Homework 10

Steve Harms

April 7, 2018

2)

a)

```
mod1 <- lm(data = Yield, yield~genotype)
blues <- mod1$coefficients + mod1$coefficients[1]
blues[1] <- blues[1]/2
blues
## (Intercept)    genotype2    genotype3    genotype4    genotype5    genotype6
##    194.9333     184.2500     191.4333     198.6000     194.1667     197.6800
##    genotype7    genotype8    genotype9   genotype10   genotype11   genotype12
##    193.5600     196.6000     202.3250     196.1000     182.1250     186.1200
##   genotype13   genotype14   genotype15   genotype16   genotype17   genotype18
##    183.9667     170.2000     189.9250     187.4200     192.6000     189.2667
##   genotype19   genotype20   genotype21   genotype22   genotype23   genotype24
##    185.6600     197.7667     199.3333     192.8500     183.1000     200.8000
##   genotype25   genotype26   genotype27   genotype28   genotype29   genotype30
##    190.5500     182.6667     192.8250     187.2000     195.5333     194.0500
##   genotype31   genotype32   genotype33   genotype34   genotype35   genotype36
##    178.6667     203.7500     185.4667     189.5750     196.0000     187.0667
##   genotype37   genotype38   genotype39   genotype40   genotype41   genotype42
##    188.2500     190.3500     185.4000     191.4500     186.2000     187.2000
##   genotype43   genotype44   genotype45   genotype46   genotype47   genotype48
##    189.7200     179.5750     189.0667     190.5000     185.6667     206.2000
##   genotype49   genotype50   genotype51   genotype52   genotype53   genotype54
##    192.2000     194.3500     197.6667     184.5000     193.2400     182.2000
##   genotype55   genotype56   genotype57   genotype58   genotype59   genotype60
##    192.1000     188.9000     185.3500     193.0500     198.3333     192.4333
##   genotype61   genotype62   genotype63   genotype64   genotype65   genotype66
##    189.2000     181.5000     192.2333     189.3000     196.0667     201.5000
##   genotype67   genotype68   genotype69   genotype70   genotype71   genotype72
##    194.0500     194.3667     181.1333     201.3000     185.4500     183.6667
##   genotype73   genotype74   genotype75   genotype76   genotype77   genotype78
##    194.6250     196.0750     196.8400     179.6500     191.5200     196.0667
##   genotype79   genotype80   genotype81   genotype82   genotype83   genotype84
##    190.1000     186.4750     194.1000     189.4333     184.5000     191.5000
##   genotype85   genotype86   genotype87   genotype88   genotype89   genotype90
##    177.4000     193.0333     194.7000     196.1500     197.4333     190.5000
##   genotype91   genotype92   genotype93   genotype94   genotype95   genotype96
##    189.9000     193.9333     177.5500     183.8500     189.3000     178.2500
##   genotype97   genotype98   genotype99  genotype100
##    200.3000     188.7750     192.9000     191.8250
```

b)

```
mod <- lme(data = Yield, yield~1, random = ~1|genotype)
mod
## Linear mixed-effects model fit by REML
##   Data: Yield
##   Log-restricted-likelihood: -1130.573
##   Fixed: yield ~ 1
## (Intercept)
##    190.6983
##
## Random effects:
##  Formula: ~1 | genotype
##         (Intercept) Residual
## StdDev:    2.686537 9.669021
##
## Number of Observations: 304
## Number of Groups: 100
sigmag <- 2.686537
sigmae <- 9.669021
```

