

alli_data

Ciera Martinez

July 11, 2017

Purpose: The purpose is to to correlate compounds to oviposition index. Mike wants to use multivariable regression or decision/regression to predict behavior. He also said to use linear or nonparametric.

```
## Library Prereqs
library(lme4)
```

```
## Warning: package 'lme4' was built under R version 3.3.2
```

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

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.3.2
```

```
library(languageR)
library(xlsx)
```

```
## Loading required package: rJava
```

```
## Loading required package: xlsxjars
```

```
library(reshape2)
```

```
## Functions
```

```
## modelcheck <- function(model, h = 8, w = 10.5) { ## because plot(lmer.obj) doesn't work rs <- residu
```

Read in Data

From .xls file.

```
sheet1 <- read.xlsx("../data/raw/alli_example_data.xlsx",
                     sheetIndex = 1)
```

```
## Check Sheet 1
head(sheet1)
```

##	peak_position	compound	Yeast.1	Yeast.2	Yeast.3
## 1	36	ethanol	1.000000	0.89141	0.955200
## 2	75	ethanol?	0.836450	0.62492	0.790530
## 3	135	ethyl acetate	0.116470	1.00000	0.058219
## 4	159	isobutanol	0.335250	0.26446	0.877350
## 5	199	acetic acid (very tiny)	0.186650	0.24482	0.000000

```
## 6          220 isoamyl acetate (very tiny) 0.095952 0.12586 0.130170
##   Yeast.4 Yeast.5   Yeast.6   variance      range      mean
## 1 0.94916 0.99384 0.0655540 7.7446e+14 70978000 61464000
## 2 0.80601 1.00000 0.0000000 1.9225e+11 1244800 841910
## 3 0.53774 0.45821 0.0099064 3.9752e+15 164080000 60227000
## 4 0.79982 1.00000 0.1306400 2.5710e+13 12009000 7845200
## 5 0.00000 1.00000 0.0000000 3.7623e+13 15803000 3770300
## 6 0.14871 1.00000 0.0892950 1.2298e+14 27995000 8146100
```

```
dim(sheet1)
```

```
## [1] 34 11
```

```
duplicated(sheet1)
```

```
## [1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [12] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [23] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [34] FALSE
```

```
### These are averaged and normalized
```

```
sheet2 <- read.xlsx("../data/raw/alli_example_data.xlsx",
                    sheetIndex = 2)
```

```
## Check Sheet 2
```

```
head(sheet2)
```

```
##   Fly.Line   Yeast X.eggs.on.yeast.side X.eggs.on.CTL.side Total.eggs
## 1  RW1005 Yeast 1             184             14           198
## 2  RW1005 Yeast 1             204              8           212
## 3  RW1005 Yeast 1             231              8           239
## 4  RW1005 Yeast 1             207             44           251
## 5  RW1005 Yeast 1             227             96           323
## 6  RW1005 Yeast 1             200             12           212
##   Oviposition.index
## 1             0.8585859
## 2             0.9245283
## 3             0.9330544
## 4             0.6494024
## 5             0.4055728
## 6             0.8867925
```

```
dim(sheet2)
```

```
## [1] 6 6
```

```
colnames(sheet2)
```

```
## [1] "Fly.Line"          "Yeast"              "X.eggs.on.yeast.side"
## [4] "X.eggs.on.CTL.side" "Total.eggs"         "Oviposition.index"
```

```
sheet3 <- read.xlsx("../data/raw/alli_example_data.xlsx",
                    sheetIndex = 3)
```

```
## Check Sheet 3
head(sheet3)
```

```
##   Fly.Line   Yeast X.eggs.on.yeast.side X.eggs.on.CTL.side Total.eggs
## 1  RW1005 Yeast 2             150             14           164
## 2  RW1005 Yeast 2             230             10           240
## 3  RW1005 Yeast 2             137             10           147
## 4  RW1005 Yeast 2             121              3           124
## 5  RW1005 Yeast 2              66              5            71
## 6  RW1005 Yeast 2             127              4           131
##   Oviposition.index
## 1             0.8292683
## 2             0.9166667
## 3             0.8639456
## 4             0.9516129
## 5             0.8591549
## 6             0.9389313
```

