Stat 414 - Day 8
Fixed vs. Random Effects

Last Time

- ICC is a way of measuring how similar observations are within the same group of a categorical variable.
- One way to model correlated errors is using generalized least squares and specifying specific correlation patterns to be estimated/

$$corr(\epsilon_{ij},\epsilon_{kl}) = \begin{matrix} \rho \; if \; j = l \\ otherwise \end{matrix}$$

	ϵ_{11}	ϵ_{21}	ϵ_{31}	ϵ_{12}	ϵ_{22}	ϵ_{23}	ϵ_{31}	ϵ_{32}	ϵ_{33}
ϵ_{11}	1	ρ	ρ	0	0	0	0	0	0
ϵ_{21}	ρ	1	ρ	0	0	0	0	0	0
ϵ_{31}	ρ	ρ	1	0	0	0	0	0	0
ϵ_{12}	0	0	0	1	ρ	ρ	0	0	0
ϵ_{22}	0	0	0	ρ	1	ρ	0	0	0
ϵ_{23}	0	0	0	ρ	ρ	1	0	0	0
ϵ_{31}	0	0	0	0	0	0	1	ρ	ρ
ϵ_{32}	0	0	0	0	0	0	ρ	1	ρ
ϵ_{33}	0	0	0	0	0	0	ρ	ρ	1

correlation matrix

- Accounting for the correlation will impact the effective sample size
- Expect $SE(\bar{y})$ to increase (by 1 + $(\bar{n} - 1)\rho$)
- $SE(\hat{\beta}_j)$ usually increases as well, especially if x_j mostly varies between clusters.
- If you have an important categorical variable, especially if you used randomized block design or cluster sampling, then that variable *must* be included in your model.
- Is possible this will also account for the correlated observations (e.g., within a region, heart disease rates from different cities are essentially independent)
- When adding a categorical variable to the model, we have the choice of treating that variable as a *fixed effect* (what we have been doing) or as a *random effect*.

Reminder:

- When variables are centered, interpret the intercept at the average value of the predictor variable(s).

Random effects models

Model equation: $Y_{ij} = \beta_0 + u_j + \epsilon_{ij}$ where $u_j \sim N(0,\tau^2)$ and $\epsilon_{ij} \sim N(0,\sigma^2)$. We also assume $cov(u_j,\epsilon_{ij}) = 0$.

Consequences

$$V(Y_{ij}) = V(\beta_0) + V(u_j) + V(\epsilon_{ij}) = \tau^2 + \sigma^2$$

$$
\begin{aligned}
Cov(Y_{ij}, Y_{kj}) &= Cov(u_j + \epsilon_{ij}, u_j + \epsilon_{kj}) = Cov(u_j, u_j) + Cov(u_j, \epsilon_{kj}) + Cov(u_j, \epsilon_{ij}) + Cov(\epsilon_{ij}, \epsilon_{kj}) \\
&= Cov(u_j, u_j) + Cov(\epsilon_{ij}, \epsilon_{kj}) = \tau^2
\end{aligned}
$$

Example 1: Finger tapping

```
fingertapstudy = read.table("http://www.isi-stats.com/isi2/data/Fingertap.txt","\t"
, header=TRUE)

#install.packages("nlme")
library(nlme)
rm1 = lme(fixed = Taps ~ 1 , random = ~1 | participant, data = fingertapstudy, meth
od="REML")

summary(rm1)
Linear mixed-effects model fit by REML
  Data: fingertapstudy
    AIC   BIC logLik
  102.3 103.5 -48.17

Random effects:
 Formula: ~1 | participant
        (Intercept) Residual
StdDev:       23.63    12.27

Fixed effects:  Taps ~ 1
            Value Std.Error DF t-value p-value
(Intercept)   474     12.34  8   38.43       0

Standardized Within-Group Residuals:
    Min      Q1      Med      Q3     Max
-1.2496 -0.7895  0.1400  0.5698  1.3015

Number of Observations: 12
Number of Groups: 4
logLik(rm1)
'log Lik.' -48.17 (df=3)
```

(a) How many parameters are estimated in this model? What are they?

3 parameters: intercept, tau, sigma

(b) What are the estimated variance components?

