

Random Effects

36-600

The Setting (At Least, Initially)

- The one-way analysis of variance (ANOVA) model is

$$Y_{ij} = \mu + \tau_i + \epsilon_{ij}$$

- μ : an overall mean response value (constant)
 - τ_i : a deterministic shift in the mean for group i (also constant)
 - ϵ_{ij} : the random variation for datum j of group i
 - Y_{ij} : the observed value for datum j of group i
- Here, there is an implicit assumption that every possible group has been identified and is represented in the dataset
 - for instance, in a particular study, we might have vaccine-taking, placebo-taking, and no treatment groups, and these would be the only three possible groups
- What if, for instance, we only have data for a subset of all possible groups?
 - for instance, in a particular study, we might examine data for all the players on m teams in a league, where the overall number of teams is $n > m$
 - the answer: we would move to the setting of random effects modeling

The Setting (At Least, Initially)

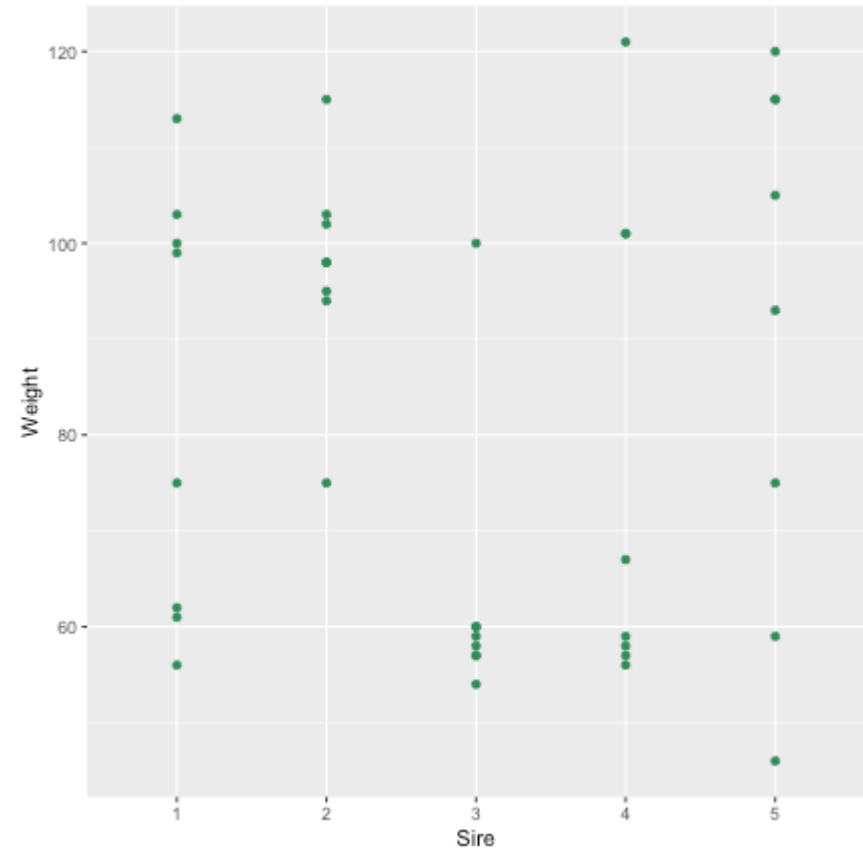
- For the setting in which we are analyzing groups of data (without regard to any other possible predictor variables), the random effects model will be

$$Y_{ij} = \mu + \tau_i + \epsilon_{ij}$$

- "uh...this is the same model as on the last slide"
- but it isn't: the ANOVA model assumes the τ_i 's are constants, whereas here we will assume that $\tau_i \sim \mathcal{N}(0, \sigma_\tau^2)$
- (in both models, $\epsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$)
- "But how does this change the model, really?"
- One-way ANOVA modeling and random effects modeling have different goals:
 - one-way ANOVA: perform the hypothesis test $\tau_1 = \dots = \tau_k$ versus "at least one of the τ 's differs from the others"
 - random effects: how large is σ_τ^2 (variance of the means across groups) compared to σ^2 (variance of the data within each group)?
- For a deeper introduction, see [this presentation](#)

Example: Calf Birth Weight

- The data shown at right come from [this book](#)
 - they show the birth weight of calves born to five different fathers ("sires")
 - since the number of possible sires is $\gg 5$, a random effects model is appropriate here



Example: Calf Birth Weight

```
# this needs to be installed!
suppressMessages(library(lme4))

# 1|sire == "a random intercept for each sire"
# (more on this notation later)
re.out <- lmer(weight ~ 1|sire,data=animals)
summary(re.out)
```

