Lab 19 R Script

Alexander Hernandez

11/03/2022

```
library(PASWR)
```

```
## Loading required package: lattice
```

```
library(missMDA)
```

```
## Warning: package 'missMDA' was built under R version 4.2.2
```

1) Biochemical Oxygen Demand from Holston River

a) Fit the regression model relating BOD to time

```
bod = read.table("C:\\repos\\STAT 50001\\Lab 19\\BOD.txt",
                 header=TRUE)
attach(bod)
```

```
## The following object is masked from package:datasets:
##
##     BOD
```

```
model1 = lm(BOD ~ Days)
model1
```

```
##
## Call:
## lm(formula = BOD ~ Days)
##
## Coefficients:
## (Intercept)         Days
##      0.6578       0.1781
```

```
# BOD = 0.6578 + 0.1781(Days)
```

b) What is the variance?

```
summary(model1)$sigma**2
```

```
## [1] 0.08253036
```

c) What is the expected BOD level at 15 days? 90% conf+pred int

```
predict(model1, data.frame(Days=15), interval="conf", level=0.9)
```

```
##        fit      lwr      upr
## 1 3.328639 3.125933 3.531346
```

```
predict(model1, data.frame(Days=15), interval="pred",  level=0.9)
```

```
##        fit      lwr      upr
## 1 3.328639 2.764355 3.892924
```

```
# With a 90% confidence interval (3.125933, 3.531346)
# and a 90% prediction interval of (2.764355, 3.892924),
# the predicted, fit value is 3.328639.
```

d) What change in mean BOD is expected when time changes by 3 days?

```
0.1781 * 3
```

```
## [1] 0.5343
```

```
# Days * 3
```

2) missMDA for Air Polution

a) Import the dataset

```
data(ozone)
```

b) Generate the list of variables included in the data

```
colnames(ozone)
```

```
##  [1] "maxO3"  "T9"     "T12"    "T15"    "Ne9"    "Ne12"   "Ne15"   "Vx9"
##  [9] "Vx12"   "Vx15"   "maxO3v" "vent"   "pluie"
```

c) Create a subset of the data with only the first 11 variables

```
oz = subset(ozone[0:11])
```

d) Fit a multiple linear regression model for maxO3 as a response var

```
attach(oz)
model2 = lm(maxO3 ~   T9 +  T12 +  T15 +  Ne9 +   Ne12 +
                    Ne15 +  Vx9 + Vx12 + Vx15 + maxO3v)
summary(model2)
```

```
##
## Call:
## lm(formula = maxO3 ~ T9 + T12 + T15 + Ne9 + Ne12 + Ne15 + Vx9 +
##     Vx12 + Vx15 + maxO3v)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -34.283  -9.348  -1.107   8.886  24.448
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.9941    25.8868   0.116   0.9089
## T9            6.0162     3.7342   1.611   0.1208
## T12          -4.2879     3.7180  -1.153   0.2606
## T15           2.2675     3.4019   0.667   0.5117
## Ne9          -1.2472     2.1251  -0.587   0.5630
## Ne12          0.1917     2.4544   0.078   0.9384
## Ne15         -0.2909     2.3791  -0.122   0.9037
## Vx9           3.1040     2.5347   1.225   0.2331
## Vx12         -0.8650     2.2006  -0.393   0.6979
## Vx15         -0.3501     2.0442  -0.171   0.8655
## maxO3v        0.3023     0.1336   2.263   0.0334 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 15.83 on 23 degrees of freedom
##   (78 observations deleted due to missingness)
## Multiple R-squared:  0.7757, Adjusted R-squared:  0.6781
## F-statistic: 7.952 on 10 and 23 DF,  p-value: 2.156e-05
```

```
# Only 'max03v' is significant with a 95% confidence.
```

