STAT 652 Assignment 1

Dhruv Patel

13/10/2021

Lecture 4 Application B
1. Once zone and make have been converted to factors, run the linear regression with per as the response and the other six variables as explanatory.
(a) Create a summary of the lm object.
i. Although you fit a model with 6 variables, how many parameters are estimated?
Answer:

```
ins = read.csv('/Users/dhruv/Downloads/Insurance-1.csv',header=TRUE)
ins$zone = as.factor(ins$zone)
ins$make=as.factor(ins$make)

ins = ins[ins$claims>0,]
dim(ins)
```

```
## [1] 1797    7
```

```
ins_per_lm = lm(per ~ ., data = ins)
summary(ins_per_lm)
```

```
##
## Call:
## lm(formula = per ~ ., data = ins)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.0994 -0.7170  0.0734  0.8393  3.7574
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.186e+01  1.321e-01  89.770  < 2e-16 ***
## km          -3.434e-01  2.064e-02 -16.641  < 2e-16 ***
## zone2       -1.376e-01  9.717e-02  -1.416    0.157
## zone3       -2.143e-02  9.753e-02  -0.220    0.826
## zone4        4.317e-01  9.692e-02   4.454 8.95e-06 ***
## zone5       -1.042e+00  1.043e-01  -9.983  < 2e-16 ***
## zone6       -4.440e-01  1.009e-01  -4.401 1.14e-05 ***
## zone7       -2.862e+00  1.378e-01 -20.767  < 2e-16 ***
## bonus        2.301e-01  1.405e-02  16.381  < 2e-16 ***
## make2       -1.403e+00  1.140e-01 -12.314  < 2e-16 ***
## make3       -1.710e+00  1.189e-01 -14.382  < 2e-16 ***
```

```
## make4       -1.834e+00  1.240e-01 -14.789  < 2e-16 ***
## make5       -1.317e+00  1.138e-01 -11.568  < 2e-16 ***
## make6       -8.253e-01  1.129e-01  -7.312 3.95e-13 ***
## make7       -1.716e+00  1.153e-01 -14.878  < 2e-16 ***
## make8       -2.070e+00  1.199e-01 -17.260  < 2e-16 ***
## make9        1.459e+00  1.209e-01  12.071  < 2e-16 ***
## insured     -5.724e-05  1.151e-05  -4.975 7.15e-07 ***
## claims       3.029e-03  3.519e-04   8.608  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.179 on 1778 degrees of freedom
## Multiple R-squared:  0.6477, Adjusted R-squared:  0.6442
## F-statistic: 181.6 on 18 and 1778 DF,  p-value: < 2.2e-16
```

```
length(ins_per_lm$coefficients)
```

```
## [1] 19
```

ii. What is the intercept of the regression when make and zone are both at their first level, 1?
Answer: To find intercept for first level everything else will be zero. So finally we will get intercept = 11.86

iii. What is the intercept of the regression when make and zone are both at their last levels, 9 and 7, respectively?
Answer: To find intercept at make9 and zone7 will be (make7) = (Intercept) + (Intercept at make9) + (Intercept at zone 7) = 11.86 - 2.862 - 1.716 = 7.282