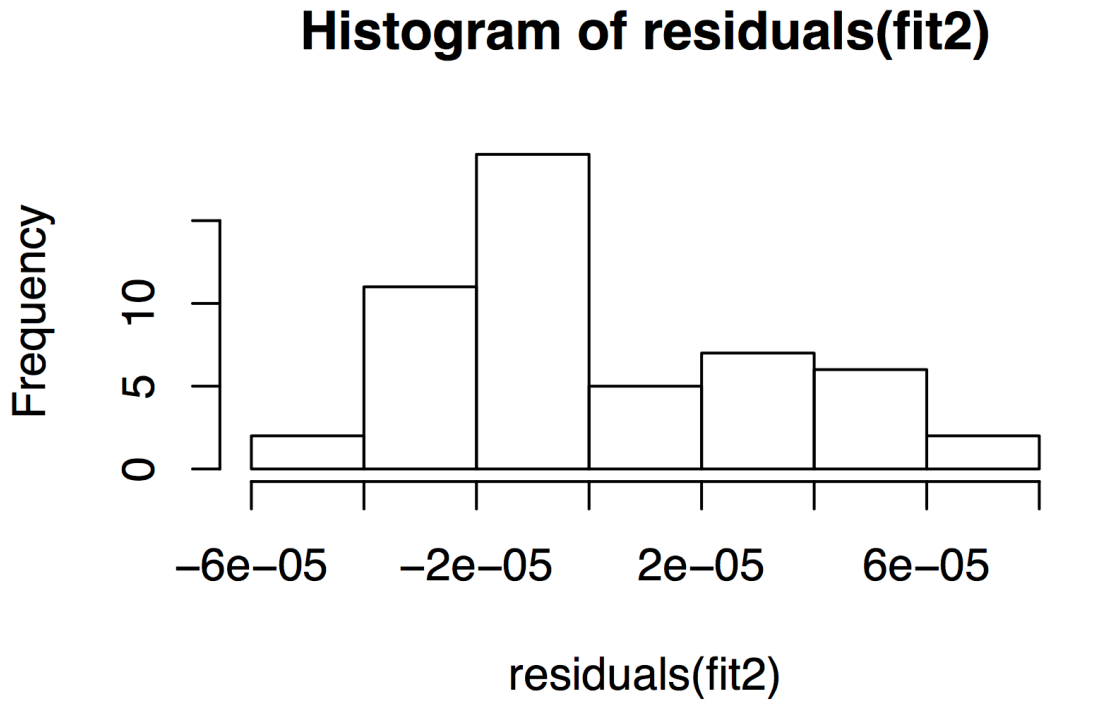
Fixed effects

Random effects

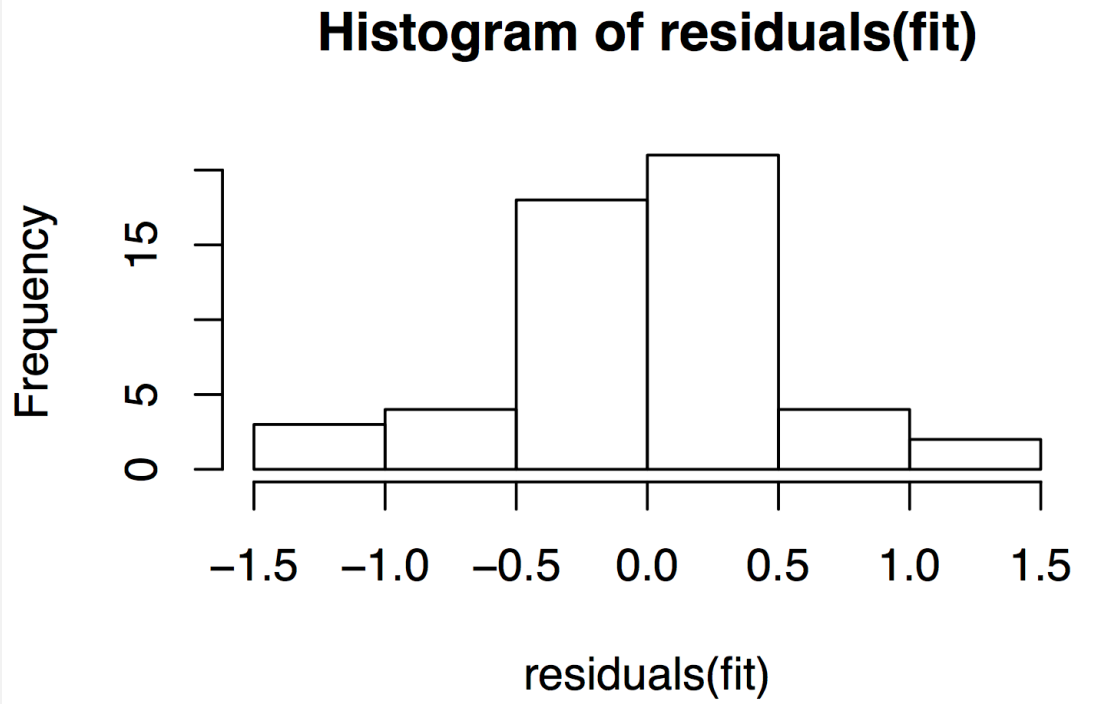
Repeated measures at sites can't be treated as independent

