yutingq2_final_CDAP

Yuting Qiu

12/12/2019

```r
library(lme4)
```

```
## Loading required package: Matrix
```

```r
library(lmerTest)
```

```
##
## Attaching package: 'lmerTest'

## The following object is masked from 'package:lme4':
##
##     lmer

## The following object is masked from 'package:stats':
##
##     step
```

```r
library(lsmeans)
```

```
## Loading required package: emmeans

## Welcome to emmeans.
## NOTE -- Important change from versions <= 1.41:
##     Indicator predictors are now treated as 2-level factors by default.
##     To revert to old behavior, use emm_options(cov.keep = character(0))

## The 'lsmeans' package is now basically a front end for 'emmeans'.
## Users are encouraged to switch the rest of the way.
## See help('transition') for more information, including how to
## convert old 'lsmeans' objects and scripts to work with 'emmeans'.
```

```r
library(reshape2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library(ggplot2)
library(glmnet)
```

```
## Loaded glmnet 3.0
```

```r
library(caret)
```

```
## Loading required package: lattice
```

```r
library(MASS)
```

```
##
## Attaching package: 'MASS'
```

```
## The following object is masked from 'package:dplyr':
##
##     select
```

```r
library(tidyverse)
```

```
## -- Attaching packages -----------------------------------------------------------------------
```

```
## v tibble  2.1.3     v purrr   0.3.2
## v tidyr   0.8.3     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.4.0
```

```
## -- Conflicts ------------------------------------------------------------------------------
## x tidyr::expand() masks Matrix::expand()
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## x purrr::lift()   masks caret::lift()
## x MASS::select()  masks dplyr::select()
```

```r
read.table("examFile.csv", sep=",",header= TRUE,na.strings = "NA")->df
# remove the genotype variable
df_clean <- df[,1:7]

# replace east with north, replace west with south
df_clean$Reg <- replace(as.character(df_clean$Reg), df_clean$Reg == "east", "north")
df_clean$Reg <- replace(as.character(df_clean$Reg), df_clean$Reg == "west", "south")
```

Obj 1. Evaluate location effects:

a. What fraction of the variation observed in yield is attributable to Location specific effects?

```r
model1 <- lm(Estimate~Loc, data=df_clean)
anova(model1) #location effect is significant
```

```
## Analysis of Variance Table
##
## Response: Estimate
##            Df Sum Sq Mean Sq F value    Pr(>F)
## Loc         5  46321  9264.2   318.1 < 2.2e-16 ***
## Residuals 410  11941    29.1
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(model1)$adj.r.squared #0.7925499
```

```
## [1] 0.7925499
```

79.25% of the variation observed in the yield is attributable to location specific effect

b. Which location seems to be the highest yield location?

```r
summary(model1)
```

```
##
## Call:
## lm(formula = Estimate ~ Loc, data = df_clean)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -17.8688  -3.4353   0.0322   3.7131  16.7053
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)        76.7404     0.6190  123.97   <2e-16 ***
## LocElkville_2019   24.4245     0.8755   27.90   <2e-16 ***
## LocHampshire_2019  34.0239     0.9080   37.47   <2e-16 ***
## LocNeoga_2018      14.7687     0.8755   16.87   <2e-16 ***
## LocPerry_2019      14.4606     0.9080   15.93   <2e-16 ***
## LocUrbana_2018     19.4130     0.9504   20.43   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.397 on 410 degrees of freedom
## Multiple R-squared:  0.795,  Adjusted R-squared:  0.7925
## F-statistic: 318.1 on 5 and 410 DF,  p-value: < 2.2e-16
```

```r
#Coefficients:
#                   Estimate Std. Error t value Pr(>|t|)
#(Intercept)        76.7404     0.6190  123.97   <2e-16 ***
#LocElkville_2019   24.4245     0.8755   27.90   <2e-16 ***
```

```
#LocHampshire_2019  34.0239     0.9080   37.47   <2e-16 ***
#LocNeoga_2018      14.7687     0.8755   16.87   <2e-16 ***
#LocPerry_2019      14.4606     0.9080   15.93   <2e-16 ***
#LocUrbana_2018     19.4130     0.9504   20.43   <2e-16 ***
```

Hampshire_2019 is the highest yield location.

Obj 2. Evaluate company/ brand effects:

a. Which company's varieties seem to perform the best across all regions?

b. Which company's varieties seem to perform the worst across all regions?

```
model2 <- lm(Estimate~Company, df_clean)
anova(model2)
```

```
## Analysis of Variance Table
##
## Response: Estimate
##            Df Sum Sq Mean Sq F value Pr(>F)
## Company    21   4988  237.51  1.7565 0.0213 *
## Residuals 394  53274  135.21
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(model2)
```

```
##
## Call:
## lm(formula = Estimate ~ Company, data = df_clean)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -30.4358  -7.6871   0.1811   7.5874  29.1441
##
## Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)                      94.3004     1.4206  66.381  < 2e-16 ***
## CompanyAgriPro                    1.4642     2.9629   0.494  0.62145
## CompanyBecks_Hybrids             -2.1933    11.7146  -0.187  0.85158
## CompanyBioTown_Seeds             -4.4905     3.3215  -1.352  0.17716
## CompanyCroplan                   -1.6176     2.9629  -0.546  0.58541
## CompanyDeRaedt_Seed               7.5558     3.9420   1.917  0.05599 .
## CompanyDyna-Gro                  -1.4364     2.2564  -0.637  0.52476
## CompanyGo_Wheat                  -9.4128     4.9552  -1.900  0.05822 .
## CompanyGreen_Valley               1.2884     5.3908   0.239  0.81123
## CompanyGROWMARK                   0.7204     2.1989   0.328  0.74336
## CompanyHoffman_Seed              -8.0885     2.9629  -2.730  0.00662 **
## CompanyKitchen_Seed_Company       0.1652     3.0871   0.054  0.95735
```

```
## CompanyKratz_Farms                    1.0786     3.2356  0.333  0.73905
## CompanyKWS_Cereals                    2.2053     3.6450  0.605  0.54551
## CompanyLEWIS                          2.3353     2.6867  0.869  0.38527
## CompanyLimagrain                     -1.0188     3.2356 -0.315  0.75302
## CompanyMiller_Bros_Farm_and_Fert.    -8.6036    11.7146 -0.734  0.46312
## CompanyMoiner_Seed                    9.1233     8.3441  1.093  0.27489
## CompanyPioneer                        4.9772     2.2886  2.175  0.03024 *
## CompanyProHarvest                    -6.7280     4.3497 -1.547  0.12272
## CompanyUSG                           -3.0548     3.1578 -0.967  0.33395
## CompanyVCIA                          -2.8520     8.3441 -0.342  0.73269
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.63 on 394 degrees of freedom
## Multiple R-squared:  0.08561,    Adjusted R-squared:  0.03687
## F-statistic: 1.757 on 21 and 394 DF,  p-value: 0.0213
```

