**Step 8: Include your R code**

Appendix

**Step 1: Introduction**

# Load the commander

> library("Rcmdr")

Dataset <- readXL("//apporto.com/dfs/UALR/Users/saoyedotun\_ualr/Desktop/AmesInput.xlsx", rownames=FALSE,

header=TRUE, na="", sheet="AmesforR (2)", stringsAsFactors=TRUE)

summary(Dataset)

Q: Does everything look okay

A: Yes

**Step 2: Ordinary Least Squares**

ordinary\_least\_squares <- lm(saleprice ~ bedroomabvgr, data=Dataset)

summary(ordinary\_least\_squares)

Call:

lm(formula = saleprice ~ bedroomabvgr, data = Dataset)

Residuals:

Min 1Q Median 3Q Max

-143109 -52672 -17719 31891 555510

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 133966 7492 17.881 < 2e-16 \*\*\*

bedroomabvgr 16381 2514 6.516 9.93e-11 \*\*\*

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 78340 on 1458 degrees of freedom

Multiple R-squared: 0.0283, Adjusted R-squared: 0.02763

F-statistic: 42.46 on 1 and 1458 DF, p-value: 9.927e-11

Q: Are the predictor variables significant? Interpret the effect of beds on sales price.

A: Yes, bedroomsabvgr is a statistically significant predictor on the sales price. For every additional bedroom to a home, the price of a home increases by 16,381.

Step 3: Ordinary Least Squares Model 2

ordinary\_least\_squares\_model\_2 <- lm(saleprice ~ yrsold + fullbath + lotarea, data = Dataset)

summary(ordinary\_least\_squares\_model\_2)

Call:

lm(formula = saleprice ~ yrsold + fullbath + lotarea, data = Dataset)

Residuals:

Min 1Q Median 3Q Max

-225996 -35133 -6785 21608 445286

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 1916033.7349 2533482.5850 0.756 0.45

yrsold -932.5626 1261.7541 -0.739 0.46

fullbath 77239.8304 3065.8527 25.194 <2e-16 \*\*\*

lotarea 1.5609 0.1692 9.225 <2e-16 \*\*\*

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 63990 on 1456 degrees of freedom

Multiple R-squared: 0.3525, Adjusted R-squared: 0.3512

F-statistic: 264.2 on 3 and 1456 DF, p-value: < 2.2e-16

Q: Are the predictor variables significant?

Interpret the effect of Lot Area and Full Bath.

A: Yes, fullbath and lotarea are statistically significant but yrsold is not. For every full bath increase the price of a home increases by 77,239.83 and for every additional lot area (square foot) to a home, the price of a home increase by 1.56

**Step 4: Ordinary Least Squares Model 3**

ordinary\_least\_squares\_model\_3 <- lm(saleprice ~ yrsold + bedroomsabvgr + lotarea + exterqual1 + exterqual2 + exterqual3 + exterqual4, data = Dataset)

summary(ordinary\_least\_squares\_model\_3)

Call:

lm(formula = saleprice ~ yrsold + bedroomabvgr + lotarea + exterqual1 +

exterqual2 + exterqual3 + exterqual4, data = Dataset)

Residuals:

Min 1Q Median 3Q Max

-290591 -29265 -4896 23605 486135

Coefficients: (1 not defined because of singularities)

Estimate Std. Error t value Pr(>|t|)

(Intercept) 3056558.5101 2117761.3581 1.443 0.14915

yrsold -1481.0740 1054.6699 -1.404 0.16044

bedroomabvgr 15536.9778 1726.0374 9.002 < 2e-16 \*\*\*

lotarea 1.6129 0.1412 11.423 < 2e-16 \*\*\*

exterqual1 218372.4909 7630.9605 28.617 < 2e-16 \*\*\*

exterqual2 -44678.0106 14396.4852 -3.103 0.00195 \*\*

exterqual3 87475.5273 2992.9232 29.227 < 2e-16 \*\*\*

exterqual4 NA NA NA NA

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 53260 on 1453 degrees of freedom

Multiple R-squared: 0.5524, Adjusted R-squared: 0.5505

F-statistic: 298.8 on 6 and 1453 DF, p-value: < 2.2e-16

Q: Are the predictor variables significant?

Interpret the effect of the External Quality variables on Sales price.

A: Yes, bedroomabvgr, and lotarea are statistically significant but yrsold is not. The external quality variables 1 and 3 increases the sale price of the home by 218,372.49 and 87,475.53, external quality variable 3 reduces the sale price of a home and external quality variable 4 does not have an effect on the sale price of a home.

