Hw04ST430Yu

Haozhe (Jerry) Yu

2023-11-04

Question 1

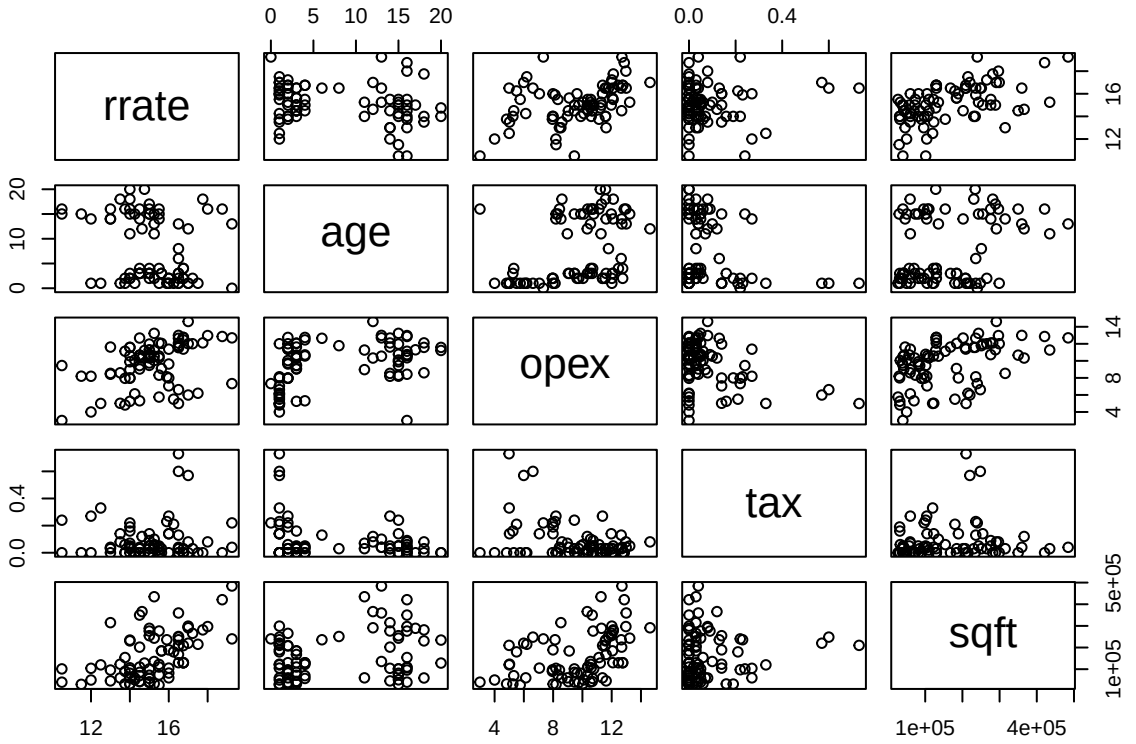
> Problem1A: For this problem, use the Commercial Property data set from KNNL Problem 6.18. The response variable is rental rates. The explanatory variables are age, operating expenses and taxes, vacancy rates, and total square footage.

```
Property <- as_tibble(read_table("https://users.stat.ufl.edu/~rrandles/sta4210/Rclassnotes/data/textdata
            col_names = c("rrate","age","opex","tax","sqft")
            ))
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   rrate = col_double(),
##   age = col_double(),
##   opex = col_double(),
##   tax = col_double(),
##   sqft = col_double()
## )
```

1. Obtain a scatterplot matrix for all 5 variables (age, operating expenses and taxes, vacancy rates, total square footage, and rental rates) and give your comments about linearity of variables.

```
pairs(Property)
```



Most of the variables seem lack linearity, except for sqft, which seems to positively linearly correlated with rental rate and operating expense, Rental Rate seems to also show a weak positive linear correlation with operating expense.

Find the correlation of all pairs of variables and give your comments about the association between variables.

```
PropertyCor <- cor(Property)
PropertyCor
```

```
##              rrate        age       opex         tax       sqft
## rrate  1.00000000 -0.2502846  0.4137872  0.06652647 0.53526237
## age   -0.25028456  1.0000000  0.3888264 -0.25266347 0.28858350
## opex   0.41378716  0.3888264  1.0000000 -0.37976174 0.44069713
## tax    0.06652647 -0.2526635 -0.3797617  1.00000000 0.08061073
## sqft   0.53526237  0.2885835  0.4406971  0.08061073 1.00000000
```

My observations seem to be borne out, as the largest correlations (abs value) seem to be between rental rate and operating expenses (0.4137872), rental rate and square footage (0.5352624), and square footage and operating expenses (0.4406971).

3) Run the multiple regressions with age, operating expenses and taxes, vacancy rates, and total square footage as the explanatory variables and rental rates as the response variable.

```
Propertym <- lm(rrate~age + opex + tax + sqft,Property)
```

a) Give the fitted (Estimated) regression equation.