[Best] Moiner_Seed variaty perform the best across all regions, but the s.e. for this company is large and thus this company doesn't have a significant effect on the yield, thus DeRaedt_Seed can be a good choice [Worst] Go_Wheat variaty perform the worst across all regions

c. Which company's varieties seem to perform the best for each location?

d. Which company's varieties seem to perform the worst for each location?

```r
as.data.frame(aggregate(x=df_clean$Estimate,
          by=list(df_clean$Loc,df_clean$Company),
          FUN=mean)) -> ag2
```

[perform the best] * Hampshire_2019 Pioneer * Elkville_2019 KWS_Cereals * Urbana_2018 Pioneer * Perry_2019 Pioneer * Neoga_2018 Croplan * Belleville_2019 Pioneer

[perfrom the worst] * Belleville_2019 ProHarvest * Neoga_2018 Miller_Bros_Farm_and_Fert. * Elkville_2019 ProHarvest * Perry_2019 ProHarvest * Urbana_2018 Kratz_Farms * Hampshire_2019 Kratz_Farms

Obj 3. Evaluate variety effects:

a. Which varieties seem to perform the best across all regions?

b. Which varieties seem to perform the worst across all regions?

```r
aggregate(x=df_clean$Estimate,
          by=list(df_clean$Variety),
          FUN=mean) %>%
  as.data.frame() -> ag3.1
```

Variety DeRaedt_24 performs the best across all regions Variety EXP18-1 performs the worst across all regions

c. Which varieties seem to perform the best for each location?

d. Which varieties seem to perform the worst for each location?

```
aggregate(x=df_clean$Estimate,
          by=list(df_clean$Variety, df_clean$Loc),
          FUN=mean) %>%
  as.data.frame() -> ag3.2
```

[varieties perform the best for each location] * WX18C at Urbana_2018 with 108.3 bu/acre * 495 at Perry_2019 with 103.7 bu/acre * H7W18 at Neoga_2018 with 105.2 bu/acre * KWS19X03 at Hampshire_2019 with 125.6 bu/acre * KWS19X09 at Elkville_2019 with 113.5 bu/acre * KWS19X07 at Belleville_2019 with 93.4 bu/acre

[varieties perform the worst for each location] * 851 and Lewis_851 at Urbana_2018 with 87.6 bu/acre * KF_15334 at Perry_2019 with 79.4 bu/acre * FS_604 at Neoga_2018 with 79.4 bu/acre * KF_15334 at Hampshire_2019 with 97.0 bu/acre * 286 at Elkville_2019 with 83.3 bu/acre * 317 at Belleville_2019 with 60.3 bu/acre

Obj 4. Evaluate regional effects(north/south) :

a. How much variation in yield does region explain alone?

b. How about together with Company?

```
model.4.1 <- lm(Estimate ~ Reg, data = df_clean)
summary(model.4.1)$adj.r.squared # 0.1623734
```

```
## [1] 0.1623734
```

```
anova(model.4.1)
```

```
## Analysis of Variance Table
##
## Response: Estimate
##            Df Sum Sq Mean Sq F value    Pr(>F)
## Reg         1   9578  9577.7  81.448 < 2.2e-16 ***
## Residuals 414  48684   117.6
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
model.4.2 <- lm(Estimate ~ Reg + Company, data = df_clean)
anova(model.4.2)
```

```
## Analysis of Variance Table
##
## Response: Estimate
##            Df Sum Sq Mean Sq F value  Pr(>F)
```

```
## Reg          1   9578  9577.7 84.3367 < 2e-16 ***
## Company     21   4053   193.0  1.6993 0.02842 *
## Residuals 393  44631   113.6
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(model.4.2)$adj.r.squared # 0.1910692
```

```
## [1] 0.1910692
```

- Region alone explain 16.24% of variation in yield.
- Together with comapany, 19.1% of variation in yield was explained.

c. Which company's varieties seem to perform the best within each region?

d. Which company's varieties seem to perform the worst within each region?

```r
library(dplyr)
as.data.frame(aggregate(x=df_clean$Estimate,
          by=list(df_clean$Reg,df$Company),
          FUN=mean)) -> ag4.1
```

[company that perform the best within each region] * company DeRaedt_Seed at south region * company Moiner_Seed at north region

[company that perform the worst within each region] * comapny ProHarvest at south region * Miller_Bros_Farm_and_Fert at north region

e. Which variety is best suited to each region?

```r
as.data.frame(aggregate(x=df_clean$Estimate,
          by=list(df_clean$Reg,df$Variety),
          FUN=mean)) -> ag4.2
```

- variety KWS19X03 best suited to north region
- variety CP9606 best suited to south region

f. Which variety is best suited to each location?

```r
as.data.frame(aggregate(x=df_clean$Estimate,
          by=list(df_clean$Loc,df$Variety),
          FUN=mean)) -> ag4.3
```