- REML = "restricted maximum likelihood"
 - a maximum likelihood estimator that acts to counteract biases in the estimation of variances
- Fixed effect: the mean birth weight is 82.55
- Random effects:
 - the sire-to-sire variance in birth weight is estimated to be $\sigma_{\tau}^2 = 116.7$ (so that the standard error is $\sqrt{116.7} = 10.81$)
 - the within-sire variance in birth weight is $\sigma^2 = 463.8$ (standard error $\sqrt{463.8} = 21.54$)

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: weight ~ 1 | sire
##      Data: animals
##
## REML criterion at convergence: 358.2
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.9593 -0.7459 -0.1581  0.8143  1.9421
##
## Random effects:
##   Groups      Name      Variance Std.Dev.
##   sire      (Intercept) 116.7     10.81
##   Residual                463.8     21.54
## Number of obs: 40, groups:  sire, 5
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)   82.550      5.911    13.96
```

Example: Calf Birth Weight

```
# this needs to be installed!
suppressMessages(library(lme4))

# 1|sire == "a random intercept for each sire"
# (more on this notation later)
re.out <- lmer(weight ~ 1|sire,data=animals)
summary(re.out)
```

- The overall variance in birth weights is $\sigma^2 + \sigma_\tau^2 = 580.5$
- The correlation (magnitude of linear dependence) between two calves...
 - with different sires: 0 (no correlation)
 - with the same sire: $\sigma_\tau^2 / (\sigma_\tau^2 + \sigma^2) = 0.20$ (this is a "weak" correlation)

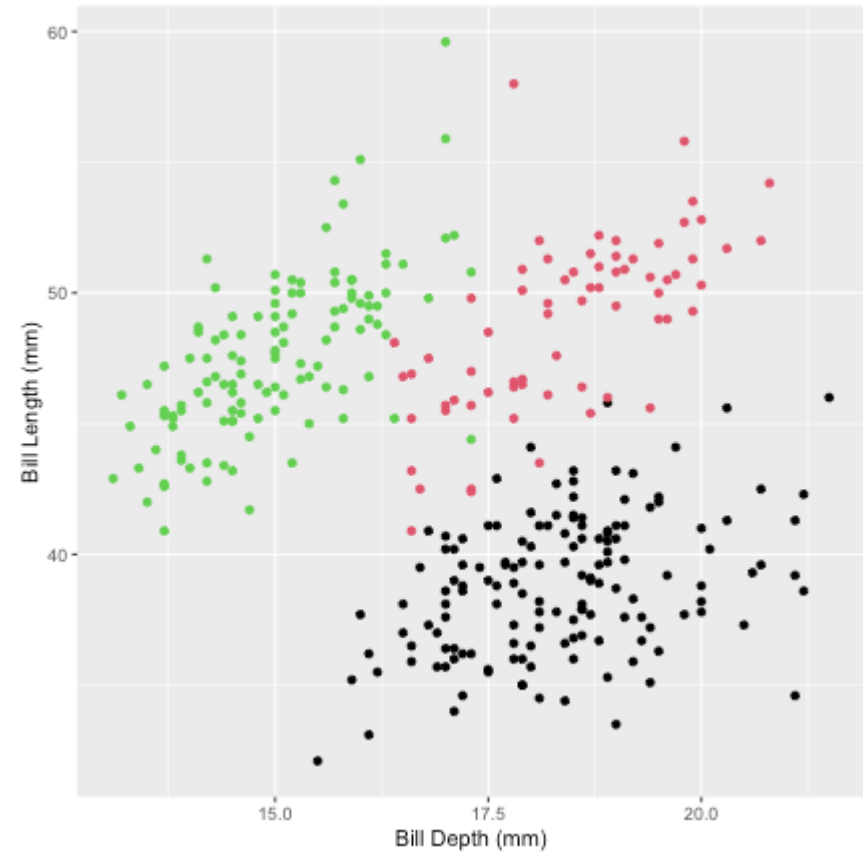
```
## Linear mixed model fit by REML ['lmerMod']
## Formula: weight ~ 1 | sire
##      Data: animals
##
## REML criterion at convergence: 358.2
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.9593 -0.7459 -0.1581  0.8143  1.9421
##
## Random effects:
##   Groups      Name      Variance Std.Dev.
##   sire      (Intercept) 116.7     10.81
##   Residual                463.8     21.54
## Number of obs: 40, groups:  sire, 5
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)   82.550      5.911   13.96
```

Example: Palmer Penguins

```
# this needs to be installed
library(palmerpenguins)

w <- complete.cases(penguins)
df <- data.frame(penguins[w,c(1,4,3)])