$\hat{\tau}$ = 23.63 (between group variability)

$\hat{\sigma}$ = 12.127 (within group variability)

(c) What is the estimated total variance in the response using this model?

$23.63^2 + 12.27^2$ = 709

(d) What is the variance-covariance matrix of the responses?

```
getVarCov(rm1, subject = "1", type = "marginal")[[1]]
      1     2     3
1 709.0 558.5 558.5
2 558.5 709.0 558.5
3 558.5 558.5 709.0
cov2cor(getVarCov(rm1, subject = "1", type = "marginal")[[1]])
       1      2      3
1 1.0000 0.7877 0.7877
2 0.7877 1.0000 0.7877
3 0.7877 0.7877 1.0000
```

Here, marginal means we are looking at the y-values rather than the errors.

```
#what does this give you?
getVarCov(rm1, type = "conditional")
participant A
Conditional variance covariance matrix
      1     2     3
1 150.5   0.0   0.0
2   0.0 150.5   0.0
3   0.0   0.0 150.5
  Standard Deviations: 12.27 12.27 12.27
```

These are the $\hat{\sigma}^2$ values (within group)

(d) How much of the (estimated) total variation is due to the groups?

$$\hat{\tau}^2/(\hat{\tau}^2 + \hat{\sigma}^2)$$

This number should look familiar:

```
#install.packages("performance")
library(performance)
performance(rm1)
# Indices of model performance

AIC   |  AICc |   BIC | R2 (cond.) | R2 (marg.) |   ICC |   RMSE |  Sigma
-------------------------------------------------------------------------
102.3 | 105.3 | 103.8 |      0.788 |          0 | 0.788 | 10.170 | 12.268
icc(rm1)
# Intraclass Correlation Coefficient

    Adjusted ICC: 0.788
  Unadjusted ICC: 0.788
```

(e) Are they any additional model assumptions to worry about?

Have the usual linearity and equal variability and now normality of both “epsilons” and “uj”

Key Ideas

- Random effects models account for both group to group variation and intraclass correlation at once.
- They are much more robust to issues like unequal group sizes, missing observations.
- Can show they often have better predictive performance than fixed effects models (see HW 4).
- You can think of the u_i as "level 2 residuals"

Example 2: Adult literacy

Recall the study on adult literacy from HW 3.

```
adultlit <- read.table("https://www.rossmanchance.com/stat414/data/adultlit.txt", h
eader=TRUE)
model1 <- lm(sessions ~ group, data = adultlit); summary(model1)

Call:
lm(formula = sessions ~ group, data = adultlit)

Residuals:
   Min     1Q Median     3Q    Max 
-6.686 -1.686  0.314  2.314  8.314 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)    5.280      0.342   15.45   <2e-16 ***
group          1.405      0.504    2.79    0.006 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.1 on 150 degrees of freedom
Multiple R-squared:  0.0493,	Adjusted R-squared:  0.043 
F-statistic: 7.78 on 1 and 150 DF,  p-value: 0.00597
```

Treatments were randomly assigned to classes, so it's important that we acknowledge this in the analysis. We could scale the standard errors using effective sample size. This gave us a SE for the treatment variable of about 0.8 rather than 0.5 so the treatment variable was no longer significant.

Alternatively, we can include the class variable in the model. Doing so as fixed effects had some issues (see HW 3), mostly due to unequal sample sizes and small sample sizes (e.g., class 361 had only 2 students).

```
table(adultlit$classid)

 57  58  84  85  86  92 114 115 116 175 176 177 178 192 242 293 361 362 401 421 
  7   8   9   7   4   5   4   6   5   7   6   7   5   6   6   3   2   7   6   5 
422 441 451 452 471 481 491 541 
  5   5   5   5   4   4   4   5 
```

But we can easily fit the model treating the classid as random effects.

```
modelRE = lme(fixed = sessions ~ 1 + group, random = ~1 | classid, data = adultlit)
summary(modelRE)
Linear mixed-effects model fit by REML
  Data: adultlit
    AIC   BIC logLik
  753.2 765.2 -372.6

Random effects:
 Formula: ~1 | classid
        (Intercept) Residual
StdDev:       1.935    2.491

Fixed effects:  sessions ~ 1 + group
            Value Std.Error  DF t-value p-value
(Intercept) 5.209    0.5904 124   8.823  0.0000
group       1.485    0.8422  26   1.763  0.0896
 Correlation:
      (Intr)
group -0.701

Standardized Within-Group Residuals:
    Min      Q1     Med      Q3     Max
-2.5208 -0.4369  0.1710  0.5866  1.9801

Number of Observations: 152
Number of Groups: 28
```

(a) What is the SE of the group variable in this model?

