

In-class Exercise

Zewei Li, Lin Wang, Liwei Fang, Qingxuan Zheng

1/28/2020

```
mydata=read.csv("bank.csv")
names(mydata)
```

```
## [1] "age"      "job"      "marital"  "education" "default"  "balance"
## [7] "housing"  "loan"     "contact"  "day"       "month"    "duration"
## [13] "campaign" "previous" "y"
```

```
typeof(mydata)
```

```
## [1] "list"
```

```
#full_mydata_model = lm(balance ~ ., data = mydata)
full_mydata_model = lm(balance ~ age + job + marital + education + default + housing + loan + contact +
#this function calculates coefficient
full=summary(full_mydata_model)
anova_1 = aov(full_mydata_model,data=mydata)# produce a error of 2934.073
fmm = coef(full)
fmm_signi <- fmm[fmm[, "Pr(>|t|)"]<0.05,]# from here we remove the insignificant parts and only left age

#recheck
model_2 = lm(balance ~ age + job + marital + education + default + loan + month , data = mydata)
#only left significant as test
anova2 = aov(model_2,data = mydata)#produce a error of 2932.844
anova2
```

```
## Call:
## aov(formula = model_2, data = mydata)
##
## Terms:
##          age          job      marital  education    default
## Sum of Squares 287649608 379477971 125500428 103553969 178912827
## Deg. of Freedom      1          11          2           3           1
##          loan      month  Residuals
## Sum of Squares 139371125 1106273367 38621067008
## Deg. of Freedom      1          11          4490
##
## Residual standard error: 2932.844
## Estimated effects may be unbalanced
```

```

model_3 = lm(balance ~ job + marital + education + default+ loan + month , data = mydata)
#remove age as test
anova3= aov(model_3,data=mydata)#produce a erro of 2936.964(since the error increase so we keep it)
anova3

```

```

## Call:
## aov(formula = model_3, data = mydata)
##
## Terms:
##              job      marital  education    default      loan
## Sum of Squares  547747657    66497705   100194770   186490782   145940017
## Deg. of Freedom      11          2           3           1           1
##              month  Residuals
## Sum of Squares  1156658358 38738277015
## Deg. of Freedom      11      4491
##
## Residual standard error: 2936.964
## Estimated effects may be unbalanced

```

```

model_4=lm(balance ~ age + marital + default + education + loan + month, data = mydata)
#remove job
anova4=aov(model_4,data = mydata)# error 2934.919
anova4

```

```

## Call:
## aov(formula = model_4, data = mydata)
##
## Terms:
##              age      marital  default  education      loan
## Sum of Squares  287649608   158403113   185324090   285045657   141289159
## Deg. of Freedom      1          2           1           3           1
##              month  Residuals
## Sum of Squares  1113610658 38770484019
## Deg. of Freedom      11      4501
##
## Residual standard error: 2934.919
## Estimated effects may be unbalanced

```

```

model_5=lm(balance ~ age + job + education + default + loan + month, data = mydata)
#remove marital
anova5=aov(model_5,data = mydata)#error 2935.928
anova5

```

```

## Call:
## aov(formula = model_5, data = mydata)
##
## Terms:
##              age      job  education    default      loan
## Sum of Squares  287649608   379477971   110474643   190561014   147219739
## Deg. of Freedom      1          11           3           1           1
##              month  Residuals
## Sum of Squares  1106861867 38719561462

```

```
## Deg. of Freedom      11      4492
##
## Residual standard error: 2935.928
## Estimated effects may be unbalanced
```

```
model_6=lm(balance ~ age + job + marital + default + loan + month , data = mydata)
# remove education
anova6=aov(model_6,data=mydata) # error 2934.378
anova6
```

```
## Call:
## aov(formula = model_6, data = mydata)
##
## Terms:
##          age          job      marital      default      loan
## Sum of Squares  287649608  379477971 125500428 188096714 149119533
## Deg. of Freedom      1      11          2          1          1
##          month  Residuals
## Sum of Squares 1124648123 38687313927
## Deg. of Freedom      11      4493
##
## Residual standard error: 2934.378
## Estimated effects may be unbalanced
```

```
model_7=lm(balance ~ age + job + marital + education + default + loan , data = mydata)
#remove month
anova7=aov(model_7,data=mydata)# error 2970.915
anova7
```

```
## Call:
## aov(formula = model_7, data = mydata)
##
## Terms:
##          age          job      marital  education      default
## Sum of Squares  287649608  379477971 125500428 103553969 178912827
## Deg. of Freedom      1      11          2          3          1
##          loan  Residuals
## Sum of Squares 139371125 39727340375
## Deg. of Freedom      1      4501
##
## Residual standard error: 2970.915
## Estimated effects may be unbalanced
```

```
model_8=lm(balance ~ age + job + marital + education + loan + month , data = mydata)
#remove default
anova8 = aov(model_8,data=mydata)#error 2938.165
anova8 #error 2938.165
```

```
## Call:
## aov(formula = model_8, data = mydata)
##
## Terms:
```

```
##               age      job      marital  education      loan
## Sum of Squares 287649608 379477971 125500428 103553969 158624152
## Deg. of Freedom      1      11      2      3      1
##               month  Residuals
## Sum of Squares 1117044653 38769955523
## Deg. of Freedom      11      4491
##
## Residual standard error: 2938.165
## Estimated effects may be unbalanced
```

```
model_9= lm(balance ~ age + job + marital + education + default + month , data = mydata)
#remove loan
anova9 = aov(model_9,data=mydata)#error 2936.033
anova9
```

```
## Call:
## aov(formula = model_9, data = mydata)
##
## Terms:
##               age      job      marital  education      default
## Sum of Squares 287649608 379477971 125500428 103553969 178912827
## Deg. of Freedom      1      11      2      3      1
##               month  Residuals
## Sum of Squares 1152996783 38713714718
## Deg. of Freedom      11      4491
##
## Residual standard error: 2936.033
## Estimated effects may be unbalanced
```

```
null_mydata_model = lm(balance ~ age + job + marital + education + default + loan + month , data = mydata)
anova(full_mydata_model,null_mydata_model)
```

```
## Analysis of Variance Table
##
## Model 1: balance ~ age + job + marital + education + default + housing +
## loan + contact + day + month + duration + campaign + previous +
## y
## Model 2: balance ~ age + job + marital + education + default + loan +
## month
## Res.Df    RSS Df Sum of Sq    F Pr(>F)
## 1    4482 3.8585e+10
## 2    4490 3.8621e+10 -8 -36488089 0.5298 0.8349
```

```
# In this part we use two models to get the final result, the first one we use the full model which includes
# variables and we use Ftest to filter the insignificant variable and keep the rest to recheck the results
# Then what we get finally named null model that is our final model, since compare with the full one the
# v1=var.test(balance~y, data = mydata)
#v1
#v2=var.test(balance-default, data = mydata)
#v2
```