Lecture 4: Applications B

Yufei Yin

B. Categorical Explanatories

1.

```
# read data
ins = read.csv("Insurance.csv")

# convert categorical variables to factors
ins$zone = as.factor(ins$zone)
ins$make = as.factor(ins$make)

# filter claims larger than 0
ins = ins[ins$claims>0,]

# 1797 observations
dim(ins)
```

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

(a)

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

```
##
## 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
```

i.

```
coef(mod)
```

```
##   (Intercept)            km         zone2         zone3         zone4
##  1.186200e+01 -3.434143e-01 -1.375833e-01 -2.143003e-02  4.316718e-01
##         zone5         zone6         zone7         bonus         make2
## -1.041544e+00 -4.440303e-01 -2.862106e+00  2.300740e-01 -1.403389e+00
##         make3         make4         make5         make6         make7
## -1.709548e+00 -1.834010e+00 -1.316946e+00 -8.253261e-01 -1.716116e+00
##         make8         make9       insured        claims
## -2.069770e+00  1.459262e+00 -5.724293e-05  3.029138e-03
```

```
length(coef(mod))
```

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

There are 19 parameters are estimated.

ii.

```
cat(coef(mod)["(Intercept)"])
```

```
## 11.862
```

When `make` and `zone` are both at their first level, 1 the intercept of the regression is 11.862.

iii.

```
cat(coef(mod)["(Intercept)"] + coef(mod)["make9"] + coef(mod)["zone7"])
```

```
## 10.45915
```

When `make` and `zone` are both at their last levels, 9 and 7 respectively the intercept of the regression is 10.45915.