Homework 7 Problem 3

Lucy Lin

2022-10-20

```
data = read.csv('kc_house_data.csv')
data$condition = as.factor(data$condition)
fit = lm(log(price)~(sqft_living + waterfront + condition)^2, data = data)
```

Problem 3a There are two levels for waterfront and no waterfront (level 0) is the baseline by default. There are five levels for condition and 1 is the baseline by default.

```
summary(fit)
```

```
##
## Call:
## lm(formula = log(price) ~ (sqft_living + waterfront + condition)^2,
##     data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.72335 -0.28013  0.01822  0.25854  1.39273
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)               1.180e+01  1.576e-01  74.892  < 2e-16 ***
## sqft_living               5.362e-04  1.159e-04   4.626 3.75e-06 ***
## waterfront                1.101e+00  3.785e-01   2.908  0.00364 **
## condition2                1.260e-01  1.724e-01   0.731  0.46503
## condition3                4.349e-01  1.578e-01   2.756  0.00586 **
## condition4                3.926e-01  1.581e-01   2.483  0.01305 *
## condition5                3.602e-01  1.593e-01   2.261  0.02377 *
## sqft_living:waterfront   -4.883e-05  1.887e-05  -2.587  0.00967 **
## sqft_living:condition2   -1.073e-04  1.247e-04  -0.860  0.38980
## sqft_living:condition3   -1.578e-04  1.160e-04  -1.360  0.17371
## sqft_living:condition4   -1.211e-04  1.161e-04  -1.043  0.29698
## sqft_living:condition5   -4.774e-05  1.164e-04  -0.410  0.68170
## waterfront:condition2    -2.897e-01  5.660e-01  -0.512  0.60877
## waterfront:condition3    -2.946e-01  3.827e-01  -0.770  0.44138
## waterfront:condition4    -3.472e-01  3.832e-01  -0.906  0.36498
## waterfront:condition5    -4.587e-01  3.887e-01  -1.180  0.23806
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3709 on 21597 degrees of freedom
## Multiple R-squared:  0.5047, Adjusted R-squared:  0.5044
## F-statistic:  1467 on 15 and 21597 DF,  p-value: < 2.2e-16
```

```
data$waterfront = as.factor(data$waterfront)
data$condition = as.factor(data$condition)
levels(data$waterfront)
```

```
## [1] "0" "1"
```

```
levels(data$condition)
```

```
## [1] "1" "2" "3" "4" "5"
```

Problem 3b 21613 (observations) - 16 (lines of coefficients in summary or parameters) = 21597 DF

Problem 3c The fitted model for waterfront = 0 and condition = 1 is: Y = 1 + sqft_living = 1.180e+01 + 5.362e-04(X) The fitted model for waterfront = 1 and condition = 3 is: Y = 1 + sqft_living + waterfront + condition3 + sqft_living:waterfront + sqft_living:condition3 + waterfront:condition3 = Y = 1.180e+01 + 5.362e-04(x) + 1.101e+00(x) + 4.349e-01(X) + -4.883e-05(X) + -1.578e-04(X) + -2.946e-01(X)

Problem 3d The interaction waterfront:condition is not significant (high probability that our F critical value can be exceeded; no stars). We cannot conclude that sqft_living:condition is significant in the model because the model with sqft_living:condition is the nested model in waterfront:condition and therefore does not take waterfront:condition into account.