- Belleville_2019: variety KWS19X07

- Elkville_2019: variety KWS19X09

- Hampshire_2019: variety KWS19X03
- Neoga_2018: variety H7W18

- Perry_2019: variety 495
- Urbana_2018: variety WX18C

Obj 5. Does the seed treatments have a significant effect on the yield?

a. Which treatment seems to have the largest positive effect? Is it significant?

```
model5.1 <- lm(Estimate ~ SeedTreatment, df_clean)
anova(model5.1)
```

```
## Analysis of Variance Table
##
## Response: Estimate
##                Df Sum Sq Mean Sq F value  Pr(>F)
## SeedTreatment   3    915  305.12  2.1922 0.08839 .
## Residuals     412  57346  139.19
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(model5.1)
```

```
##
## Call:
## lm(formula = Estimate ~ SeedTreatment, data = df_clean)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -32.457  -7.116   0.047   8.748  30.549
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)      89.150      2.781  32.059   <2e-16 ***
## SeedTreatmentC    5.951      2.868   2.075   0.0386 *
## SeedTreatmentE    2.957     12.121   0.244   0.8074
## SeedTreatmentG    3.653      2.992   1.221   0.2228
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.8 on 412 degrees of freedom
## Multiple R-squared:  0.01571,    Adjusted R-squared:  0.008544
## F-statistic: 2.192 on 3 and 412 DF,  p-value: 0.08839
```

Seed treatment itself have a signficant positive effect. Treatment C is the largest with signigicant effect.

b. What fraction of the variation observed in yield is attributable to seed treatments?

```r
summary(model5.1)$adj.r.squared
```

```
## [1] 0.008544335
```

0.85% of the variarion observed in yield is attributable to seed treatment.

c. Is seed treatment recommended?

```r
model5.2 <- lm(Estimate ~ Loc + SeedTreatment, df_clean)
anova(model5.2)
```

```
## Analysis of Variance Table
##
## Response: Estimate
##                Df Sum Sq Mean Sq  F value Pr(>F)
## Loc             5  46321  9264.2 318.4971 <2e-16 ***
## SeedTreatment   3    102    34.1   1.1723   0.32
## Residuals     407  11838    29.1
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(model5.2)
```

```
##
## Call:
## lm(formula = Estimate ~ Loc + SeedTreatment, data = df_clean)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -17.8614  -3.3068   0.0064   3.7516  16.4488
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)        74.5229     1.3712  54.350   <2e-16 ***
## LocElkville_2019   24.4245     0.8749  27.917   <2e-16 ***
## LocHampshire_2019  33.8568     0.9161  36.957   <2e-16 ***
## LocNeoga_2018      14.7535     0.8781  16.802   <2e-16 ***
## LocPerry_2019      14.2936     0.9161  15.602   <2e-16 ***
## LocUrbana_2018     19.2478     0.9591  20.068   <2e-16 ***
## SeedTreatmentC      2.4741     1.3256   1.866   0.0627 .
## SeedTreatmentE      2.8306     5.5650   0.509   0.6113
## SeedTreatmentG      2.2102     1.3704   1.613   0.1076
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.393 on 407 degrees of freedom
## Multiple R-squared:  0.7968, Adjusted R-squared:  0.7928
## F-statistic: 199.5 on 8 and 407 DF,  p-value: < 2.2e-16
```

seed treatment is not signifiant any more when look together with other factors, thus there is no need to apply seed treatment

Obj 6. Give your best prediction of maximum yield under these best case scenario conditions.