Implementing a mixed effects model

```
fit2 <- lme(beyeri ~ oaks + jun + elev,
            random = list(~1|site, ~1|trip),
            data=dat)
```




```
fit <- lme(sqrt(beyeri) ~ oaks + jun + elev,
           random = list(~1|site,~1|trip),
           data=dat)
```



Implementing a mixed effects model

Mixed effect models can be fit using the LME function from the package nlme.

```
fit <- lme(sqrt(beyeri) ~ oaks + jun + elev,
           random = list(~1|site,~1|trip),
           data=dat)
> summary(fit)
Linear mixed-effects model fit by REML
 Data: dat
       AIC      BIC    logLik
  153.6247 166.7231 -69.81233

Random effects:
 Formula: ~1 | site
        (Intercept)
StdDev: 2.272341e-05

 Formula: ~1 | trip %in% site
       (Intercept)   Residual
StdDev:   0.8169537 0.00264004

Fixed effects: sqrt(beyeri) ~ oaks + jun + elev
                Value Std.Error DF   t-value p-value
(Intercept) -1.6788177 1.5125428 26 -1.109931  0.2772
oaks         0.9707025 0.2943403 22  3.297892  0.0033
jun         -0.0860503 0.2460159 22 -0.349775  0.7298
elev         0.0012513 0.0008072 22  1.550167  0.1354
 Correlation:
     (Intr) oaks   jun
oaks  0.530
jun   0.197 -0.097
elev -0.993 -0.584 -0.250
```

```
fit <- lme(sqrt(beyeri) ~ oaks + elev,
           random = list(~1|site,~1|trip),
           data=dat)
> summary(fit)
Linear mixed-effects model fit by REML
 Data: dat
       AIC      BIC   logLik
  150.7722 162.1231 -69.3861

Random effects:
 Formula: ~1 | site
        (Intercept)
StdDev: 2.191491e-05

 Formula: ~1 | trip %in% site
       (Intercept)    Residual
StdDev:   0.8096043 0.002598713

Fixed effects: sqrt(beyeri) ~ oaks + elev
                Value Std.Error DF   t-value p-value
(Intercept) -1.5744919 1.4695032 26 -1.071445  0.2938
oaks         0.9607576 0.2903283 23  3.309211  0.0031
elev         0.0011807 0.0007746 23  1.524404  0.1410
 Correlation:
     (Intr) oaks
oaks  0.563
elev -0.994 -0.631
```

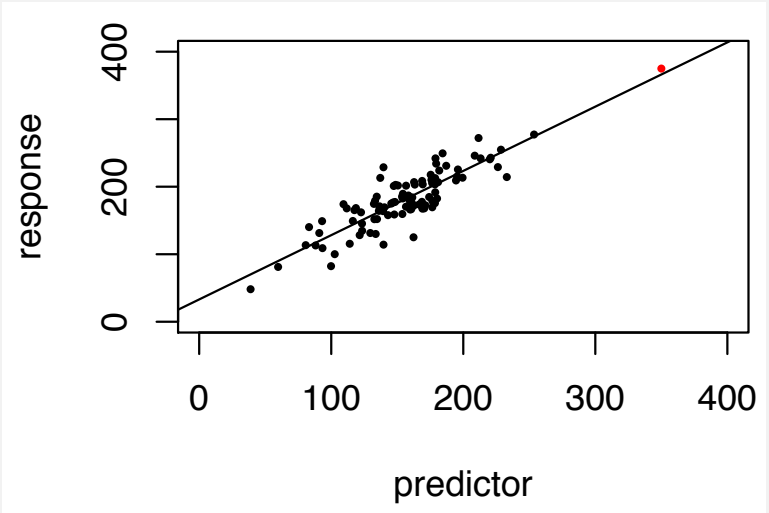
```
fit <- lme(sqrt(beyeri) ~ oaks,
           random = list(~1|site,~1|trip),
           data=dat)
> summary(fit)
Linear mixed-effects model fit by REML
 Data: dat
       AIC      BIC    logLik
  138.5902 148.1504 -64.29512

Random effects:
 Formula: ~1 | site
        (Intercept)
StdDev: 2.309943e-05

 Formula: ~1 | trip %in% site
       (Intercept)    Residual
StdDev:   0.8202517 0.002667513

Fixed effects: sqrt(beyeri) ~ oaks
                Value Std.Error DF  t-value p-value
(Intercept) 0.6514133 0.1674341 26 3.890566   6e-04
oaks        1.2400606 0.2281742 24 5.434710   0e+00
 Correlation:
     (Intr)
oaks -0.734
```