.84, very similar to the .8 you got in the homework by tweaking based on effective sample size

(b) Does it make sense to treat the classid as random effects in this study?

Random vs. Fixed Effects

- do you care about that factor (e.g., don't necessarily want to compare the participants to each other, but do want to compare the treatments to each other)
- is it reasonable to consider the "level 2" units as a random sample or was that an intention feature of the study design? If so, RE will allow you to generalize from the units in your study to the population of units

classid in adult literacy study: random

Finger tapping study participants: random

heart disease (pace of life) regions: fixed

Here is another way of fitting a model with 'random intercepts'

```
#install.packages("lme4")
library(lme4)
```

```
modelRE2 = lmer(sessions ~ 1 + group + (1 | classid), data = adultlit)
summary(modelRE2)
Linear mixed model fit by REML ['lmerMod']
Formula: sessions ~ 1 + group + (1 | classid)
   Data: adultlit

REML criterion at convergence: 745.2

Scaled residuals:
   Min     1Q Median     3Q    Max
-2.521 -0.437  0.171  0.587  1.980

Random effects:
 Groups   Name        Variance Std.Dev.
 classid  (Intercept) 3.74     1.94
 Residual             6.20     2.49
Number of obs: 152, groups:  classid, 28

Fixed effects:
            Estimate Std. Error t value
(Intercept)    5.209      0.590    8.82
group          1.485      0.842    1.76

Correlation of Fixed Effects:
      (Intr)
group -0.701
```

(c) What is the syntax for specifying the fixed and random effects? Does this function use ML or REML by default? What are the estimated variable components? What is missing from the output? Is this a problem? What is unnecessary in the output?

syntax: (1 | groupid) convey the component

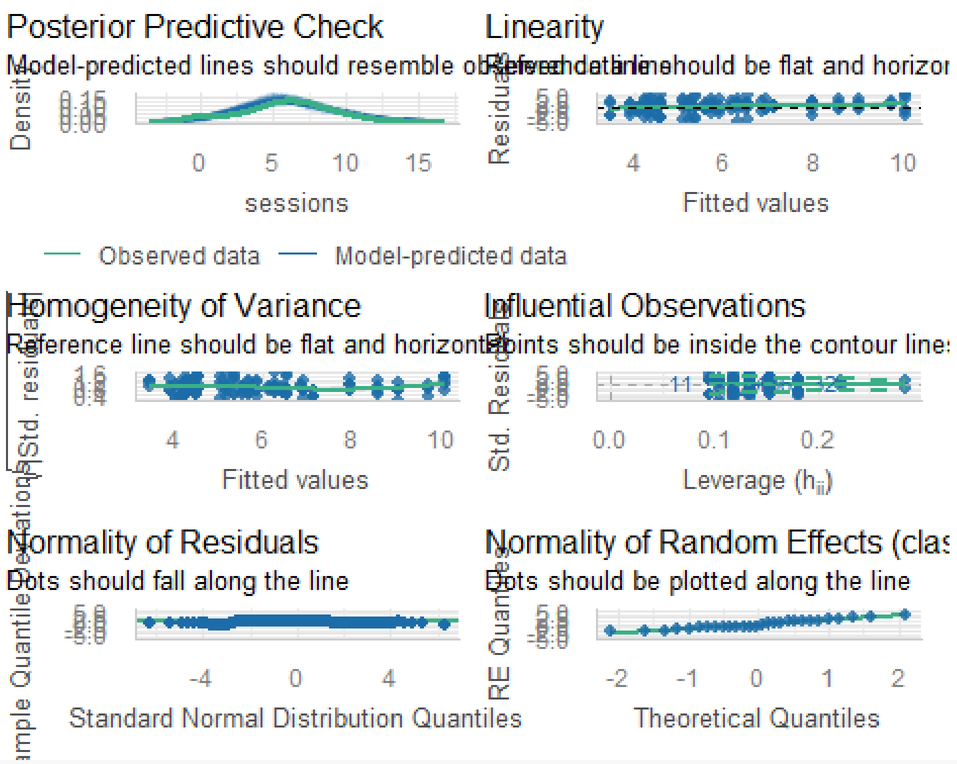
default: REML

same estimates but now both variances and standard deviations for variance components and no p-values for fixed effects (but that's ok, can use 2 as a cutoff for significant t-tests)

will soon stop outputting the 'correlation of fixed effects'