a. Which location variety combinations should be used to get the maximum yield?

```r
# split the data
df_south <- df_clean[df_clean$Reg=="south",]
df_north <- df_clean[df_clean$Reg=="north",]

# find the best model
fit_int <- lm(Estimate ~ 1, data=df_clean)
fit_full <- lm(Estimate ~ Company + Loc + Reg + SeedTreatment + Variety, data=df_clean)
step(fit_int,scope = list(upper = formula(fit_full), lower = formula(fit_int)), direction = 'both')
```

```
## Start:  AIC=2057.88
## Estimate ~ 1
##
##                  Df Sum of Sq   RSS    AIC
## + Loc             5     46321 11941 1408.5
## + Reg             1      9578 48684 1985.2
## + SeedTreatment   3       915 57346 2057.3
## <none>                        58262 2057.9
## + Company        21      4988 53274 2062.7
## + Variety       130     10877 47385 2231.9
##
## Step:  AIC=1408.52
## Estimate ~ Loc
##
##                  Df Sum of Sq   RSS    AIC
## + Variety       130      6417  5523 1347.8
## + Company        21      2385  9556 1357.8
## <none>                        11941 1408.5
## + SeedTreatment   3       102 11838 1410.9
## - Loc             5     46321 58262 2057.9
##
## Step:  AIC=1347.81
## Estimate ~ Loc + Variety
##
##                  Df Sum of Sq   RSS    AIC
## <none>                         5523 1347.8
## + SeedTreatment   2         0  5523 1351.8
## - Variety       130      6417 11941 1408.5
## - Loc             5     41861 47385 2231.9
```

```
##
## Call:
```

```
## lm(formula = Estimate ~ Loc + Variety, data = df_clean)
##
## Coefficients:
##         (Intercept)      LocElkville_2019     LocHampshire_2019
##             74.4087               24.4245               34.2028
##       LocNeoga_2018         LocPerry_2019        LocUrbana_2018
##             14.0123               14.6396               19.2457
##         Variety25R25          Variety25R40          Variety25R61
##              6.3953                4.4953                9.0188
##         Variety25R74          Variety25R77            Variety286
##              6.2438                3.4457               -4.9781
##          Variety317           Variety3197           Variety3228
##             -5.3281               -1.2543                0.0258
##         Variety3329           Variety3404           Variety3448
##              4.7357                5.5346                8.7258
##         Variety3536            Variety413            Variety438
##             -1.5500                5.0953                2.4508
##          Variety444            Variety446            Variety454
##              2.7383                6.4124                3.9457
##          Variety463            Variety473            Variety475
##              1.8094                2.9723                4.2994
##          Variety480            Variety485            Variety486
##             -0.1139               -3.3516                3.2984
##          Variety495            Variety628            Variety658
##              6.5219               -3.3257                1.5743
##          Variety65X            Variety668            Variety66X
##              4.5743                5.3743                1.4743
##          Variety828            Variety829            Variety833
##             -1.0714                5.4438               -3.6181
##          Variety839            Variety851           Variety9522
##              2.3397               -2.0723                7.7124
##         Variety9552           Variety9701           Variety9750
##              6.2122                3.4410                0.2306
##         Variety9811           Variety9862           Variety9932
##              0.1861               -2.3784                5.7219
##         Variety9941           Variety9980       VarietyBeck_726
##              2.4469               -2.6000                3.6861
##       VarietyCP8550         VarietyCP8800         VarietyCP9415
##              1.8557                1.4000                5.5835
##       VarietyCP9606          VarietyD496W          VarietyD497W
##             10.5002               -3.8876                3.7190
##        VarietyD498W          VarietyD510W     VarietyDeRaedt_11
##              4.5861                0.2724                3.2397
##    VarietyDeRaedt_17     VarietyDeRaedt_24    VarietyDiener_497W
##              2.7938                7.7938                3.7190
##  VarietyDiener_D510W       VarietyExp_1884       VarietyExp_1892
##              0.2724                1.0802                6.7302
##      VarietyExp_1899       VarietyExp_1902       VarietyExp_1905
##             12.8743                1.7969               -1.8281
##      VarietyExp_1906       VarietyExp_1913        VarietyEXP18-1
##             -3.0000               -0.6500               -6.8000
##       VarietyEXP18-2         VarietyFS_599         VarietyFS_601
##              1.6000                0.8469               -2.0781
##        VarietyFS_603         VarietyFS_604         VarietyFS_615
```

```
##                4.6151               -1.5629                7.0983
##         VarietyFS_619         VarietyFS_624          VarietyH7W15
##                0.2956                3.0094                1.9568
##          VarietyH7W16          VarietyH7W17          VarietyH7W18
##               -4.1498               -6.1242               16.7861
##          VarietyH7W28       VarietyHilliard        VarietyKF_15144
##               -3.3476                5.0790                0.5743
##        VarietyKF_15241        VarietyKF_15334        VarietyKF_15639
##                0.1438               -8.9556                2.9277
##          VarietyKF_553          VarietyKF_667          VarietyKF_727
##               -6.3062               -0.3062               -1.3556
##         VarietyKSC_416         VarietyKSC_417         VarietyKSC_418
##               -0.3729                3.2494                4.0317
##       VarietyKWS19X03       VarietyKWS19X07       VarietyKWS19X09
##                1.7969                5.6719                3.8719
##         VarietyL11548         VarietyL11549         VarietyL11617
##                7.9024                3.5861               -3.3139
##         VarietyL11713         VarietyL11719           VarietyL214
##               -6.1500                1.9246               -2.7242
##      VarietyLewis_828      VarietyLewis_833      VarietyLewis_839
##               -1.0714               -3.6181                2.3397
##      VarietyLewis_851          VarietyMW857 VarietyPioneer_25R25
##               -2.0723                4.5938                6.3953
## VarietyPioneer_25R40 VarietyPioneer_25R61 VarietyPioneer_25R74
##                4.4953                9.0188                6.2438
## VarietyPioneer_25R77        VarietySRW_8550        VarietySRW_9415
##                3.4457                1.8557                5.5835
##        VarietySRW_9606         VarietySY_100          VarietySY_547
##               10.5002                3.2616                0.4983
##          VarietySY_576        VarietySY_Viper       VarietyVA12W-68
##                6.1219                5.1219                0.9758
##         VarietyWX17775         VarietyWX17778         VarietyWX18416
##                3.0802                6.1861                4.6000
##         VarietyWX18724           VarietyWX18A           VarietyWX18B
##                2.6302               -0.7698                5.8302
##           VarietyWX18C         VarietyWX19711         VarietyWX19713
##                7.4802                2.2438                3.9500
##         VarietyWX19714           VarietyWX19A           VarietyWX19B
##               -3.7500                7.5969                1.7719
##         VarietyXW_1802
##               10.5861
```