If you report your df with your F-statistic the reviewer will know if you did the right type of model

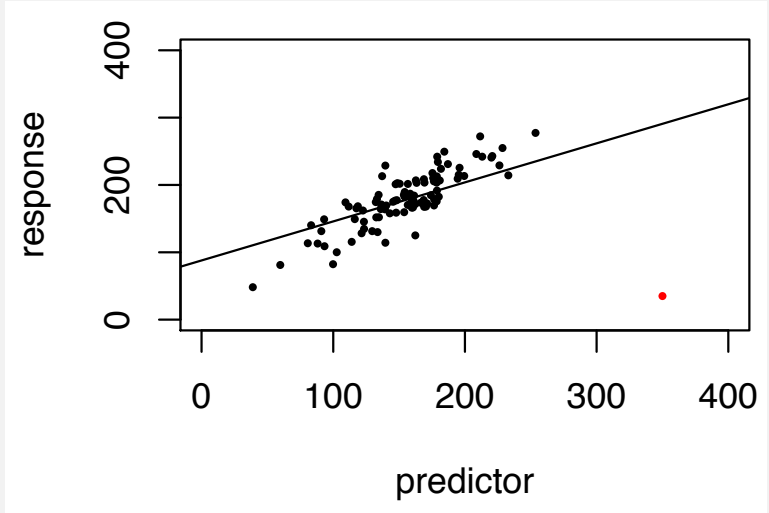
4 Types of outliers

1) Obviously erroneously recorded or measured data

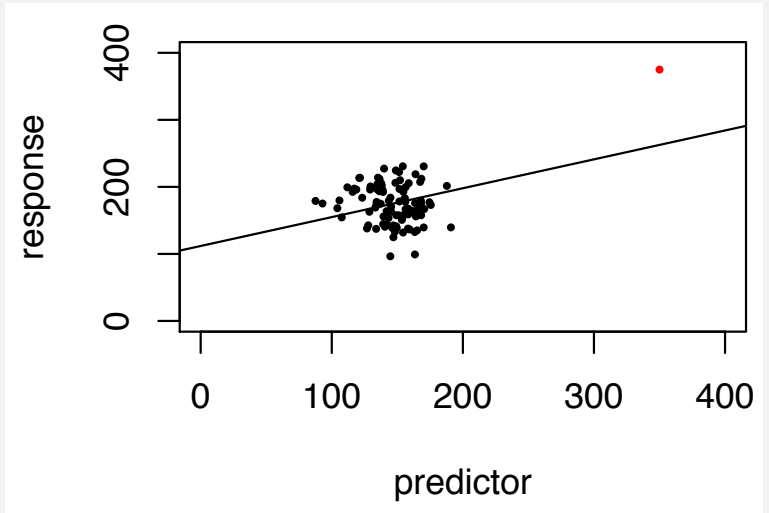
2) Extreme data point that doesn't impact statistic of interest.



3) Extreme data point that impacts statistic of interest.



4) Extreme data point that creates significance.



Considerations for models with random effects

- Most software will assume that all factors are fixed unless you specify them as mixed.
- Designating factors as random effects takes extra work.
- The lm function treats all predictors as fixed effects.
- Treating random effects as fixed effects is fundamentally wrong.

Links

MCMCMglmm: Fit mixed models with phylogenetic or pedigree information in a Bayesian framework.

Outlier Package: Apply outlier tests to identify possible outlier datapoints - I don’t recommend this.

For Thursday

Bring laptop to class!

Plans for the rest of the semester.

Heath Blackmon
BSBW 309
coleoguy@gmail.com
@coleoguy

Models

"Remember that all models are wrong; the practical question is how wrong do they have to be to not be useful."

George Box