report

ZJH

2022/4/12

```r
set.seed(100)
tea <- readRDS('nonglin_tea.RDS')
analyze_data <- readRDS('analyze.RDS')
library(tidyverse)
```
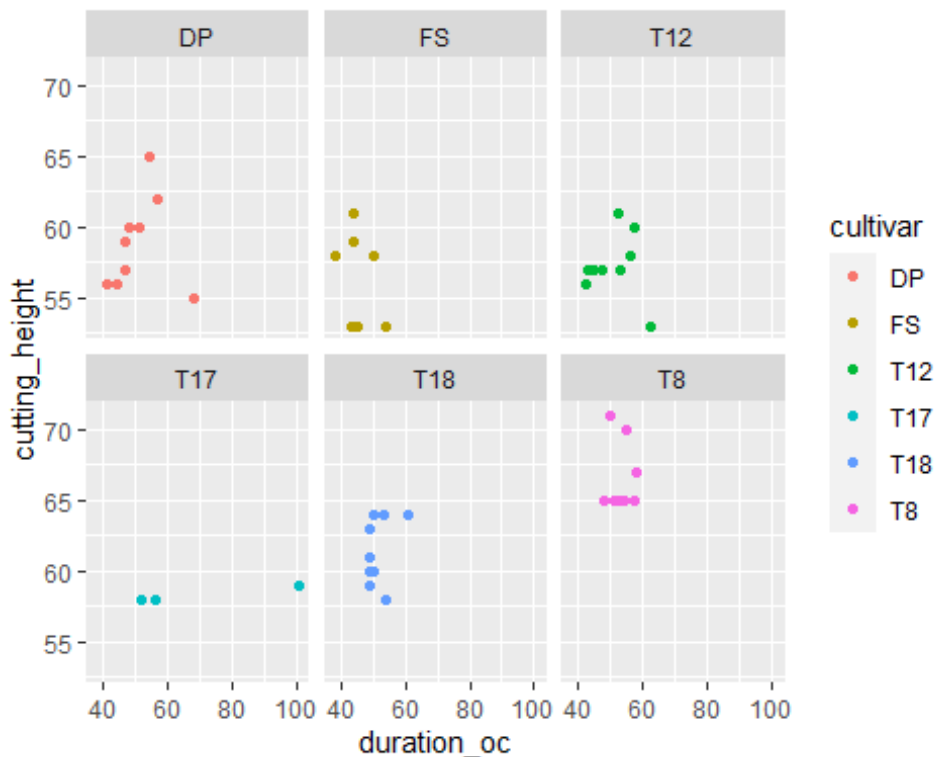


```r
numeric_dat <- analyze_data %>% select(.,-c(key,level,observe_ys,observe_date:sample_label,G1L:O9L,FAA:total_catechins,year,season,temp_differ))
train_idx <- sample(1:141,70)
test_idx <- !(1:141 %in%train_idx)
train <- numeric_dat[train_idx,]
test <- numeric_dat[test_idx,]
xtrain <- model.matrix(polyphenol~., train)[,-1]
ytrain <- train$polyphenol

ytest <- test$polyphenol
xtest <- model.matrix(polyphenol~., test)[,-1]
```

```
lambdas_to_try <- 10^seq(-3, 7, length.out = 100)
lasso_cv <- cv.glmnet(xtrain, ytrain, alpha = 1, lambda = lambdas_to_try)
plot(lasso_cv)
```
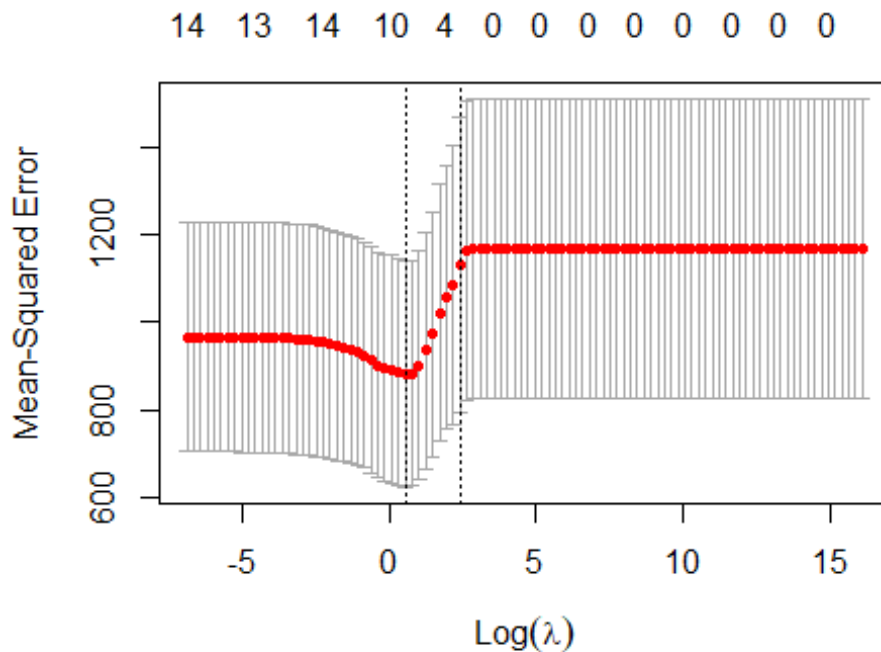

採前調查的開面比占了很重要的比重，另外裁切高度居然也相關，這倒是蠻令我意外的。

```
best_lambda_lasso <- lasso_cv$lambda.min
best_lambda_lasso

## [1] 1.707353

lasso_mod <- glmnet(xtrain, ytrain, alpha = 1, lambda = best_lambda_lasso)
predict.glmnet(lasso_mod, type = 'coefficients')

## 15 x 1 sparse Matrix of class "dgCMatrix"
##                            s0
## (Intercept)        70.01201772
## duration_oc         0.28406478
## Total_leaf         -0.69405640
## GORatio             .
## OTRatio           -42.31734045
## cutting_height      1.37797957
## tea_buds_cm        -2.32960496
## leaf_number        -3.23305674
```

```
## avg_inter_node     .
## open_plane         .
## buds_weight_100    0.15163174
## acu_mean_temp      0.00166651
## rain               .
## growth_mean_temp   0.03616287
## Growth_length      .
```

```
fit <- lm(polyphenol~Total_leaf+OTRatio+cutting_height+leaf_number+open
_plane+buds_weight_100+acu_mean_temp+growth_mean_temp,data = test)
lasso_pred <- predict(lasso_mod,  newx = xtest)
lm_pred <- predict(fit)

mean((lasso_pred-ytest)^2)
```

```
## [1] 928.667
```

```
mean((lm_pred-ytest)^2)
```

```
## [1] 806.6858
```

太大了，一個標準差就跟自己差不多