```r
# fit the best model for each region seperately
## south
model.6.south <- lm(Estimate ~ Loc + Variety, data=df_south)
summary(model.6.south)
```

```
##
## Call:
## lm(formula = Estimate ~ Loc + Variety, data = df_south)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -11.086  -1.625   0.000   1.588  10.922
```

```
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)           74.4087     3.1356  23.731  < 2e-16 ***
## LocElkville_2019      24.4245     0.7147  34.177  < 2e-16 ***
## LocPerry_2019         14.3431     0.9563  14.999  < 2e-16 ***
## Variety25R25           5.1815     5.4680   0.948  0.34525
## Variety25R40           4.5482     5.4680   0.832  0.40720
## Variety25R61           9.8465     4.0325   2.442  0.01608 *
## Variety25R74           5.6577     4.0325   1.403  0.16321
## Variety25R77           4.8790     4.4055   1.108  0.27031
## Variety286            -4.4514     4.0325  -1.104  0.27187
## Variety317            -8.8180     4.0325  -2.187  0.03072 *
## Variety3329            4.6605     3.8152   1.222  0.22429
## Variety3404            4.6457     3.8152   1.218  0.22576
## Variety3536           -1.5500     4.4055  -0.352  0.72558
## Variety413             6.4149     5.4680   1.173  0.24307
## Variety438             2.9815     5.4680   0.545  0.58659
## Variety444             2.0458     4.0325   0.507  0.61287
## Variety454             4.8624     4.4055   1.104  0.27194
## Variety463             1.1910     4.0325   0.295  0.76824
## Variety473             2.6610     4.0325   0.660  0.51059
## Variety475             3.8910     4.0325   0.965  0.33654
## Variety485            -4.0569     4.0325  -1.006  0.31643
## Variety486             6.0431     4.0325   1.499  0.13662
## Variety495             7.8820     4.0325   1.955  0.05297 .
## Variety828            -6.0518     5.4680  -1.107  0.27062
## Variety829             4.1459     5.4680   0.758  0.44983
## Variety833            -5.0185     5.4680  -0.918  0.36058
## Variety839             2.2149     5.4680   0.405  0.68616
## Variety851             4.2084     5.4680   0.770  0.44304
## Variety9522            7.3790     4.4055   1.675  0.09656 .
## Variety9701            4.0465     4.0325   1.003  0.31766
## Variety9750            1.6290     4.4055   0.370  0.71220
## Variety9862           -3.3090     4.0325  -0.821  0.41352
## Variety9932            5.6820     4.0325   1.409  0.16143
## Variety9941            2.6820     4.0325   0.665  0.50728
## Variety9980           -2.6000     4.4055  -0.590  0.55619
## VarietyCP8550         -1.8043     4.4055  -0.410  0.68287
## VarietyCP8800          1.4000     4.4055   0.318  0.75120
## VarietyCP9415          6.4624     4.4055   1.467  0.14504
## VarietyCP9606         13.2624     4.4055   3.010  0.00319 **
## VarietyD497W           2.4355     4.4055   0.553  0.58141
## VarietyD510W           0.1957     4.4055   0.044  0.96464
## VarietyDeRaedt_11      7.2815     4.4939   1.620  0.10781
## VarietyDeRaedt_17      0.7459     5.4680   0.136  0.89173
## VarietyDeRaedt_24      6.3459     5.4680   1.161  0.24815
## VarietyDiener_497W     2.4355     4.4055   0.553  0.58141
## VarietyDiener_D510W    0.1957     4.4055   0.044  0.96464
## VarietyExp_1902        2.9820     4.0325   0.739  0.46107
## VarietyExp_1905       -3.9514     4.0325  -0.980  0.32913
## VarietyExp_1906       -3.0000     4.4055  -0.681  0.49721
## VarietyExp_1913       -0.6500     4.4055  -0.148  0.88295
## VarietyEXP18-1        -6.8000     4.4055  -1.544  0.12535
```

```
## VarietyEXP18-2         1.6000     4.4055   0.363  0.71711
## VarietyFS_599          0.6820     4.0325   0.169  0.86599
## VarietyFS_601         -0.3847     4.0325  -0.095  0.92415
## VarietyFS_603          6.0431     4.0325   1.499  0.13662
## VarietyFS_604         -1.4979     4.0325  -0.371  0.71095
## VarietyFS_615          8.4243     4.0325   2.089  0.03883 *
## VarietyFS_624          4.2021     4.0325   1.042  0.29949
## VarietyH7W15           2.5290     3.8152   0.663  0.50869
## VarietyH7W16          -5.8376     3.8152  -1.530  0.12865
## VarietyH7W28          -3.0645     3.8152  -0.803  0.42345
## VarietyKF_15241        1.8459     5.4680   0.338  0.73628
## VarietyKF_15334       -9.3916     5.4680  -1.718  0.08848 .
## VarietyKF_15639        4.9584     5.4680   0.907  0.36635
## VarietyKF_553         -2.4541     5.4680  -0.449  0.65438
## VarietyKF_667         -3.0541     5.4680  -0.559  0.57752
## VarietyKF_727          1.8584     5.4680   0.340  0.73456
## VarietyKSC_416         0.2243     4.0325   0.056  0.95573
## VarietyKSC_417         3.3910     4.0325   0.841  0.40208
## VarietyKSC_418         4.5098     4.0325   1.118  0.26566
## VarietyKWS19X03       -3.1847     4.0325  -0.790  0.43123
## VarietyKWS19X07        8.1153     4.0325   2.012  0.04643 *
## VarietyKWS19X09        7.5153     4.0325   1.864  0.06483 .
## VarietyL11548          6.6605     3.8152   1.746  0.08343 .
## VarietyL11713         -6.1500     4.4055  -1.396  0.16531
## VarietyL11719          3.8459     4.4939   0.856  0.39383
## VarietyLewis_828      -6.0518     5.4680  -1.107  0.27062
## VarietyLewis_833      -5.0185     5.4680  -0.918  0.36058
## VarietyLewis_839       2.2149     5.4680   0.405  0.68616
## VarietyLewis_851       4.2084     5.4680   0.770  0.44304
## VarietyMW857           4.1459     5.4680   0.758  0.44983
## VarietyPioneer_25R25   5.1815     5.4680   0.948  0.34525
## VarietyPioneer_25R40   4.5482     5.4680   0.832  0.40720
## VarietyPioneer_25R61   9.8465     4.0325   2.442  0.01608 *
## VarietyPioneer_25R74   5.6577     4.0325   1.403  0.16321
## VarietyPioneer_25R77   4.8790     4.4055   1.108  0.27031
## VarietySRW_8550       -1.8043     4.4055  -0.410  0.68287
## VarietySRW_9415        6.4624     4.4055   1.467  0.14504
## VarietySRW_9606       13.2624     4.4055   3.010  0.00319 **
## VarietySY_100          2.2577     4.0325   0.560  0.57662
## VarietySY_547         -2.2757     4.0325  -0.564  0.57358
## VarietySY_576          5.7153     4.0325   1.417  0.15900
## VarietySY_Viper        4.8486     4.0325   1.202  0.23160
## VarietyWX18416         4.6000     4.4055   1.044  0.29853
## VarietyWX19711         4.4459     5.4680   0.813  0.41780
## VarietyWX19713         3.9500     4.4055   0.897  0.37173
## VarietyWX19714        -3.7500     4.4055  -0.851  0.39636
## VarietyWX19A           9.4153     4.0325   2.335  0.02123 *
## VarietyWX19B           2.7153     4.0325   0.673  0.50203
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.405 on 119 degrees of freedom
## Multiple R-squared:  0.9226, Adjusted R-squared:  0.8588
## F-statistic: 14.47 on 98 and 119 DF,  p-value: < 2.2e-16
```

```r
best.v.l_1 <- data.frame(
   Loc = "Elkville_2019",Variety="CP9606"
)
best.v.l_2<- data.frame(
   Loc = "Elkville_2019",Variety="SRW_9606"
)
predict(model.6.south, best.v.l_1)
```

```
##        1
## 112.0956
```

```r
# 112.0956
predict(model.6.south, best.v.l_2)
```

```
##        1
## 112.0956
```

```r
# 112.0956

## north
model.6.north <- lm(Estimate ~ Loc + Variety, data=df_north)
summary(model.6.north)
```

```
##
## Call:
## lm(formula = Estimate ~ Loc + Variety, data = df_north)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.3249 -0.9369  0.0000  0.7006 10.3303
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)           1.158e+02  3.102e+00  37.314  < 2e-16 ***