```
dim(sheet3)
```

```
## [1] 6 6
```

Cleaning, transforming, and merging

There were two different experiments that were performed. 1. Measuring of the yeast chemical compounds present, as detailed in `sheet1` 2. Measurements of fly behavior laying eggs with presense of said yeast `sheet2` and `sheet3`

I believe to do multivariate analysis on trying to find correlations between chemical compounds present and Oviposition.index, we just have to assume it is one experiment i.e. merge the data together.

Each measurement will also have additional columns corresponding to what is measured for each yeast type.

```
#####
## Sheet 1
#####
```

```
## What are these variance, range, mean columns? These don't make sense to me.
##           I am going to remove.
```

```
# Peak position doesn't really make sense either (prob describes chemical component peak?), but is we a
# Update: Looks like a peak can call an identical compound and this causes problems later
#           therefore I will merge this column with
```

```
## Check
colnames(sheet1)
```

```
## [1] "peak_position" "compound"      "Yeast.1"      "Yeast.2"
## [5] "Yeast.3"       "Yeast.4"       "Yeast.5"       "Yeast.6"
## [9] "variance"      "range"         "mean"
```

```
# Remove
sheet1 <- sheet1[,-c(9:11)]

# add new column
sheet1$compound_peak <- paste(sheet1$compound, sheet1$peak_position, sep = "_")

## Check
head(sheet1)
```

```
##   peak_position      compound Yeast.1 Yeast.2 Yeast.3
## 1          36      ethanol 1.000000 0.89141 0.955200
## 2          75 ethanol? 0.836450 0.62492 0.790530
## 3         135    ethyl acetate 0.116470 1.00000 0.058219
## 4         159    isobutanol 0.335250 0.26446 0.877350
## 5         199  acetic acid (very tiny) 0.186650 0.24482 0.000000
## 6         220 isoamyl acetate (very tiny) 0.095952 0.12586 0.130170
##   Yeast.4 Yeast.5 Yeast.6      compound_peak
## 1 0.94916 0.99384 0.0655540      ethanol_36
## 2 0.80601 1.00000 0.0000000      ethanol?_75
## 3 0.53774 0.45821 0.0099064    ethyl acetate_135
## 4 0.79982 1.00000 0.1306400    isobutanol_159
## 5 0.00000 1.00000 0.0000000  acetic acid (very tiny)_199
## 6 0.14871 1.00000 0.0892950 isoamyl acetate (very tiny)_220
```

```
## Now get rid of 1 and 2
sheet1 <- sheet1[,-c(1,2)]

# Transform Sheet1 first melt the Yeast columns from wide to long
sheet1_melt <- melt(sheet1, id.vars = "compound_peak",
  variable.name = "yeast")

## Check
head(sheet1_melt)
```

```
##           compound_peak  yeast  value
## 1      ethanol_36 Yeast.1 1.000000
## 2      ethanol?_75 Yeast.1 0.836450
## 3    ethyl acetate_135 Yeast.1 0.116470
## 4    isobutanol_159 Yeast.1 0.335250
## 5  acetic acid (very tiny)_199 Yeast.1 0.186650
## 6 isoamyl acetate (very tiny)_220 Yeast.1 0.095952
```

```
# Transform Sheet 1 compound components from long to wide
sheet1_cast <- dcast(sheet1_melt, yeast ~ compound_peak, value.var = "value")