**Step 5: Ordinary Least Squares Model 4**

Dataset$loglotarea <- with(Dataset, log(lotarea))

ordinary\_least\_squares\_model\_4 <- lm(saleprice ~ yrsold + fullbath + loglotarea + exterqual1 + exterqual2 + exterqual3 + exterqual4, data = Dataset)

summary(ordinary\_least\_squares\_model\_4)

Call:

lm(formula = saleprice ~ yrsold + fullbath + loglotarea + exterqual1 +

exterqual2 + exterqual3 + exterqual4, data = Dataset)

Residuals:

Min 1Q Median 3Q Max

-272747 -27756 -2218 21562 447783

Coefficients: (1 not defined because of singularities)

Estimate Std. Error t value Pr(>|t|)

(Intercept) 2240348.1 1933929.2 1.158 0.2469

yrsold -1263.0 962.8 -1.312 0.1898

fullbath 37979.5 2691.4 14.111 <2e-16 \*\*\*

loglotarea 42712.9 2516.0 16.977 <2e-16 \*\*\*

exterqual1 179952.7 7237.7 24.863 <2e-16 \*\*\*

exterqual2 -38819.9 13146.9 -2.953 0.0032 \*\*

exterqual3 65336.5 3076.6 21.237 <2e-16 \*\*\*

exterqual4 NA NA NA NA

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 48620 on 1453 degrees of freedom

Multiple R-squared: 0.6269, Adjusted R-squared: 0.6254

F-statistic: 406.9 on 6 and 1453 DF, p-value: < 2.2e-16

Q: Are the predictor variables significant?

Interpret the effect of the Log Lot Area variables on Sales price

A: Yes, fullbath, and loglotarea are statistically significant but yrsold is not. For every full bath increase the price of a home increases by 37,979.5 and for every additional loglotarea (square foot) to a home, the price of a home increase by 42,712.9

**Step 6: Ordinary Least Squares Model 5**

Dataset$logsaleprice <- with(Dataset, log(saleprice))

ordinary\_least\_squares\_model\_5 <- lm(logsaleprice ~ yrsold + bedroomabvgr + lotarea + loglotarea + exterqual1 + exterqual2 + exterqual3 + exterqual4, data = Dataset)

summary(ordinary\_least\_squares\_model\_5)

Call:

lm(formula = logsaleprice ~ yrsold + bedroomabvgr + lotarea +

loglotarea + exterqual1 + exterqual2 + exterqual3 + exterqual4,

data = Dataset)

Residuals:

Min 1Q Median 3Q Max

-1.28565 -0.13634 0.00534 0.15326 1.00640

Coefficients: (1 not defined because of singularities)

Estimate Std. Error t value Pr(>|t|)

(Intercept) 20.23610906974 10.20073068828 1.984 0.0475 \*

yrsold -0.00531903344 0.00507788080 -1.047 0.2950

bedroomabvgr 0.06723437151 0.00865472762 7.769 1.49e-14 \*\*\*

lotarea -0.00000005958 0.00000094522 -0.063 0.9497

loglotarea 0.22979653560 0.01899387047 12.098 < 2e-16 \*\*\*

exterqual1 0.85043593191 0.03701063670 22.978 < 2e-16 \*\*\*

exterqual2 -0.46937282737 0.06930044995 -6.773 1.83e-11 \*\*\*

exterqual3 0.46689077039 0.01442327099 32.371 < 2e-16 \*\*\*

exterqual4 NA NA NA NA

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2564 on 1452 degrees of freedom

Multiple R-squared: 0.59, Adjusted R-squared: 0.588

F-statistic: 298.5 on 7 and 1452 DF, p-value: < 2.2e-16

Q: Are the predictor variables significant?

Interpret the effect of the Log Lot Area variables on Sales price

Based on the models you have estimated, which model do you think is the best? Why?

A: Yes, bedroomabvgr, and loglotarea are statistically significant but yrsold and lotarea are not. For every loglotarea to a home, the sale price of a home increase by 0.22

**Step 7: Ordinary Least Squares Model 6**

ordinary\_least\_squares\_model\_6 <- lm(bedroomabvgr ~ yrsold + fullbath + loglotarea + exterqual1 + exterqual2 + exterqual3 + exterqual4, data = Dataset)

summary(ordinary\_least\_squares\_model\_6)

Call:

lm(formula = bedroomabvgr ~ yrsold + fullbath + loglotarea +

exterqual1 + exterqual2 + exterqual3 + exterqual4, data = Dataset)

Residuals:

Min 1Q Median 3Q Max

-2.4901 -0.5162 0.0611 0.3809 4.8103

Coefficients: (1 not defined because of singularities)

Estimate Std. Error t value Pr(>|t|)

(Intercept) 22.73046 28.41033 0.800 0.424

yrsold -0.01192 0.01414 -0.843 0.399

fullbath 0.67161 0.03954 16.986 < 2e-16 \*\*\*

loglotarea 0.35034 0.03696 9.479 < 2e-16 \*\*\*

exterqual1 -0.71741 0.10632 -6.747 2.17e-11 \*\*\*

exterqual2 -0.21037 0.19313 -1.089 0.276

exterqual3 -0.41686 0.04520 -9.223 < 2e-16 \*\*\*

exterqual4 NA NA NA NA

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7143 on 1453 degrees of freedom

Multiple R-squared: 0.2365, Adjusted R-squared: 0.2333

F-statistic: 75 on 6 and 1453 DF, p-value: < 2.2e-16

Q: Rerun Equation 5 with another predictor variable of your own choosing.

A: with bedroomabvgr as dependent variable, it can be said that fullbath, loglotarea are statistically significant predictors, while yrsold is not.