Assignment #3

April 2020

Part 1)

1) 67 + 4 + 1 = 72 number of observations

2) H0:B1 = 0

Ha:B1 = 0

3) 2.186/0.4104 = 5.3265. a = .01 = 2.6524 < t1 which means we reject H0, and the null hypothesis of B1 = 0. X is then stated as a valid indicator for Y.

4) R2 = SSR/SSt = SSR/SSR + SSE = (1974.53 + 118.8642 + 32.4701 + .4356) / (1974.53 + 118.8642 + 32.4701 + .4356 + 630.36) = .7713. Variation is 77.13%.

5) .7713 – (1 - .7713) * 4/(72-4-1) = .7577. Here, adjusted R-squared is different from R-squared values because R-squared assumes that all independent variables are necessary for explaining variation in dependent variables, while adjusted R-squared penalizes and lowers the score for adding independent variables that do not help in predicant dependent variables.

6) H0:B1 = B2 = B3 = B4 = 0

Ha:B1 =/ 0 (not equal to zero), for I E 1,2,3,4

7) ((1974.53 + 118.8642 + 32.4701 + .4356)/4) / ((630.36/72-4-1)) = 56.5003

When p=4 and freedom =67, p value <.0001. This means that at least one of the slope parameters is zero and we can reject the null hypothesis.

Part 2)

8) Model 1 nests Model 2 because there are far more explanatory variables in model 2 that are no in Model 1, whilst all of the variables in Model 1 are in Model 2.

9) H0:B5 = B6 =0

Ha:Bi =/ (not equal to) 0, for I E 5, 6

10) F = ((630.36 – 572.6091)/(7 - 5)) / ((572.6091 / (72 – 6 -1)) = 3.2778

F95,2,65 = 3.1381

Again the null hypothesis is reject as model 1 is still more powerful than model 2.

11) The 10 continuous variables:

Sales Price, Year Sold, Sub Class, Lot Area, Neighborhood, Building Type, House Style, Overall Condition, Year Built, and Garage cars.

It is interesting to put these into different sets as on the surface they variables seem very disparate. However I chose to break it into: Main House characteristics: Lot Area, Building type, House Style, Garage Cars, Overall Condition, and Year Built. The other is Ancillary Housing conditions: Sales Price, Year Sold, Sub Class, Neighborhood. The reason I broke them into these two sets was the actual

structure and bones of the homes which are main house characteristics, while ancillary characteristics are variables like neighborhood that describe the are of the house and not the structure itself.

12)

```
Call:
lm(formula = SalePrice ~ LotArea + BldgType + HouseStyle + GarageCars +
    OverallCond + YearBuilt, data = new)

Residuals:
    Min      1Q  Median      3Q     Max
-224786  -33046   -7862   20073  466456

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)      -2.176e+06  9.172e+04 -23.729  < 2e-16 ***
LotArea           1.654e+00  1.328e-01  12.459  < 2e-16 ***
BldgType2fmCon   -4.067e+03  7.009e+03  -0.580   0.5618
BldgTypeDuplex   -2.630e+04  5.346e+03  -4.919 9.17e-07 ***
BldgTypeTwnhs    -3.862e+04  5.662e+03  -6.821 1.09e-11 ***
BldgTypeTwnhsE   -1.082e+04  3.893e+03  -2.778   0.0055 **
HouseStyle1.5Unf -1.563e+04  1.256e+04  -1.245   0.2134
HouseStyle1Story -1.819e+04  3.624e+03  -5.019 5.50e-07 ***
HouseStyle2.5Fin  9.032e+04  1.906e+04   4.737 2.27e-06 ***
HouseStyle2.5Unf  5.057e+04  1.131e+04   4.471 8.08e-06 ***
HouseStyle2Story -1.932e+03  3.888e+03  -0.497   0.6192
HouseStyleSFoyer -4.126e+04  6.882e+03  -5.994 2.29e-09 ***
HouseStyleSLvl   -3.623e+04  5.790e+03  -6.258 4.48e-10 ***
GarageCars        4.219e+04  1.599e+03  26.391  < 2e-16 ***
OverallCond       8.116e+03  9.651e+02   8.410  < 2e-16 ***
YearBuilt         1.134e+03  4.683e+01  24.222  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 53110 on 2913 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.5605,    Adjusted R-squared:  0.5582
F-statistic: 247.7 on 15 and 2913 DF,  p-value: < 2.2e-16
```

B1 is Lot Area

H0:B1 = 0

Ha:B1 =/ 0

T1 = 1.654 / 1.328 = 1.2455

t-test with a = .05

ta = 69.708>|t1|, reject the null

B2 is GarageCars

T1 = 4.219/1.599 = 2.6385

Ta = 125.72 >|t1|, reject the null

B3 is Overall Condition

T1 = 8.116/9.651 = .8409

Ta = 270.91> |t1|, reject the null

B4 is YearBuilt

T1 = 1.134/4.683 = .2421

Ta = 3528.1 > |t1|, reject the null

b)

H0: Beata =Beta2 = Beta3 = Beta4 = o versus Ha:B1 for E 1,2,3,4

F-statistic = 247.7, reject the null

Part 4)