## [x] Something wrong. Is there duplicate rows? Yes. See above.

## Check
head(sheet1_cast)
```

```
##   yeast 1-hexanol, 2-ethyl_1765 1, 3 dioxolane, 2,4,5-trimethyl_347
## 1 Yeast.1                      0                      0.34992
```

## 2 Yeast.2	0	0.41668
## 3 Yeast.3	0	0.17985
## 4 Yeast.4	1	1.00000
## 5 Yeast.5	0	0.00000
## 6 Yeast.6	0	0.00000
## 2-methoxy-4-vinylphenol (vanilla, coffee, clove)_3379		
## 1	1	
## 2	0	
## 3	0	
## 4	0	
## 5	0	
## 6	0	
## 3(2H)-Thiophenone, dihydro-2-methyl (blackberry)_1514		
## 1	1	
## 2	0	
## 3	0	
## 4	0	
## 5	0	
## 6	0	
## 7-octenoic acid, ethyl ester_2692 acetic acid (very tiny)_199		
## 1	1	0.18665
## 2	0	0.24482
## 3	0	0.00000
## 4	0	0.00000
## 5	0	1.00000
## 6	0	0.00000
## acetic acid, 2-phenylethyl ester _3100		
## 1	0	
## 2	0	
## 3	0	
## 4	0	
## 5	1	
## 6	0	
## acetic acid, 2-phenylethyl ester_3043		
## 1	0	
## 2	0	
## 3	0	
## 4	0	
## 5	1	
## 6	0	
## acetic acid, 2-phenylethyl ester_3071 butanoic acid, ethyl ester_589		
## 1	0.0032137	0.9772
## 2	0.0043946	0.0000
## 3	0.0000000	0.0000
## 4	0.0037545	1.0000
## 5	1.0000000	0.0000
## 6	0.0000000	0.0000
## decanoic acid, ethyl ester_3810 ethane, 1, -diethoxy (very tiny)_357		
## 1	0.0000	0.34992
## 2	0.0000	0.41668
## 3	1.0000	0.17985
## 4	0.0000	1.00000
## 5	0.4113	0.84443
## 6	0.0000	0.00000

```

## ethanol_36 ethanol?_75 ethyl acetate_135
## 1 1.000000 0.83645 0.1164700
## 2 0.891410 0.62492 1.0000000
## 3 0.955200 0.79053 0.0582190
## 4 0.949160 0.80601 0.5377400
## 5 0.993840 1.00000 0.4582100
## 6 0.065554 0.00000 0.0099064
## Hex-5-enoic acid, ethyl ester_1526 Hexanoic acid, ethyl ester_1595
## 1 1 1.0000
## 2 0 0.0000
## 3 0 0.2921
## 4 0 0.0000
## 5 0 0.0000
## 6 0 0.0000
## isoamyl acetate (very tiny)_220 isoamyl acetate (very tiny)_422
## 1 0.095952 0.00000
## 2 0.125860 0.00000
## 3 0.130170 0.00000
## 4 0.148710 1.00000
## 5 1.000000 0.93075
## 6 0.089295 0.00000
## isoamyl acetate_924 isoamyl acetate_937 isoamyl acetate_945
## 1 0.73309 0.0038196 0
## 2 0.42385 0.0020915 0
## 3 0.51934 0.0192910 0
## 4 1.00000 0.0024193 0
## 5 0.00000 1.0000000 1
## 6 0.20197 0.0075024 0
## isoamyl alcohol_373 isobutanol_159 isobutyl acetate_490
## 1 0.76947 0.33525 0.00000
## 2 0.39356 0.26446 0.00000
## 3 1.00000 0.87735 0.00000
## 4 0.95575 0.79982 0.80669
## 5 0.45979 1.00000 1.00000
## 6 0.12203 0.13064 0.00000
## methyl anthranilate (concord grapes)_3522 Nonanal_2202
## 1 1.00000 0.88033
## 2 0.00000 0.91702
## 3 0.67333 0.71627
## 4 0.58270 1.00000
## 5 0.00000 0.57656
## 6 0.56424 0.30628
## octanoic acid ethyl ester_2741 oxime, methoxyphenyl (small)_1092
## 1 1.00000 0.00000
## 2 0.31833 0.67012
## 3 0.76345 0.00000
## 4 0.00000 1.00000
## 5 0.00000 0.00000
## 6 0.00000 0.00000
## phenylethyl alcohol_2157 phenylethyl alcohol_2249 phenylethyl alcohol_2291
## 1 0.00000 0.58718 0.00000
## 2 0.00000 0.33826 0.65361
## 3 0.33585 0.73174 0.67828
## 4 1.00000 1.00000 1.00000

```