Rental Rate $(Y_i) = 12.201(beta_0)$ + -0.142age$(beta_1)$ + 0.282Operating_Expenses$(beta_2)$ + 0.619Tax$(beta_3)$ + 7.9243019×10^{-6}Square_Footage$(beta_4)$

b) Interpret the estimated parameters in the context of the problem

For each incremental increase in Age, holding all other variables constant, there is a -0.142 decrease in rental rate.

For each incremental increase in Operating Expenses, holding all other variables constant, there is a 0.282increase in rental rate.

For each incremental increase in Tax, holding all other variables constant, there is a 0.619 decrease in rental rate.

For each increase in 100,000 Square Feet, holding all other variables constant, there is a 0.792 increase in rental rate.

c) Give the value of R2 and Adj R2

$r^2 = 0.5847496$

$r^2_{adj} = 0.5628943$

d) Give the results of the significance test for the null hypothesis that the four regression coefficients for the explanatory variables are all zero. Be sure to give the null and alternative hypotheses, the test

statistic and its associated degrees of freedom, the p-value, and a brief conclusion in words.

```
summary(Propertym) #%>% str()
```

```
##
## Call:
## lm(formula = rrate ~ age + opex + tax + sqft, data = Property)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.1872 -0.5911 -0.0910  0.5579  2.9441
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.220e+01  5.780e-01  21.110  < 2e-16 ***
## age         -1.420e-01  2.134e-02  -6.655 3.89e-09 ***
```

```
## opex         2.820e-01  6.317e-02   4.464 2.75e-05 ***
## tax          6.193e-01  1.087e+00   0.570     0.57
## sqft         7.924e-06  1.385e-06   5.722 1.98e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.137 on 76 degrees of freedom
## Multiple R-squared:  0.5847, Adjusted R-squared:  0.5629
## F-statistic: 26.76 on 4 and 76 DF,  p-value: 7.272e-14
```

```
7.272e-14
```

```
## [1] 7.272e-14
```

```
1-pf(summary(Propertym)$fstat[1],summary(Propertym)$fstat[2],summary(Propertym)$fstat[3])
```

```
##        value
## 7.271961e-14
```

```
round(summary(Propertym)$coeff[1,3],3)
```

```
## [1] 21.11
```

```
summary(Propertym)$fstat
```

```
##    value    numdf    dendf
## 26.75553  4.00000 76.00000
```

```
Propertytable <- data.frame("H0"=c("b1 = b2 = b3 = b4 =0",
                                   "b0=0",
                                   "b1-0",
                                   "b2=0",
                                   "b3=0",
                                   "b4-0"),
                            "Test Statistic" = c (round(summary(Propertym)$fstat[1],3),
                                                  round(summary(Propertym)$coeff[1,3],3),
                                                  round(summary(Propertym)$coeff[2,3],3),
                                                  round(summary(Propertym)$coeff[3,3],3),
                                                  round(summary(Propertym)$coeff[4,3],3),
                                                  round(summary(Propertym)$coeff[5,3],3)

                            ),
                            "p-value" = c(1-pf(summary(Propertym)$fstat[1],summary(Propertym)$fstat[2],s
                                                  summary(Propertym)$coeff[1,4],
                                                  summary(Propertym)$coeff[2,4],
                                                  summary(Propertym)$coeff[3,4],
                                                  summary(Propertym)$coeff[4,4],
                                                  summary(Propertym)$coeff[5,4]),
                            "Reject H0?" = c("Yes",
                                             "Yes",
                                             "Yes",
```

```
                                            "Yes",
                                            "No",
                                            "Yes")
                              ) %>% as_tibble()

Propertytable
```

```
## # A tibble: 6 x 4
##   H0                    Test.Statistic  p.value Reject.H0.
##   <chr>                          <dbl>    <dbl> <chr>
## 1 b1 = b2 = b3 = b4 =0           26.8  7.27e-14 Yes
## 2 b0=0                           21.1  1.60e-33 Yes
## 3 b1-0                           -6.66 3.89e- 9 Yes
## 4 b2=0                            4.46 2.75e- 5 Yes
## 5 b3=0                            0.57 5.70e- 1 No
## 6 b4-0                            5.72 1.98e- 7 Yes
```

Conclusion

- row1: as $p<0.05$, we reject H0 and conclude that there is evidence to support a linear relationship between at least one of the predictor variables and the response variable
- row2: as $p<0.05$, we reject H0 and conclude that there is evidence to support the claim that controlling for all other variables, the intercept is nonzero.
- row2: as $p<0.05$, we reject H0 and conclude that there is evidence to support the claim that controlling for all other variables, there exists a linear relationship between age and rental rate.
- row2: as $p<0.05$, we reject H0 and conclude that there is evidence to support the claim that controlling for all other variables, there exists a linear relationship between operating expenses and rental rate.
- row2: as $p>0.05$, we reject H0 and conclude that there is evidence to support the claim that controlling for all other variables, there exists a linear relationship between Tax and rental rate.
- row2: as $p<0.05$, we reject H0 and conclude that there is evidence to support the claim that controlling for all other variables, there exists a linear relationship between apartment square footage and rental rate.