## LocNeoga_2018        -2.064e+01  1.128e+00 -18.293  < 2e-16 ***
## LocUrbana_2018       -1.496e+01  9.488e-01 -15.768  < 2e-16 ***
## Variety25R40         -2.533e+00  4.336e+00  -0.584 0.560745
## Variety25R61          1.288e+00  3.972e+00   0.324 0.746623
## Variety25R74         -7.323e-02  3.972e+00  -0.018 0.985337
## Variety25R77         -6.126e+00  5.402e+00  -1.134 0.260249
## Variety286           -1.341e+01  5.331e+00  -2.516 0.013953 *
## Variety317           -1.714e+00  5.331e+00  -0.321 0.748761
## Variety3197          -7.960e+00  5.402e+00  -1.474 0.144678
## Variety3228          -6.679e+00  5.402e+00  -1.237 0.220005
## Variety3329          -1.819e+00  4.447e+00  -0.409 0.683628
## Variety3404           6.071e-01  4.447e+00   0.137 0.891777
## Variety3448           2.021e+00  5.402e+00   0.374 0.709389
## Variety413           -2.567e+00  4.336e+00  -0.592 0.555611
## Variety438           -4.817e+00  4.336e+00  -1.111 0.270074
## Variety444           -3.003e+00  4.368e+00  -0.688 0.493832
## Variety446           -2.929e-01  5.402e+00  -0.054 0.956893
## Variety454           -4.593e+00  5.402e+00  -0.850 0.397802
```

```
## Variety463           -4.475e+00  3.972e+00  -1.127 0.263326
## Variety473           -3.620e+00  3.972e+00  -0.911 0.364966
## Variety475           -2.195e+00  3.972e+00  -0.553 0.582028
## Variety480           -6.819e+00  5.402e+00  -1.262 0.210602
## Variety485           -9.550e+00  3.972e+00  -2.404 0.018605 *
## Variety486           -6.350e+00  3.972e+00  -1.599 0.113993
## Variety495           -4.414e+00  5.331e+00  -0.828 0.410315
## Variety628           -1.047e+01  5.331e+00  -1.965 0.053064 .
## Variety658           -5.574e+00  5.331e+00  -1.046 0.299031
## Variety65X           -2.574e+00  5.331e+00  -0.483 0.630554
## Variety668           -1.774e+00  5.331e+00  -0.333 0.740172
## Variety66X           -5.674e+00  5.331e+00  -1.064 0.290507
## Variety828           -5.583e+00  4.336e+00  -1.288 0.201705
## Variety829           -1.137e-01  5.331e+00  -0.021 0.983046
## Variety833           -9.920e+00  4.336e+00  -2.288 0.024887 *
## Variety839           -4.600e+00  4.336e+00  -1.061 0.292044
## Variety851           -1.221e+01  4.336e+00  -2.817 0.006153 **
## Variety9522           1.674e+00  5.402e+00   0.310 0.757506
## Variety9552          -9.366e-01  5.331e+00  -0.176 0.861019
## Variety9701          -4.068e+00  3.972e+00  -1.024 0.308975
## Variety9750          -8.095e+00  4.372e+00  -1.851 0.067956 .
## Variety9811          -6.519e+00  5.402e+00  -1.207 0.231167
## Variety9862          -8.351e+00  3.972e+00  -2.103 0.038770 *
## Variety9932          -1.014e+00  5.331e+00  -0.190 0.849709
## Variety9941          -5.114e+00  5.331e+00  -0.959 0.340488
## VarietyBeck_726      -3.019e+00  5.402e+00  -0.559 0.577823
## VarietyCP8550         2.471e+00  5.402e+00   0.457 0.648696
## VarietyCP9415        -2.879e+00  5.402e+00  -0.533 0.595515
## VarietyCP9606        -1.729e+00  5.402e+00  -0.320 0.749701
## VarietyD496W         -1.059e+01  5.402e+00  -1.961 0.053487 .
## VarietyD497W         -4.192e-01  5.402e+00  -0.078 0.938343
## VarietyD498W         -2.119e+00  5.402e+00  -0.392 0.695901
## VarietyD510W         -6.279e+00  5.402e+00  -1.163 0.248616
## VarietyDeRaedt_11    -5.783e+00  3.755e+00  -1.540 0.127617
## VarietyDeRaedt_17    -2.014e+00  5.331e+00  -0.378 0.706698
## VarietyDeRaedt_24     2.386e+00  5.331e+00   0.448 0.655703
## VarietyDiener_497W   -4.192e-01  5.402e+00  -0.078 0.938343
## VarietyDiener_D510W  -6.279e+00  5.402e+00  -1.163 0.248616
## VarietyExp_1884      -5.847e+00  4.372e+00  -1.337 0.185092
## VarietyExp_1892      -1.968e-01  4.372e+00  -0.045 0.964213
## VarietyExp_1899       5.726e+00  5.331e+00   1.074 0.286215
## VarietyExp_1902      -8.614e+00  5.331e+00  -1.616 0.110267
## VarietyExp_1905      -2.314e+00  5.331e+00  -0.434 0.665530
## VarietyFS_599        -5.514e+00  5.331e+00  -1.034 0.304294
## VarietyFS_601        -1.401e+01  5.331e+00  -2.628 0.010349 *
## VarietyFS_603        -3.716e+00  3.972e+00  -0.936 0.352382
## VarietyFS_604        -8.531e+00  3.972e+00  -2.148 0.034868 *
## VarietyFS_615        -1.131e+00  3.972e+00  -0.285 0.776599
## VarietyFS_619        -6.631e+00  4.372e+00  -1.517 0.133451
## VarietyFS_624        -5.087e+00  3.972e+00  -1.281 0.204156
## VarietyH7W15         -5.893e+00  4.447e+00  -1.325 0.189061
## VarietyH7W16         -7.479e+00  4.447e+00  -1.682 0.096652 .
## VarietyH7W17         -1.283e+01  5.402e+00  -2.375 0.020032 *
## VarietyH7W18          1.008e+01  5.402e+00   1.866 0.065813 .
```

```
## VarietyH7W28         -1.062e+01  4.447e+00  -2.388 0.019399 *
## VarietyHilliard      -1.626e+00  5.402e+00  -0.301 0.764173
## VarietyKF_15144      -6.574e+00  5.331e+00  -1.233 0.221277
## VarietyKF_15241      -8.414e+00  5.331e+00  -1.578 0.118640
## VarietyKF_15334      -1.574e+01  4.336e+00  -3.630 0.000508 ***
## VarietyKF_15639      -5.090e+00  4.336e+00  -1.174 0.244072
## VarietyKF_553        -1.701e+01  5.331e+00  -3.191 0.002051 **
## VarietyKF_667        -4.414e+00  5.331e+00  -0.828 0.410315
## VarietyKF_727        -9.965e+00  4.336e+00  -2.298 0.024266 *
## VarietyKSC_416       -7.873e+00  3.972e+00  -1.982 0.051015 .
## VarietyKSC_417       -3.795e+00  3.972e+00  -0.956 0.342264
## VarietyKSC_418       -3.350e+00  3.972e+00  -0.843 0.401658
## VarietyKWS19X03       9.886e+00  5.331e+00   1.854 0.067521 .
## VarietyKWS19X07      -8.514e+00  5.331e+00  -1.597 0.114392
## VarietyKWS19X09      -1.391e+01  5.331e+00  -2.610 0.010884 *
## VarietyL11548         3.681e+00  4.447e+00   0.828 0.410417
## VarietyL11549        -3.119e+00  5.402e+00  -0.577 0.565317
## VarietyL11617        -1.002e+01  5.402e+00  -1.855 0.067446 .
## VarietyL11719        -5.816e+00  3.793e+00  -1.534 0.129211
## VarietyL214          -9.429e+00  5.402e+00  -1.746 0.084856 .
## VarietyLewis_828     -5.583e+00  4.336e+00  -1.288 0.201705
## VarietyLewis_833     -9.920e+00  4.336e+00  -2.288 0.024887 *
## VarietyLewis_839     -4.600e+00  4.336e+00  -1.061 0.292044
## VarietyLewis_851     -1.221e+01  4.336e+00  -2.817 0.006153 **
## VarietyMW857         -1.814e+00  5.331e+00  -0.340 0.734648
## VarietyPioneer_25R25 -7.033e-14  4.336e+00   0.000 1.000000
## VarietyPioneer_25R40 -2.533e+00  4.336e+00  -0.584 0.560745
## VarietyPioneer_25R61  1.288e+00  3.972e+00   0.324 0.746623
## VarietyPioneer_25R74 -7.323e-02  3.972e+00  -0.018 0.985337
## VarietyPioneer_25R77 -6.126e+00  5.402e+00  -1.134 0.260249
## VarietySRW_8550       2.471e+00  5.402e+00   0.457 0.648696
## VarietySRW_9415      -2.879e+00  5.402e+00  -0.533 0.595515
## VarietySRW_9606      -1.729e+00  5.402e+00  -0.320 0.749701
## VarietySY_100        -2.638e+00  3.972e+00  -0.664 0.508620
## VarietySY_547        -3.631e+00  3.972e+00  -0.914 0.363471
## VarietySY_576         4.863e-01  5.331e+00   0.091 0.927555
## VarietySY_Viper      -9.137e-01  5.331e+00  -0.171 0.864382
## VarietyVA12W-68      -5.729e+00  5.402e+00  -1.061 0.292144
## VarietyWX17775       -3.847e+00  4.372e+00  -0.880 0.381709
## VarietyWX17778       -5.192e-01  5.402e+00  -0.096 0.923676
## VarietyWX18724       -4.297e+00  4.372e+00  -0.983 0.328827
## VarietyWX18A         -7.697e+00  4.372e+00  -1.760 0.082325 .
## VarietyWX18B         -1.097e+00  4.372e+00  -0.251 0.802598
## VarietyWX18C          5.532e-01  4.372e+00   0.127 0.899653
## VarietyWX19711       -6.814e+00  5.331e+00  -1.278 0.205087
## VarietyWX19A         -4.714e+00  5.331e+00  -0.884 0.379385
## VarietyWX19B         -7.914e+00  5.331e+00  -1.484 0.141805
## VarietyXW_1802        3.881e+00  5.402e+00   0.718 0.474653
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.336 on 77 degrees of freedom
## Multiple R-squared:  0.9232, Adjusted R-squared:  0.8035
## F-statistic: 7.714 on 120 and 77 DF,  p-value: < 2.2e-16
```

```
# Hampshire_2019    VarietyKWS19X03
best.v.l_3 <- data.frame(
   Loc = "Hampshire_2019",Variety="KWS19X03"
)
predict(model.6.north, best.v.l_3)
```

```
##        1
## 125.6498
```

```
# 125.6498
```