There is another nice graph you can use with lmer:

```
performance::check_model(modelRE2) #install see package
```



```
#qqnorm(residuals(modelRE2))
```

(d) What do you learn?

Model assumptions appear to be met

Shrinkage Estimation

Even though we say we are not all that interested in the individual u_j and that they aren't really parameters but "unobservable latent effects," we do still get estimates for them that are used to estimate τ and it might still be interesting to explore those estimates (e.g., do they appear to be normally distributed?) But how are they estimated differently?

Example 3: Back to finger tapping

```
#The output no longer gives us the estimated effects for the participants, but R do
es store them for us.
ranef(rm1)
  (Intercept)
A      -15.60
B      -13.76
C       -3.67
D       33.03
#Fitted values, prediction for each participant
fits=predict(rm1); fits
    A     A     A     B     B     B     C     C     C     D     D     D
458.4 458.4 458.4 460.2 460.2 460.2 470.3 470.3 470.3 507.0 507.0 507.0
```

```
attr(,"label")
[1] "Fitted values"
fitted.values(rm1)
    A     A     A     B     B     B     C     C     C     D     D     D 
458.4 458.4 458.4 460.2 460.2 460.2 470.3 470.3 470.3 507.0 507.0 507.0 
attr(,"label")
[1] "Fitted values"
#How do these compare to the participant means?
tapply(fingertapstudy$Taps, fingertapstudy$participant, mean);
  A   B   C   D 
457 459 470 510 
tapply(fits, fingertapstudy$participant, mean)
    A     B     C     D 
458.4 460.2 470.3 507.0 
#Storing these for later
#library(tidyverse)
pmeans <- fingertapstudy |>
  group_by(participant) |>
  mutate(Tapmean = mean(Taps)) |>
  ungroup()
#pmeans
```

(a) Why are these called *shrinkage estimates*? What are they being shrunk towards?

shrinking towards the grand mean

Compare to the fixed effect model (with effect coding):

```
modelB <- lm(Taps ~ participant, contrasts = list(participant = contr.sum), data = 
fingertapstudy)
fitted.values(modelB)
  1   2   3   4   5   6   7   8   9  10  11  12 
457 457 457 459 459 459 470 470 470 510 510 510 
```

Definitions

One way to estimate a participant's effect is to ignore all the other participant, call this *no pooling*. Another way is to ignore the player to player differences and use the overall mean, call this *complete pooling*. Treating the player as a random effect creates *partial pooling*. We can think of each predicted group mean as being a weighted average of the group mean and the overall mean: $w(group\ mean) + (1-w)(overall\ mean)$ where the weight for group j, (w_j), depends on the relative sizes of the variance components and on the group size, $w_j = \tau^2/\left(\tau^2 + \sigma^2/n_j\right)$). The weights reflect the "reliability" of the group.