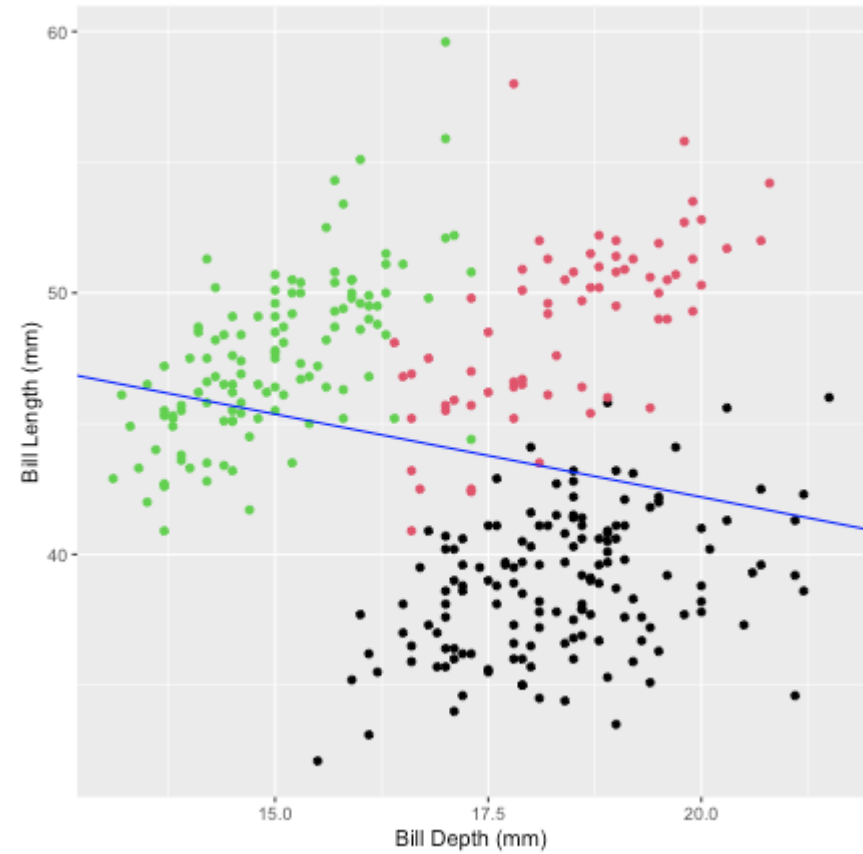
ggplot(data=df,
        mapping=aes(x=bill_depth_mm,y=bill_length_mm)
        geom_point(col=as.numeric(df$species)) +
        xlab("Bill Depth (mm)") + ylab("Bill Length (mm)"))
```



Example: Palmer Penguins

```
# regress bill length upon bill depth  
lm.out <- lm(bill_length_mm ~ bill_depth_mm, data=df)  
AIC(lm.out)
```

```
## [1] 2063.692
```



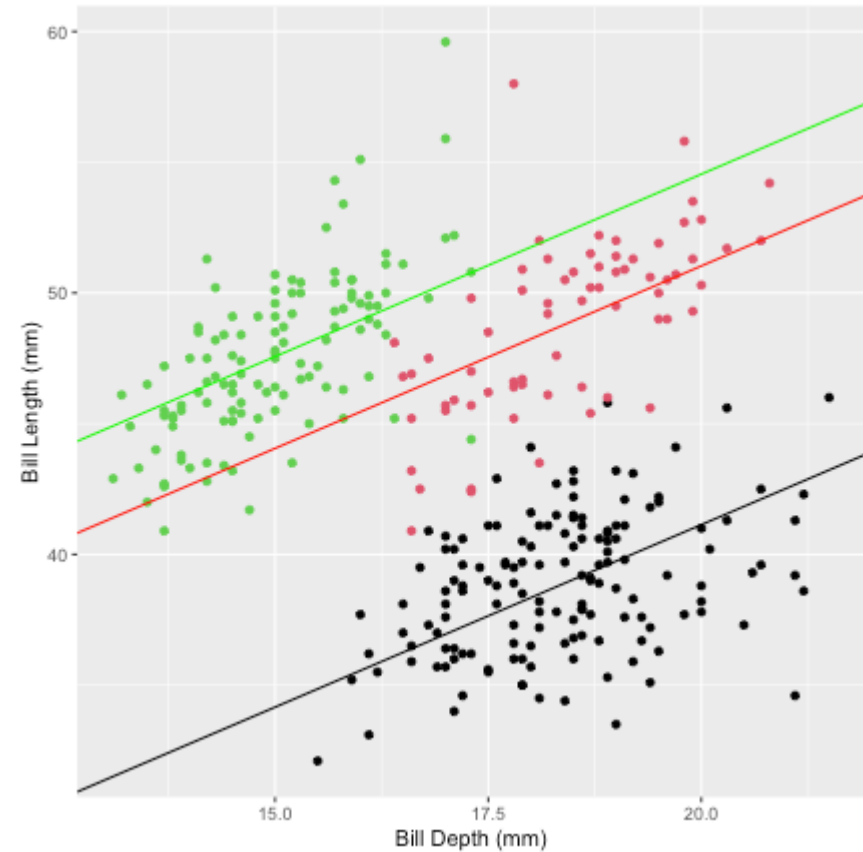
Example: Palmer Penguins - Random Intercepts