13)

a)

```
Call:
lm(formula = SalePrice ~ LotArea + BldgType + HouseStyle + GarageCars +
    OverallCond + YearBuilt + YrSold + SubClass + Neighborhood,
    data = new)

Residuals:
    Min      1Q  Median      3Q     Max
-184578  -23443   -3146   16341  401435

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)         -5.251e+05  1.223e+06  -0.429 0.667761
LotArea              1.340e+00  1.139e-01  11.762  < 2e-16 ***
BldgType2fmCon       1.532e+04  1.320e+04   1.161 0.245862
BldgTypeDuplex      -8.055e+03  6.049e+03  -1.332 0.183107
BldgTypeTwnhs       -4.832e+04  1.058e+04  -4.566 5.17e-06 ***
BldgTypeTwnhsE      -2.724e+04  9.330e+03  -2.920 0.003529 **
HouseStyle1.5Unf    -1.761e+04  1.020e+04  -1.726 0.084460 .
HouseStyle1Story    -1.564e+04  3.716e+03  -4.207 2.66e-05 ***
HouseStyle2.5Fin     8.210e+04  1.575e+04   5.213 1.99e-07 ***
HouseStyle2.5Unf     4.175e+04  9.230e+03   4.523 6.35e-06 ***
HouseStyle2Story     5.318e+03  3.501e+03   1.519 0.128831
HouseStyleSFoyer    -1.289e+04  6.205e+03  -2.078 0.037778 *
HouseStyleSLvl      -9.941e+03  5.609e+03  -1.772 0.076419 .
GarageCars           2.747e+04  1.396e+03  19.679  < 2e-16 ***
OverallCond          8.107e+03  7.912e+02  10.246  < 2e-16 ***
YearBuilt            9.609e+02  6.102e+01  15.748  < 2e-16 ***
YrSold              -6.262e+02  6.052e+02  -1.035 0.300876
SubClass            -1.522e+02  8.813e+01  -1.727 0.084309 .
NeighborhoodBlueste -2.234e+04  1.611e+04  -1.387 0.165585
NeighborhoodBrDale  -2.127e+04  1.242e+04  -1.711 0.087100 .
NeighborhoodBrkSide -3.054e+04  1.043e+04  -2.927 0.003452 **
NeighborhoodClearCr -8.642e+03  1.116e+04  -0.775 0.438599
NeighborhoodCollgCr -3.262e+04  9.077e+03  -3.593 0.000332 ***
NeighborhoodCrawfor  2.200e+04  1.003e+04   2.192 0.028450 *
NeighborhoodEdwards -3.862e+04  9.584e+03  -4.029 5.74e-05 ***
NeighborhoodGilbert -5.227e+04  9.461e+03  -5.525 3.59e-08 ***
NeighborhoodGreens   4.281e+04  1.744e+04   2.455 0.014139 *
NeighborhoodGrnHill  1.286e+05  3.136e+04   4.101 4.22e-05 ***
NeighborhoodIDOTRR  -4.024e+04  1.065e+04  -3.780 0.000160 ***
NeighborhoodLandmrk -2.969e+04  4.392e+04  -0.676 0.499144
NeighborhoodMeadowV -2.207e+04  1.150e+04  -1.920 0.054983 .
NeighborhoodMitchel -4.844e+04  9.673e+03  -5.007 5.85e-07 ***
NeighborhoodNAmes   -3.862e+04  9.251e+03  -4.174 3.08e-05 ***
NeighborhoodNoRidge  6.752e+04  1.020e+04   6.623 4.19e-11 ***
NeighborhoodNPkVill -1.133e+04  1.248e+04  -0.908 0.363964
NeighborhoodNridgHt  7.462e+04  9.094e+03   8.206 3.41e-16 ***
NeighborhoodNWAmes  -3.200e+04  9.603e+03  -3.332 0.000873 ***
NeighborhoodOldTown -3.181e+04  1.013e+04  -3.142 0.001698 **
NeighborhoodSawyer  -4.867e+04  9.672e+03  -5.031 5.17e-07 ***
NeighborhoodSawyerW -3.745e+04  9.504e+03  -3.941 8.32e-05 ***
NeighborhoodSomerst  1.166e+03  9.025e+03   0.129 0.897229
NeighborhoodStoneBr  9.792e+04  1.023e+04   9.572  < 2e-16 ***
NeighborhoodSWISU   -1.647e+04  1.153e+04  -1.428 0.153387
NeighborhoodTimber  -2.664e+03  1.010e+04  -0.264 0.791965
NeighborhoodVeenker  2.517e+04  1.225e+04   2.054 0.040034 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

B1 is Year Sold

H0:B1 = 0

Ha:B1 =/ 0

T1 = -6.262 / 6.052 = 1.0347

t-test with a = .05

ta = 82456>|t1|, reject the null

B2 is Subclass

T1 = -1.522 / 8.813 = -.1727

t-test with a = .05

ta = 72.84>|t1|, reject the null

b)

F-statistic = 166.9, reject the null

H0: Beta =Beta2 = Beta3 = Beta4 = Beta5 = Beta6 = 0 versus Ha:B1 for E 1,2,3,4, 5, 6

14)

H0: Beta5 =Beta6 = Beta7 = Beta8 = Beta9 = Beta10 - 0 versus Ha:B1 for E 5,6,7,8,9,10

F-statistic = 127.62, reject the null