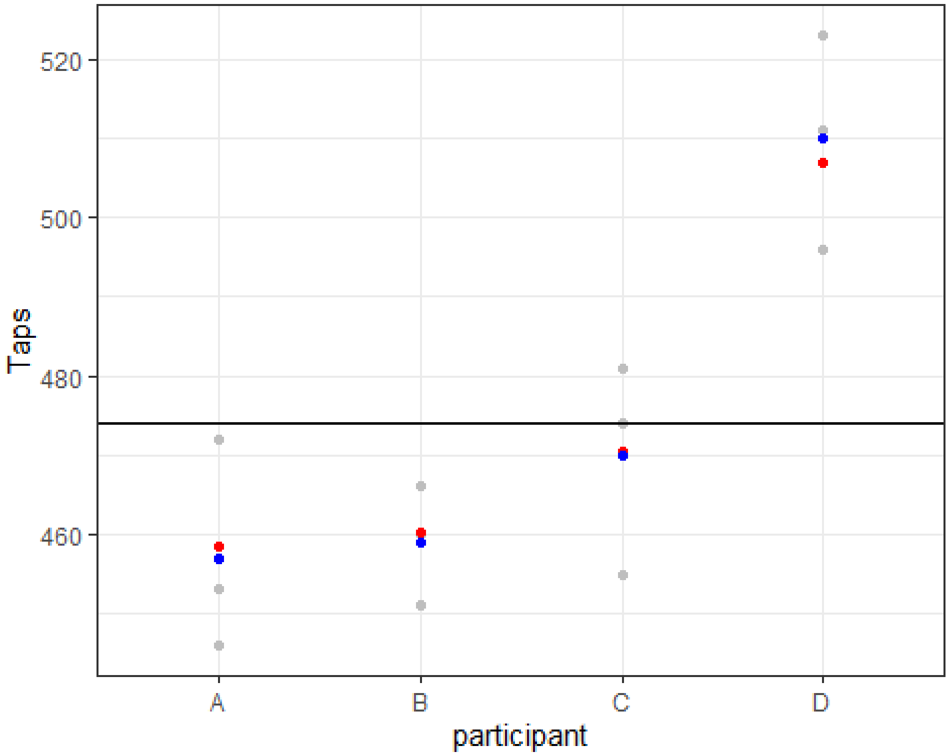
(b) Calculate the weight for participant 1? How will it compare to the weight for participant 2?

$23.63^2/(23.63^2 + 12.27^2/3)$ = .9175

```
participant 2 is the same because n_j = 3 for each participant
```

(c) Summarize what you learn from the following graph

```
ggplot(fingertapstudy, aes(x=participant, y = pmeans$Tapmean, group=participant)) +
  geom_point(aes(y = Taps), col="grey") +
  geom_point(aes(y=fits),  col="red") +
  geom_point(aes(y = pmeans$Tapmean), col="blue") +
  geom_hline(aes(yintercept = mean(Taps))) +
  theme_bw()
```



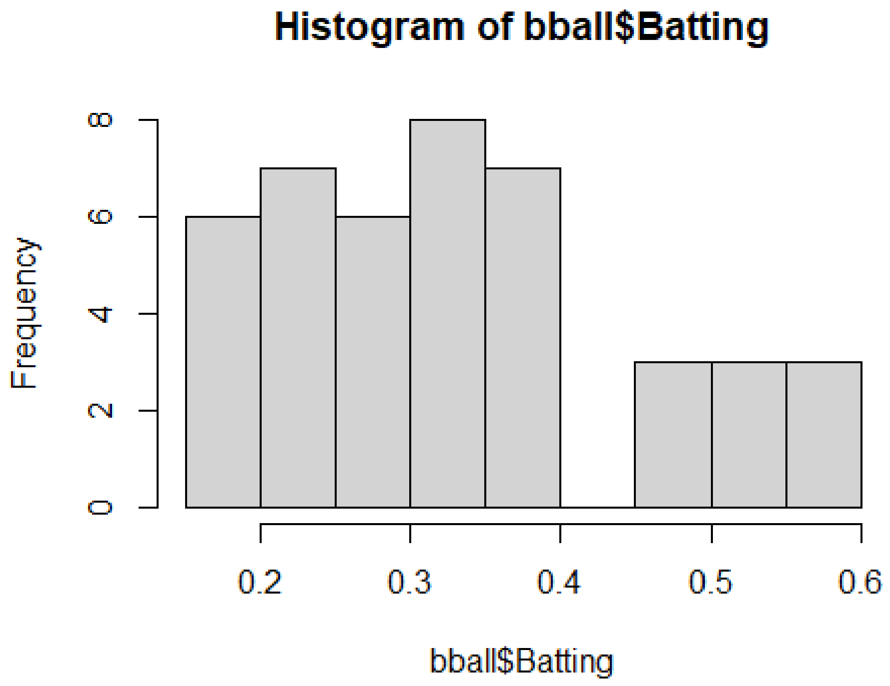
(d) Which mean appears to change the most and which the least? Why is that?

participant D just because further away to ybar (estimate of overall mean) participant C just because closer to estimate for overall mean

