```r
# orates.r
d0 = read.csv("rates.csv")
head(d0)
```

```
##   row years quarters orates
## 1   1  2009        1  0.560
## 2   2  2009        2  0.702
## 3   3  2009        3  0.800
## 4   4  2009        4  0.568
## 5   5  2010        1  0.575
## 6   6  2010        2  0.738
```

```r
# plot
yb = c(0,1)
xb = c(0,30)
plot(orates~row,d0,type='l',ylim=yb,xlim=xb,ylab='occupancy rate',xlab='quarter')
grid()

# regression with predictor row
model1 = lm(orates~row,d0)
abline(model1,col="red")
summary(model1)
```

```
##
## Call:
## lm(formula = orates ~ row, data = d0)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.10229 -0.08678 -0.03420  0.07243  0.19201
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.639168   0.046211  13.832 4.97e-11 ***
## row         0.005260   0.003858   1.364     0.19
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.09948 on 18 degrees of freedom
## Multiple R-squared:  0.09363,    Adjusted R-squared:  0.04327
## F-statistic: 1.859 on 1 and 18 DF,  p-value: 0.1895
```

```r
# not good fit

# regression with all predictors
model1b = lm(orates~.,d0)
summary(model1b)
```

```
##
## Call:
## lm(formula = orates ~ ., data = d0)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.11036 -0.08354 -0.02756  0.07336  0.18848
##
## Coefficients: (1 not defined because of singularities)
##              Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)  45.54864  167.30106   0.272    0.789
## row           0.01064    0.02043   0.521    0.609
## years        -0.02236    0.08330  -0.268    0.792
## quarters           NA         NA      NA       NA
##
## Residual standard error: 0.1021 on 17 degrees of freedom
## Multiple R-squared:  0.09745,    Adjusted R-squared:  -0.008731
## F-statistic: 0.9178 on 2 and 17 DF,  p-value: 0.4183
```

```
# singularities when predictors are correlated
cor(d0)
```

```
##                row     years  quarters    orates
## row      1.0000000 0.9810229 0.1938917 0.3059836
## years    0.9810229 1.0000000 0.0000000 0.2881844
## quarters 0.1938917 0.0000000 1.0000000 0.1200054
## orates   0.3059836 0.2881844 0.1200054 1.0000000
```

```
# remove row column

# regression with years and quarters
model1c = lm(orates~years+quarters,d0)
summary(model1c)
```

```
##
## Call:
## lm(formula = orates ~ years + quarters, data = d0)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.11036 -0.08354 -0.02756  0.07336  0.18848
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -39.95440   32.47913  -1.230    0.235
## years         0.02020    0.01615   1.251    0.228
## quarters      0.01064    0.02043   0.521    0.609
##
## Residual standard error: 0.1021 on 17 degrees of freedom
## Multiple R-squared:  0.09745,    Adjusted R-squared:  -0.008731
## F-statistic: 0.9178 on 2 and 17 DF,  p-value: 0.4183
```

```
# not good fit

# dataframe d1 -quarter as categorical var, and no column row-
d1 = d0
d1$row = NULL
d1$quarters = as.factor(d1$quarters)
str(d1)
```

```
## 'data.frame':    20 obs. of  3 variables:
##  $ years   : int  2009 2009 2009 2009 2010 2010 2010 2010 2011 2011 ...
##  $ quarters: Factor w/ 4 levels "1","2","3","4": 1 2 3 4 1 2 3 4 1 2 ...
##  $ orates  : num  0.56 0.702 0.8 0.568 0.575 0.738 0.868 0.605 0.594 0.738 ...
```

```
head(d1)
```

```
##   years quarters orates
## 1  2009        1  0.560
## 2  2009        2  0.702
```

```
## 3  2009        3  0.800
## 4  2009        4  0.568
## 5  2010        1  0.575
## 6  2010        2  0.738
```

```
# regression with categorical var
model2 = lm(orates~.,d1)
summary(model2)
```

```
##
## Call:
## lm(formula = orates ~ ., data = d1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.0862 -0.0083 -0.0023  0.0153  0.0730
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -40.019000  12.103047  -3.307  0.00479 **
## years         0.020200   0.006018   3.356  0.00433 **
## quarters2     0.141000   0.024074   5.857 3.15e-05 ***
## quarters3     0.212000   0.024074   8.806 2.59e-07 ***
## quarters4     0.011800   0.024074   0.490  0.63111
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.03806 on 15 degrees of freedom