```
## 5          0.00000          0.26423          0.00000
## 6          0.00000          0.23312          0.00000
##   propanoic acid, ethyl ester_308 triacetin_3589
## 1          0.14003          0
## 2          0.38125          0
## 3          0.00000          0
## 4          0.69186          1
## 5          1.00000          0
## 6          0.00000          0
```

```
colnames(sheet1_cast) #There are 35 compounds measured for each yeast
```

```
## [1] "yeast"
## [2] "1-hexanol, 2-ethyl_1765"
## [3] "1, 3 dioxolane, 2,4,5-trimethyl_347"
## [4] "2-methoxy-4-vinylphenol (vanilla, coffee, clove)_3379"
## [5] "3(2H)-Thiophenone, dihydro-2-methyl (blackberry)_1514"
## [6] "7-octenoic acid, ethyl ester_2692"
## [7] "acetic acid (very tiny)_199"
## [8] "acetic acid, 2-phenylethyl ester _3100"
## [9] "acetic acid, 2-phenylethyl ester_3043"
## [10] "acetic acid, 2-phenylethyl ester_3071"
## [11] "butanoic acid, ethyl ester_589"
## [12] "decanoic acid, ethyl ester_3810"
## [13] "ethane, 1, -diethoxy (very tiny)_357"
## [14] "ethanol_36"
## [15] "ethanol?_75"
## [16] "ethyl acetate_135"
## [17] "Hex-5-enoic acid, ethyl ester_1526"
## [18] "Hexanoic acid, ethyl ester_1595"
## [19] "isoamyl acetate (very tiny)_220"
## [20] "isoamyl acetate (very tiny)_422"
## [21] "isoamyl acetate_924"
## [22] "isoamyl acetate_937"
## [23] "isoamyl acetate_945"
## [24] "isoamyl alcohol_373"
## [25] "isobutanol_159"
## [26] "isobutyl acetate_490"
## [27] "methyl anthranilate (concord grapes)_3522"
## [28] "Nonanal_2202"
## [29] "octanoic acid ethyl ester_2741"
## [30] "oxime, methoxyphenyl (small)_1092"
## [31] "phenylethyl alcohol_2157"
## [32] "phenylethyl alcohol_2249"
## [33] "phenylethyl alcohol_2291"
## [34] "propanoic acid, ethyl ester_308"
## [35] "triacetin_3589"
```

```
## rename back to original
```

```
sheet1 <- sheet1_cast
```

```
## Get rid of Yeast 3:6
```

```
sheet1 <- sheet1[-c(3:6),]
head(sheet1)
```

```
##      yeast 1-hexanol, 2-ethyl_1765 1, 3 dioxolane, 2,4,5-trimethyl_347
## 1 Yeast.1                      0                      0.34992
## 2 Yeast.2                      0                      0.41668
##      2-methoxy-4-vinylphenol (vanilla, coffee, clove)_3379
## 1                      1
## 2                      0
##      3(2H)-Thiophenone, dihydro-2-methyl (blackberry)_1514
## 1                      1
## 2                      0
##      7-octenoic acid, ethyl ester_2692 acetic acid (very tiny)_199
## 1                      1                      0.18665
## 2                      0                      0.24482
##      acetic acid, 2-phenylethyl ester _3100
## 1                      0
## 2                      0
##      acetic acid, 2-phenylethyl ester_3043
## 1                      0
## 2                      0
##      acetic acid, 2-phenylethyl ester_3071 butanoic acid, ethyl ester_589
## 1                      0.0032137                      0.9772
## 2                      0.0043946                      0.0000
##      decanoic acid, ethyl ester_3810 ethane, 1, -diethoxy (very tiny)_357
## 1                      0                      0.34992
## 2                      0                      0.41668
##      ethanol_36 ethanol?_75 ethyl acetate_135
## 1      1.00000      0.83645      0.11647
## 2      0.89141      0.62492      1.00000
##      Hex-5-enoic acid, ethyl ester_1526 Hexanoic acid, ethyl ester_1595
## 1                      1                      1
## 2                      0                      0
##      isoamyl acetate (very tiny)_220 isoamyl acetate (very tiny)_422
## 1                      0.095952                      0
## 2                      0.125860                      0
##      isoamyl acetate_924 isoamyl acetate_937 isoamyl acetate_945
## 1                      0.73309      0.0038196      0
## 2                      0.42385      0.0020915      0
##      isoamyl alcohol_373 isobutanol_159 isobutyl acetate_490
## 1                      0.76947      0.33525      0
## 2                      0.39356      0.26446      0
##      methyl anthranilate (concord grapes)_3522 Nonanal_2202
## 1                      1      0.88033
## 2                      0      0.91702
##      octanoic acid ethyl ester_2741 oxime, methoxyphenyl (small)_1092
## 1                      1.00000                      0.00000
## 2                      0.31833                      0.67012
##      phenylethyl alcohol_2157 phenylethyl alcohol_2249 phenylethyl alcohol_2291
## 1                      0                      0.58718      0.00000
## 2                      0                      0.33826      0.65361
##      propanoic acid, ethyl ester_308 triacetin_3589
## 1                      0.14003      0
```