Example 4: Fake baseball

Suppose the data in bball.txt are the batting averages for 6 players over several seasons.

```
bball <-read.table("http://www.rossmanchance.com/stat414/data/bball.txt",header=T)
tail(bball)
      Player Batting
38 Rodriguez    0.60
39 Rodriguez    0.50
40 Rodriguez    0.55
41    Suarez    0.60
42    Suarez    0.50
43    Suarez    0.55
hist(bball$Batting)
```



```
load(url("http://www.rossmanchance.com/iscam4/ISCAM.RData"))
iscamsummary(bball$Batting)
  Missing  n  Min    Q1 Median  Q3 Max  Mean    SD Skewness
1       0 43 0.15 0.235  0.325 0.4 0.6 0.344 0.129    0.548
iscamsummary(bball$Batting, bball$Player)
          Missing  n  Min    Q1 Median    Q3  Max  Mean    SD Skewness
Anderson        0  6 0.24 0.285  0.300 0.315 0.33 0.295 0.032   -0.736
Jones           0 11 0.15 0.190  0.210 0.222 0.24 0.202 0.029   -0.612
Mitchell        0  6 0.20 0.300  0.320 0.385 0.40 0.323 0.075   -0.511
Rodriguez       0  6 0.50 0.512  0.550 0.587 0.60 0.550 0.045    0.000
Smith           0 11 0.30 0.325  0.325 0.400 0.40 0.357 0.042    0.082
Suarez          0  3 0.50 0.525  0.550 0.575 0.60 0.550 0.050    0.000
```

(a) Do you expect a high or low ICC value? Explain.

group means differ by almost .2 but the group SDs are below .1 so maybe a larger ICC

(b) Opinion: Do you really think Rodriguez and Suarez are that much better than everyone else? What else could be going on? Which averages do you find the most/least "trustworthy"? Why?

means based on small samples sizes are not terribly trustworthy

```
#Library(nlme)
model2 = lme(fixed = Batting ~ 1, random = ~ 1 | Player, data = bball)
summary(model2)
Linear mixed-effects model fit by REML
  Data: bball
```

```
     AIC    BIC logLik
  -111.1 -105.9  58.55

Random effects:
 Formula: ~1 | Player
        (Intercept) Residual
StdDev:      0.1402  0.04481

Fixed effects:  Batting ~ 1
             Value Std.Error DF t-value p-value
(Intercept) 0.3788   0.05772 37   6.564       0

Standardized Within-Group Residuals:
     Min       Q1      Med       Q3      Max
-2.77325 -0.71463  0.08025  0.77774  1.69027

Number of Observations: 43
Number of Groups: 6
ranef(model2)
          (Intercept)
Anderson     -0.08244
Jones        -0.17495
Mitchell     -0.05458
Rodriguez     0.16829
Smith        -0.02183
Suarez        0.16552
#compare the fixed effects
bball$Player = factor(bball$Player) #I believe default is now alphabetical
contrasts(bball$Player) <- contr.sum
summary(model1gls <- gls(Batting ~ Player, data = bball))$coefficients
(Intercept)     Player1     Player2     Player3     Player4     Player5
    0.37957    -0.08457    -0.17730    -0.05624     0.17043    -0.02275
```

(c) How do these player estimated effects compare to model 1? (You may need to find Suarez.)

The 'effect's are closer to zero with the random effects model

We can also convert the effects to the predicted values for each player

```
#compare the estimated means
tapply(model2$fitted[,2], bball$Player, mean) #random
 Anderson     Jones  Mitchell Rodriguez     Smith    Suarez
   0.2964    0.2039    0.3243    0.5471    0.3570    0.5444
tapply(model1gls$fitted,bball$Player, mean) #fixed
 Anderson     Jones  Mitchell Rodriguez     Smith    Suarez
   0.2950    0.2023    0.3233    0.5500    0.3568    0.5500
```

(d) Whose estimates (Jones or Suarez) changed more? Why does that make sense for these data?