c)

```
blups <- mod$coefficients$fixed +
  mod$coefficients$random$genotype + mod$coefficients$random$genotype[1]
blups[1] <- mod$coefficients$fixed[1] + mod$coefficients$random$genotype[1]
blups
##      (Intercept)
## 1       191.4947
## 2       190.6322
## 3       191.6329
## 4       192.9806
## 5       192.5927
## 6       193.4391
## 7       192.2917
## 8       192.2841
## 9       194.2379
## 10      192.2172
## 11      189.4719
## 12      190.2196
## 13      190.2288
## 14      188.7531
## 15      191.3122
## 16      190.5817
## 17      191.7491
## 18      191.2255
## 19      190.0915
## 20      192.8239
## 21      194.2283
## 22      191.7825
## 23      190.9502
## 24      192.8458
## 25      191.4597
## 26      189.9844
## 27      191.9965
## 28      190.8369
## 29      192.4039
## 30      191.9430
## 31      189.2322
## 32      193.2404
## 33      190.5109
## 34      191.2297
## 35      192.2038
## 36      190.8118
## 37      191.1672
## 38      191.4481
## 39      190.7861
## 40      191.5952
## 41      191.1723
## 42      191.2440
## 43      191.2222
## 44      188.8702
## 45      191.1879
## 46      191.4574
## 47      190.5485
## 48      192.6057
## 49      191.6023
## 50      191.9831
## 51      192.8051
## 52      191.0505
## 53      192.2026
## 54      190.3581
## 55      191.7583
## 56      191.1565
## 57      189.8016
## 58      191.8092
## 59      192.9305
## 60      191.8210
## 61      191.2129
```

```
## 62      189.3244
## 63      191.7834
## 64      191.1648
## 65      192.5042
## 66      192.9394
## 67      191.9430
## 68      192.1845
## 69      189.6960
## 70      193.4883
## 71      190.2564
## 72      190.1724
## 73      192.4212
## 74      192.7633
## 75      193.2052
## 76      190.0170
## 77      191.7235
## 78      192.5042
## 79      191.3535
## 80      190.4982
## 81      192.1344
## 82      191.2568
## 83      191.0505
## 84      191.6455
## 85      188.9940
## 86      191.9338
## 87      191.7815
## 88      192.7810
## 89      192.7612
## 90      191.4574
## 91      191.3879
## 92      192.1030
## 93      189.7361
## 94      190.5787
## 95      191.2318
## 96      189.8297
## 97      193.7602
## 98      191.0409
## 99      191.7892
## 100     191.7605
```

d)

The BLUEs and BLUPs do not have a perfectly linear relationship because when we take into account the within-genotype variance, we can get better estimates of the yield for each genotype. The variation from the line y=x is due to taking the random effect into account in prediction. However, note that there are approximately the same number of points above and below the line, because the expected value of our random effect is 0.

```r
plot(x=blues, y = blups) + abline(a = 0, b = 1, col = "blue", lwd = 1)
```
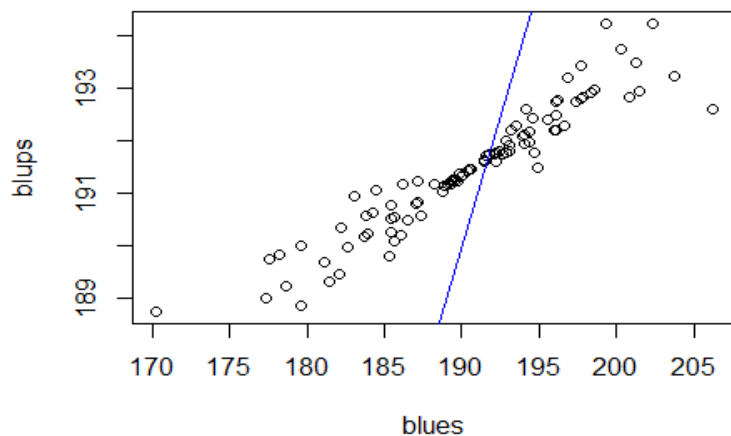
e)

```
ests <- data.frame(unique(Yield$genotype),as.numeric(blues), blups)
names(ests) <- c("genotype", "blues", "blups")
ests %>% arrange(desc(blues)) %>% head()
##    genotype   blues    blups
## 1        48 206.200 192.6057
## 2        32 203.750 193.2404
## 3         9 202.325 194.2379
## 4        66 201.500 192.9394
## 5        70 201.300 193.4883
## 6        24 200.800 192.8458
```

f)

```
ests%>% arrange(desc(blups)) %>% head()
##    genotype    blues    blups
## 1         9 202.3250 194.2379
## 2        21 199.3333 194.2283
## 3        97 200.3000 193.7602
## 4        70 201.3000 193.4883
## 5         6 197.6800 193.4391
## 6        32 203.7500 193.2404
```

g)

The top BLUEs do not match the top BLUPs because some of the genotypes are penalized for having higher variance within their groups (genotype 9, for example). After we take the random effect due to genotype into account, our best estimates of group means change.