At location Hampshire_2019 use variety "KWS19X03" can get the best yield prediction

b. Is paying a premium for Seed Treatment justified?

```
model.6.south <- lm(Estimate ~ Loc + Variety, data=df_south)
model.6.south_sd <- lm(Estimate ~ Loc + Variety + SeedTreatment, data=df_south)
anova(model.6.south, model.6.south_sd)
```

```
## Analysis of Variance Table
##
## Model 1: Estimate ~ Loc + Variety
## Model 2: Estimate ~ Loc + Variety + SeedTreatment
##   Res.Df    RSS Df Sum of Sq  F Pr(>F)
## 1    119 2309.6
## 2    117 2309.6  2         0  0      1
```

```
model.6.north <- lm(Estimate ~ Loc + Variety, data=df_north)
model.6.north_sd <- lm(Estimate ~ Loc + Variety + SeedTreatment, data=df_north)
anova(model.6.north, model.6.north_sd)
```

```
## Analysis of Variance Table
##
## Model 1: Estimate ~ Loc + Variety
## Model 2: Estimate ~ Loc + Variety + SeedTreatment
##   Res.Df    RSS Df Sum of Sq  F Pr(>F)
## 1     77 1447.6
## 2     75 1447.6  2         0  0      1
```

No differnce in terms of the efficiency of the model for both south and north region, no need to use seed treatment