Suarez is a bit more different because was further away from grand mean to beging with

(e) Calculate the weights for Jones and Suarez. Why are the weights pretty large? Which is larger? Why?

$$\hat{\tau} = .140; \hat{\sigma} = .0448$$

Jones: $.1401714^2/(.1401714^2 + .04480776^2/11)$ = .9907

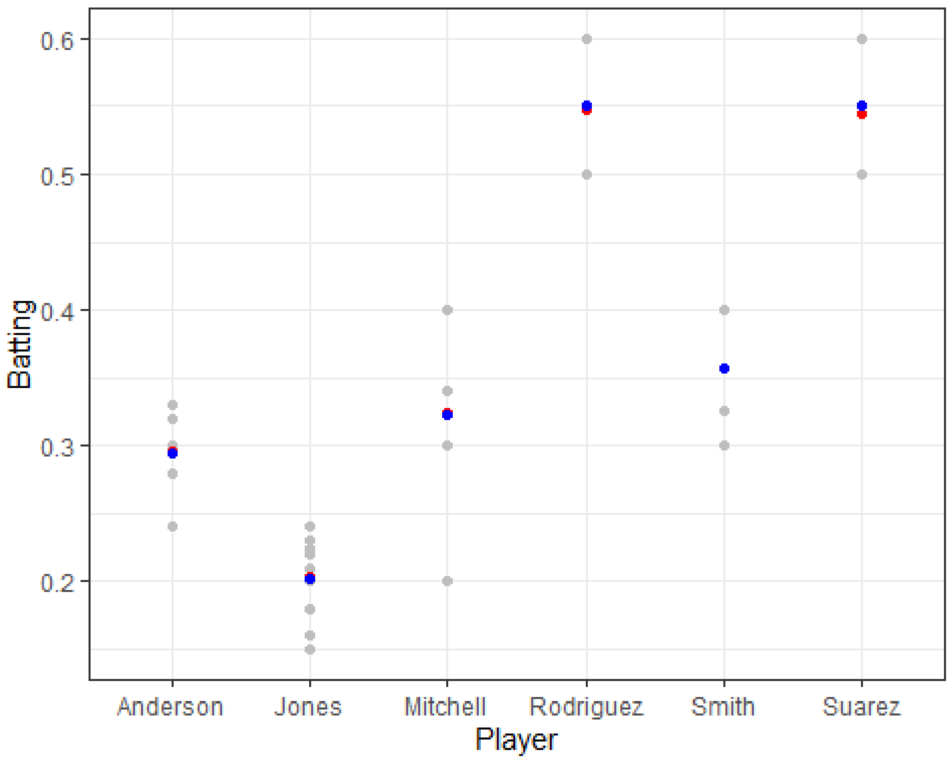
Suarez: $.1401714^2/(.1401714^2 + .04480776^2/3)$ = .967 (more shrinkage because smaller sample size)

(f) Verify the estimated group means for Jones and Suarez using these weights. Which changes (from the group mean) more? Why?

Jones: .9907 x .202 + (1-.9907) x .378847 = .2036

Suarez: .967 x .550 + (1-.967) x .378847 = .544

```
playermeans <- bball |>
  group_by(Player) |>
  mutate(Batmean = mean(Batting)) |>
  ungroup()
playermeans
# A tibble: 43 × 3
   Player Batting Batmean
   <fct>    <dbl>   <dbl>
 1 Smith    0.3     0.357
 2 Smith    0.4     0.357
 3 Smith    0.325   0.357
 4 Smith    0.4     0.357
 5 Smith    0.325   0.357
 6 Smith    0.4     0.357
 7 Smith    0.325   0.357
 8 Smith    0.4     0.357
 9 Smith    0.325   0.357
10 Smith    0.4     0.357
# ℹ 33 more rows
ggplot(bball, aes(x=Player, y = playermeans$Batmean, group=Player)) +
  geom_point(aes(y = Batting), col="grey") +
  geom_point(aes(y=predict(model2)),  col="red") +
  geom_point(aes(y = playermeans$Batmean), col="blue") +
  theme_bw()
```



Notes

- Degree of shrinkage depends on the variance of the effect (τ) and the number of observations per level in the effect. With large variance estimates, there is little shrinkage.
  - You can consider fixed effects as a special case of random effects where the variance component is very large.
  - If variance component is small, then more shrinkage.
- If the variance component is zero, the effect levels are shrunk to exactly zero. It is even possible to obtain highly negative variance components where the shrinkage is reversed.
  - If very few observations per level, then more shrinkage.
  - If many observations per level, the estimates shrink less.
  - You can consider fixed effects as a special case with infinitely many observations.