## Multiple R-squared:  0.8894, Adjusted R-squared:  0.8599
## F-statistic: 30.16 on 4 and 15 DF,  p-value: 5.159e-07
```

```
# model improved

# dataframe d2 created by adding column of fitted values to original dataframe d0
d2 = d0
pred = model2$fitted.values
d2$predictions = pred
head(d2)
```

```
##   row years quarters orates predictions
## 1   1  2009        1  0.560      0.5628
## 2   2  2009        2  0.702      0.7038
## 3   3  2009        3  0.800      0.7748
## 4   4  2009        4  0.568      0.5746
## 5   5  2010        1  0.575      0.5830
## 6   6  2010        2  0.738      0.7240
```

```
# plot fitted values
lines(predictions~row,d2,col="blue")

# predict 2014-2015 with model m2

years = rep(2014:2015,each = 4)
quarters = rep(1:4,times=2)
quarters = as.factor(quarters)
newval = data.frame(years,quarters)
newval
```

```
##   years quarters
```

```
## 1  2014        1
## 2  2014        2
## 3  2014        3
## 4  2014        4
## 5  2015        1
## 6  2015        2
## 7  2015        3
## 8  2015        4
```

```
predictions = predict(model2,newval)
d3 = data.frame(newval,predictions)
d3
```

```
##   years quarters predictions
## 1  2014        1      0.6638
## 2  2014        2      0.8048
## 3  2014        3      0.8758
## 4  2014        4      0.6756
## 5  2015        1      0.6840
## 6  2015        2      0.8250
## 7  2015        3      0.8960
## 8  2015        4      0.6958
```

```
#
# remove cols row and response from d2
d2$orates = NULL
d2$row = NULL
# combine dataframes
d5 = rbind(d2,d3)
d5
```

```
##     years quarters predictions
## 1    2009        1      0.5628
## 2    2009        2      0.7038
## 3    2009        3      0.7748
## 4    2009        4      0.5746
## 5    2010        1      0.5830
## 6    2010        2      0.7240
## 7    2010        3      0.7950
## 8    2010        4      0.5948
## 9    2011        1      0.6032
## 10   2011        2      0.7442
## 11   2011        3      0.8152
## 12   2011        4      0.6150
## 13   2012        1      0.6234
## 14   2012        2      0.7644
## 15   2012        3      0.8354
## 16   2012        4      0.6352
## 17   2013        1      0.6436
## 18   2013        2      0.7846
## 19   2013        3      0.8556
## 20   2013        4      0.6554
## 110  2014        1      0.6638
## 21   2014        2      0.8048
## 31   2014        3      0.8758
## 41   2014        4      0.6756
## 51   2015        1      0.6840
## 61   2015        2      0.8250
```

```
## 71   2015        3      0.8960
## 81   2015        4      0.6958
```

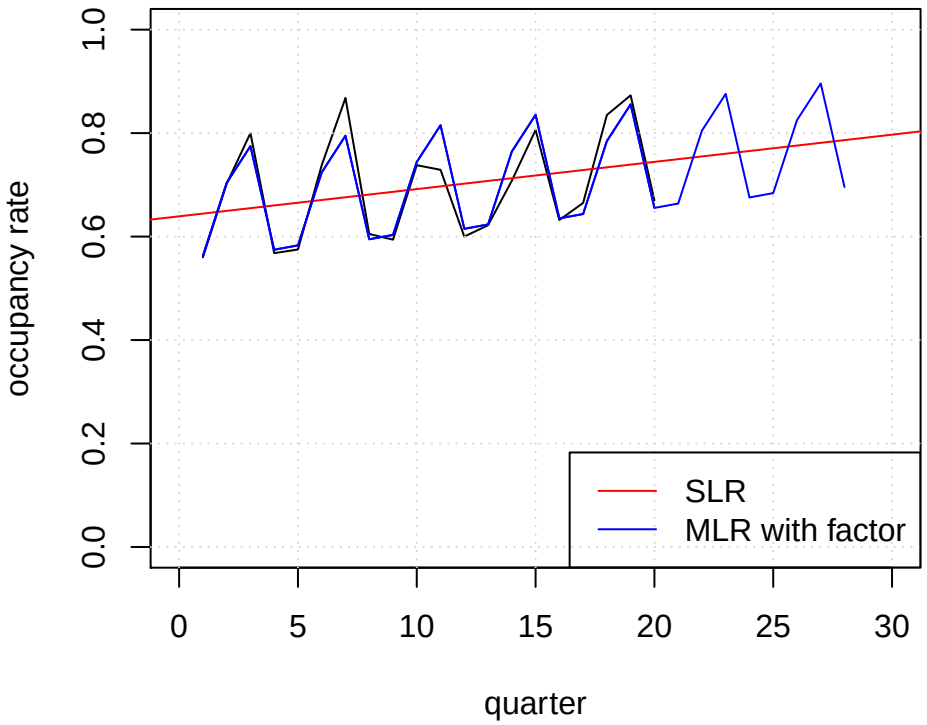
```
# rownames adjusted to prevent duplicated rownames
#
# new rownames
nrow(d5)
```

```
## [1] 28
```

```
rownames(d5)=1:28
d5
```

```
##    years quarters predictions
## 1   2009        1      0.5628
## 2   2009        2      0.7038
## 3   2009        3      0.7748
## 4   2009        4      0.5746
## 5   2010        1      0.5830
## 6   2010        2      0.7240
## 7   2010        3      0.7950
## 8   2010        4      0.5948
## 9   2011        1      0.6032
## 10  2011        2      0.7442
## 11  2011        3      0.8152
## 12  2011        4      0.6150
## 13  2012        1      0.6234
## 14  2012        2      0.7644
## 15  2012        3      0.8354
## 16  2012        4      0.6352
## 17  2013        1      0.6436
## 18  2013        2      0.7846
## 19  2013        3      0.8556
## 20  2013        4      0.6554
## 21  2014        1      0.6638
## 22  2014        2      0.8048
## 23  2014        3      0.8758
## 24  2014        4      0.6756
## 25  2015        1      0.6840
## 26  2015        2      0.8250
## 27  2015        3      0.8960
## 28  2015        4      0.6958
```

```
# plot predictions
row = 1:28
lines(d5$predictions~row,col="blue")
legend("bottomright", legend=c("SLR", "MLR with factor"),
       col=c("red", "blue"),lty = c(1,1))
```


occupancy rate
0.0
0.2
0.4
0.6
0.8
1.0
0
5
10
15
20
25
30
quarter
SLR
MLR with factor