```
## 2                                0.38125                                0
```

Sheet 2 and 3

```
#####
## sheet2 and sheet3
#####

# make sure they have same columns
colnames(sheet3)

## [1] "Fly.Line"          "Yeast"          "X.eggs.on.yeast.side"
## [4] "X.eggs.on.CTL.side" "Total.eggs"     "Oviposition.index"

colnames(sheet2)

## [1] "Fly.Line"          "Yeast"          "X.eggs.on.yeast.side"
## [4] "X.eggs.on.CTL.side" "Total.eggs"     "Oviposition.index"

# add together
sheet23 <- rbind(sheet2, sheet3)

## Now merge all three by yeast
## First make sure they are named the same

colnames(sheet23)

## [1] "Fly.Line"          "Yeast"          "X.eggs.on.yeast.side"
## [4] "X.eggs.on.CTL.side" "Total.eggs"     "Oviposition.index"

summary(sheet23$Yeast)

## Yeast 1 Yeast 2
##      6      6

summary(sheet1$yeast)

## Yeast.1 Yeast.2 Yeast.3 Yeast.4 Yeast.5 Yeast.6
##      1      1      0      0      0      0

## Fix
colnames(sheet23)[2] <- "yeast"

sheet23$yeast <- gsub("Yeast 1", "Yeast.1", sheet23$yeast)
sheet23$yeast <- gsub("Yeast 2", "Yeast.2", sheet23$yeast)

## Now we can merge

## Check
dim(sheet23)
```

```
## [1] 12 6
```

```
dim(sheet1)
```

```
## [1] 2 35
```

```
## Merge
sheet123 <- merge(sheet23, sheet1, by = "yeast")