c. Report the best prediction of maximum yield under these best case scenario conditions for each variety for each region and location.

```r
# south region (Perry_2019; Belleville_2019; Elkville_2019)

# Elkville_2019
best.v.l_1 <- data.frame(
   Loc = "Elkville_2019",Variety="CP9606"
)
best.v.l_2<- data.frame(
   Loc = "Elkville_2019",Variety="KWS19X09"
)
predict(model.6.south, best.v.l_1)# 112.0956  # selected by summary
```

```
##        1
## 112.0956
```

```r
predict(model.6.south, best.v.l_2)# 106.3485 selected by mean
```

```
##        1
## 106.3485
```

```r
# Perry_2019
best.v.l_4 <- data.frame(
   Loc = "Perry_2019",Variety="CP9606"
)
best.v.l_5<- data.frame(
   Loc = "Perry_2019",Variety="25R61"
)
best.v.l_5.1<- data.frame(
   Loc = "Perry_2019",Variety="495"
)
predict(model.6.south, best.v.l_4) # 102.0142(not show up)  # selected by summary
```

```
##        1
## 102.0142
```

```r
predict(model.6.south, best.v.l_5) # 98.59835 the second best variety and also tested in this location
```

```
##        1
## 98.59835
```

```r
predict(model.6.south, best.v.l_5.1) # 96.63376 select by mean
```

```
##        1
## 96.63376
```

```r
# Belleville_2019
best.v.l_6 <- data.frame(
   Loc = "Belleville_2019",Variety="CP9606"
)
best.v.l_7<- data.frame(
   Loc = "Belleville_2019",Variety="KWS19X07"
)
predict(model.6.south, best.v.l_6) # 87.67107  # selected by summary
```

```
##         1
## 87.67107
```

```r
predict(model.6.south, best.v.l_7) # 82.52397  # selected by mean
```

```
##         1
## 82.52397
```

```r
# north region (Hampshire_2019; Urbana_2018; Neoga_2018)
# Hampshire_2019
best.v.l_8<- data.frame(
   Loc = "Hampshire_2019",Variety="KWS19X03"
)
predict(model.6.north, best.v.l_8)# 125.6498 selected by summary and mean
```

```
##         1
## 125.6498
```

```r
# Urbana_2018
best.v.l_9 <- data.frame(
   Loc = "Urbana_2018",Variety="WX18C"
)
best.v.l_10<- data.frame(
   Loc = "Urbana_2018",Variety="KWS19X03"
)
best.v.l_10.1<- data.frame(
   Loc = "Urbana_2018",Variety="WX18C"
)
predict(model.6.north, best.v.l_9)# 101.3564 selected by mean
```

```
##         1
## 101.3564
```

```r
predict(model.6.north, best.v.l_10)# 110.6896 selected by summary_not show up
```

```
##         1
## 110.6896
```

```r
# Neoga_2018
best.v.l_11 <- data.frame(
   Loc = "Neoga_2018",Variety="H7W18"
)
best.v.l_12<- data.frame(
   Loc = "Neoga_2018",Variety="KWS19X03"
)
best.v.l_12.1<- data.frame(
   Loc = "Neoga_2018",Variety="3404"
)
predict(model.6.north, best.v.l_11) #  105.2071 selected by mean
```

```
##        1
## 105.2071
```

```r
predict(model.6.north, best.v.l_12) # 105.0126 selected by summary_not show up
```

```
##        1
## 105.0126
```

```r
predict(model.6.north, best.v.l_12.1) # 95.73333 as second high by summary
```

```
##        1
## 95.73333
```

- In the north region, at location Hampshire_2019, use variety "KWS19X03" can get the best yield, being 125.65 bu/acre.
- In the south region, at location Elkville_2019, use either CP9606 or SRW_9606 can get the maximum yield, being 112.1 bu/acre.