3) Realtor

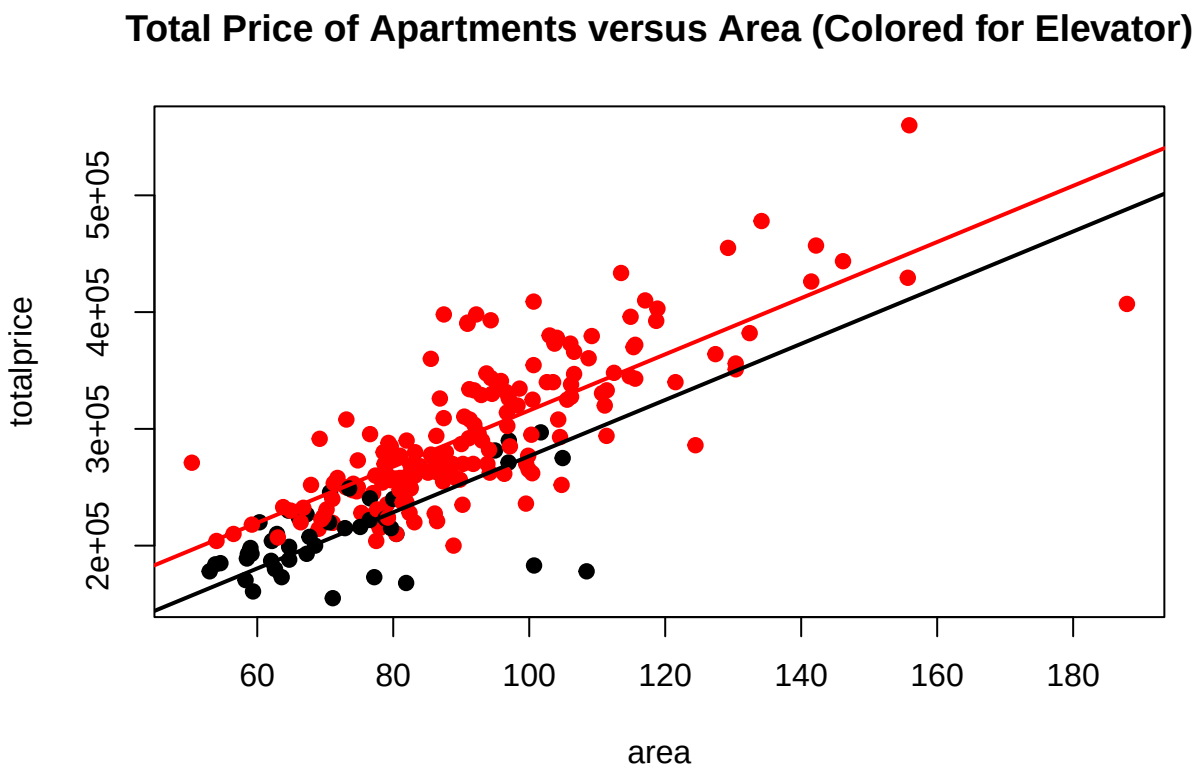
```
attach(vit2005)
plot(totalprice ~ area,
     col=ifelse(elevator=="1", "red", "black"),
     main= "Total Price of Apartments versus Area (Colored for Elevator)",
     pch=19)

model3 = lm(totalprice ~ area + elevator)
model3
```

```
##
## Call:
## lm(formula = totalprice ~ area + elevator)
##
## Coefficients:
## (Intercept)         area     elevator
##       36174         2405        39091
```

```
# totalprice = 36174 + 2405(area) + 39091(elevator)

abline(36174,       2405, col="black", lwd=2)
abline(36174+39091, 2405, col="red", lwd=2)
```



```
summary(model3)
```

```
##
## Call:
## lm(formula = totalprice ~ area + elevator)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -120265  -20224   -2567   18281  112406
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  36173.6    11434.8   3.163  0.00178 **
## area          2405.4      136.3  17.652  < 2e-16 ***
## elevator     39091.1     7022.8   5.566 7.71e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 38240 on 215 degrees of freedom
## Multiple R-squared:  0.6983, Adjusted R-squared:  0.6955
## F-statistic: 248.8 on 2 and 215 DF,  p-value: < 2.2e-16
```

```
 # Both 'area' and presence of 'elevator' have a significant impact
# on apartment pricing in Victoria, Spain.
```