##Check
dim(sheet123)
```

```
## [1] 12 40
```

```
head(sheet123)
```

```
##      yeast Fly.Line X.eggs.on.yeast.side X.eggs.on.CTL.side Total.eggs
## 1 Yeast.1   RW1005                184                14        198
## 2 Yeast.1   RW1005                204                 8        212
## 3 Yeast.1   RW1005                231                 8        239
## 4 Yeast.1   RW1005                207                44        251
## 5 Yeast.1   RW1005                227                96        323
## 6 Yeast.1   RW1005                200                12        212
##      Oviposition.index 1-hexanol, 2-ethyl_1765
## 1          0.8585859                0
## 2          0.9245283                0
## 3          0.9330544                0
## 4          0.6494024                0
## 5          0.4055728                0
## 6          0.8867925                0
##      1, 3 dioxolane, 2,4,5-trimethyl_347
## 1                0.34992
## 2                0.34992
## 3                0.34992
## 4                0.34992
## 5                0.34992
## 6                0.34992
##      2-methoxy-4-vinylphenol (vanilla, coffee, clove)_3379
## 1                1
## 2                1
## 3                1
## 4                1
## 5                1
## 6                1
##      3(2H)-Thiophenone, dihydro-2-methyl (blackberry)_1514
## 1                1
## 2                1
## 3                1
## 4                1
## 5                1
## 6                1
##      7-octenoic acid, ethyl ester_2692 acetic acid (very tiny)_199
```

## 1	1	0.18665
## 2	1	0.18665
## 3	1	0.18665
## 4	1	0.18665
## 5	1	0.18665
## 6	1	0.18665
##	acetic acid, 2-phenylethyl ester _3100	
## 1	0	
## 2	0	
## 3	0	
## 4	0	
## 5	0	
## 6	0	
##	acetic acid, 2-phenylethyl ester _3043	
## 1	0	
## 2	0	
## 3	0	
## 4	0	
## 5	0	
## 6	0	
##	acetic acid, 2-phenylethyl ester _3071 butanoic acid, ethyl ester _589	
## 1	0.0032137	0.9772
## 2	0.0032137	0.9772
## 3	0.0032137	0.9772
## 4	0.0032137	0.9772
## 5	0.0032137	0.9772
## 6	0.0032137	0.9772
##	decanoic acid, ethyl ester _3810 ethane, 1, -diethoxy (very tiny) _357	
## 1	0	0.34992
## 2	0	0.34992
## 3	0	0.34992
## 4	0	0.34992
## 5	0	0.34992
## 6	0	0.34992
##	ethanol _36 ethanol? _75 ethyl acetate _135	
## 1	1 0.83645	0.11647
## 2	1 0.83645	0.11647
## 3	1 0.83645	0.11647
## 4	1 0.83645	0.11647
## 5	1 0.83645	0.11647
## 6	1 0.83645	0.11647
##	Hex-5-enoic acid, ethyl ester _1526 Hexanoic acid, ethyl ester _1595	
## 1	1	1
## 2	1	1
## 3	1	1
## 4	1	1
## 5	1	1
## 6	1	1
##	isoamyl acetate (very tiny) _220 isoamyl acetate (very tiny) _422	
## 1	0.095952	0
## 2	0.095952	0
## 3	0.095952	0
## 4	0.095952	0
## 5	0.095952	0

```

## 6          0.095952          0
## isoamyl acetate_924 isoamyl acetate_937 isoamyl acetate_945
## 1          0.73309          0.0038196          0
## 2          0.73309          0.0038196          0
## 3          0.73309          0.0038196          0
## 4          0.73309          0.0038196          0
## 5          0.73309          0.0038196          0
## 6          0.73309          0.0038196          0
## isoamyl alcohol_373 isobutanol_159 isobutyl acetate_490
## 1          0.76947          0.33525          0
## 2          0.76947          0.33525          0
## 3          0.76947          0.33525          0
## 4          0.76947          0.33525          0
## 5          0.76947          0.33525          0
## 6          0.76947          0.33525          0
## methyl anthranilate (concord grapes)_3522 Nonanal_2202
## 1          1          0.88033
## 2          1          0.88033
## 3          1          0.88033
## 4          1          0.88033
## 5          1          0.88033
## 6          1          0.88033
## octanoic acid ethyl ester_2741 oxime, methoxyphenyl (small)_1092
## 1          1          0
## 2          1          0
## 3          1          0
## 4          1          0
## 5          1          0
## 6          1          0
## phenylethyl alcohol_2157 phenylethyl alcohol_2249 phenylethyl alcohol_2291
## 1          0          0.58718          0
## 2          0          0.58718          0
## 3          0          0.58718          0
## 4          0          0.58718          0
## 5          0          0.58718          0
## 6          0          0.58718          0
## propanoic acid, ethyl ester_308 triacetin_3589
## 1          0.14003          0
## 2          0.14003          0
## 3          0.14003          0
## 4          0.14003          0
## 5          0.14003          0
## 6          0.14003          0

```

```

write.csv(sheet123, "../data/output/combinedData.csv")

```