```
# regress bill length upon bill depth
# make species a random effect (random intercept)
lm.out <- lmer(bill_length_mm ~ bill_depth_mm +
               (1|species),data=df)
coef(lm.out)$species
```

```
##              (Intercept) bill_depth_mm
## Adelie           13.22455      1.395653
## Chinstrap        23.12096      1.395653
## Gentoo           26.63174      1.395653
```

```
AIC(lm.out) # much lower - better model
```

```
## [1] 1585.28
```



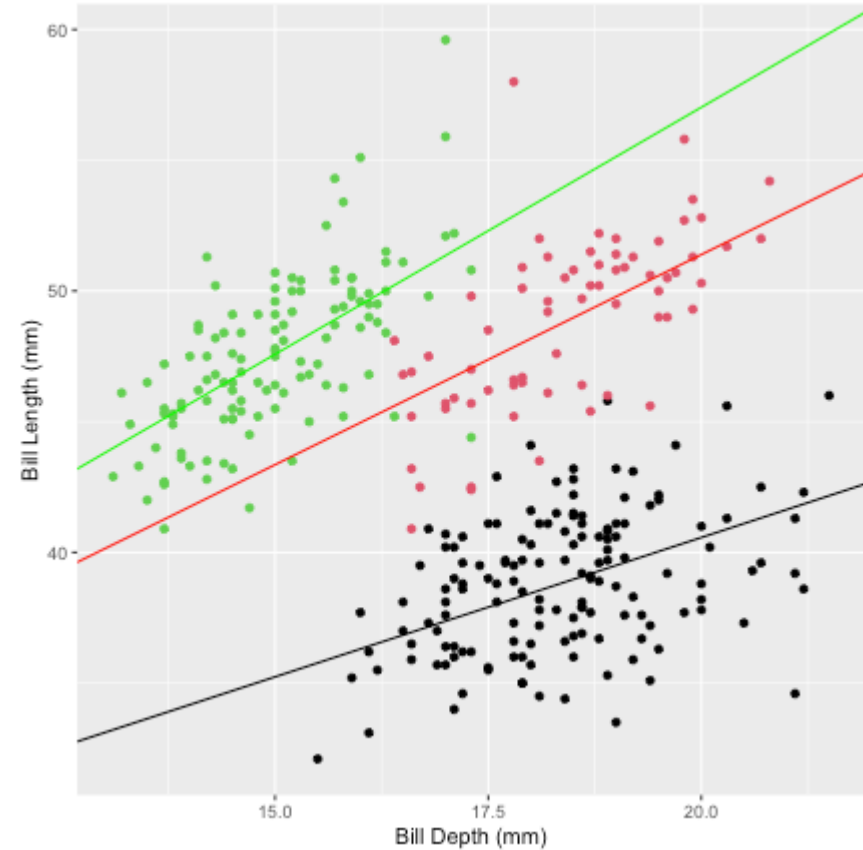
Example: Palmer Penguins - Random Slopes

```
# regress bill length upon bill depth
# make species a random effect (random slopes)
lm.out <- lmer(bill_length_mm ~ bill_depth_mm +
              (0 + bill_depth_mm|species),data=df)
coef(lm.out)$species
```

```
##              (Intercept) bill_depth_mm
## Adelie           19.24352         1.066549
## Chinstrap        19.24352         1.607421
## Gentoo           19.24352         1.888989
```

```
AIC(lm.out) # sufficiently lower-better model still
```

```
## [1] 1566.371
```



Example: Palmer Penguins - Random Intercepts and Slopes

```
# regress bill length upon bill depth
# make species a random effect
# (random intercepts and slopes)
lm.out <- lmer(bill_length_mm ~ bill_depth_mm +
               (bill_depth_mm|species),data=df)
```

```
## boundary (singular) fit: see help('isSingular')
```

```
coef(lm.out)$species
```

```
##           (Intercept) bill_depth_mm
## Adelie      22.72927      0.8773012
## Chinstrap   18.23566      1.6623990
## Gentoo      15.52816      2.1354384
```

```
AIC(lm.out) # higher than before - not best model
```

```
## [1] 1567.762
```

