205615894_stats101a_hw7

Takao

2/23/2022

```
winequality <- read.csv("winequality-red.csv")
head(winequality)
```

```
##   fixed.acidity volatile.acidity citric.acid residual.sugar chlorides
## 1           7.4             0.70        0.00            1.9     0.076
## 2           7.8             0.88        0.00            2.6     0.098
## 3           7.8             0.76        0.04            2.3     0.092
## 4          11.2             0.28        0.56            1.9     0.075
## 5           7.4             0.70        0.00            1.9     0.076
## 6           7.4             0.66        0.00            1.8     0.075
##   free.sulfur.dioxide total.sulfur.dioxide density   pH sulphates alcohol
## 1                  11                   34  0.9978 3.51      0.56     9.4
## 2                  25                   67  0.9968 3.20      0.68     9.8
## 3                  15                   54  0.9970 3.26      0.65     9.8
## 4                  17                   60  0.9980 3.16      0.58     9.8
## 5                  11                   34  0.9978 3.51      0.56     9.4
## 6                  13                   40  0.9978 3.51      0.56     9.4
##   quality
## 1       5
## 2       5
## 3       5
## 4       6
## 5       5
## 6       5
```

1

```
mlr.model1 <- lm(quality~fixed.acidity + volatile.acidity + citric.acid + residual.sugar + chlorides + f
mlr.model1
```

```
##
## Call:
## lm(formula = quality ~ fixed.acidity + volatile.acidity + citric.acid +
##     residual.sugar + chlorides + free.sulfur.dioxide + total.sulfur.dioxide +
##     density + pH + sulphates + alcohol, data = winequality)
##
## Coefficients:
##          (Intercept)         fixed.acidity      volatile.acidity
##            22.207788              0.024800              -1.080667
```

```
##          citric.acid        residual.sugar             chlorides
##            -0.179253              0.016264            -1.875460
##  free.sulfur.dioxide  total.sulfur.dioxide               density
##             0.004521             -0.003304           -18.130969
##                   pH             sulphates               alcohol
##            -0.410411              0.917006             0.275673
```

2

```
summary(mlr.model1)
```

```
##
## Call:
## lm(formula = quality ~ fixed.acidity + volatile.acidity + citric.acid +
##     residual.sugar + chlorides + free.sulfur.dioxide + total.sulfur.dioxide +
##     density + pH + sulphates + alcohol, data = winequality)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.69261 -0.36656 -0.04891  0.45239  2.02508
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)           2.221e+01  2.124e+01   1.045   0.2960
## fixed.acidity         2.480e-02  2.607e-02   0.951   0.3415
## volatile.acidity     -1.081e+00  1.213e-01  -8.909  < 2e-16 ***
## citric.acid          -1.793e-01  1.476e-01  -1.214   0.2249
## residual.sugar        1.626e-02  1.503e-02   1.082   0.2793
## chlorides            -1.875e+00  4.201e-01  -4.465 8.59e-06 ***
## free.sulfur.dioxide   4.521e-03  2.187e-03   2.067   0.0389 *
## total.sulfur.dioxide -3.304e-03  7.321e-04  -4.513 6.88e-06 ***
## density              -1.813e+01  2.168e+01  -0.836   0.4032
## pH                   -4.104e-01  1.920e-01  -2.137   0.0327 *
## sulphates             9.170e-01  1.145e-01   8.007 2.26e-15 ***
## alcohol               2.757e-01  2.654e-02  10.388  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6488 on 1582 degrees of freedom
## Multiple R-squared:  0.3603, Adjusted R-squared:  0.3559
## F-statistic:    81 on 11 and 1582 DF,  p-value: < 2.2e-16
```

This F-statistic is shown at the bottom of the output above. The p-value is calculated using a reference distribution F with 11 and 1582 degrees of freedom. At a level of $\alpha = 0.05$, we *reject* the null hypothesis since the p-value is less than 2.2e-16 which is smaller than this value of α. Therefore, we conclude that at least one of the predictors in the model has a significant explanatory power.

3

```
summary(mlr.model1)
```

```
##
## Call:
## lm(formula = quality ~ fixed.acidity + volatile.acidity + citric.acid +
##     residual.sugar + chlorides + free.sulfur.dioxide + total.sulfur.dioxide +
##     density + pH + sulphates + alcohol, data = winequality)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.69261 -0.36656 -0.04891  0.45239  2.02508
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)           2.221e+01  2.124e+01   1.045   0.2960
## fixed.acidity         2.480e-02  2.607e-02   0.951   0.3415
## volatile.acidity     -1.081e+00  1.213e-01  -8.909  < 2e-16 ***
## citric.acid          -1.793e-01  1.476e-01  -1.214   0.2249
## residual.sugar        1.626e-02  1.503e-02   1.082   0.2793
## chlorides            -1.875e+00  4.201e-01  -4.465 8.59e-06 ***
## free.sulfur.dioxide   4.521e-03  2.187e-03   2.067   0.0389 *
## total.sulfur.dioxide -3.304e-03  7.321e-04  -4.513 6.88e-06 ***
## density              -1.813e+01  2.168e+01  -0.836   0.4032
## pH                   -4.104e-01  1.920e-01  -2.137   0.0327 *
## sulphates             9.170e-01  1.145e-01   8.007 2.26e-15 ***
## alcohol               2.757e-01  2.654e-02  10.388  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6488 on 1582 degrees of freedom
## Multiple R-squared:  0.3603, Adjusted R-squared:  0.3559
## F-statistic:    81 on 11 and 1582 DF,  p-value: < 2.2e-16
```

1. The fixed acidity does not have a significant association with the wine quality, when the predictors volatile acidity, citric acid, residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, density, pH, sulphates, and alcohol are in the model.
2. The volatile acidity has a significant association with the wine quality, when the predictors fixed acidity, citric acid, residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, density, pH, sulphates, and alcohol are in the model.
3. Citric acid does not have a significant association with the wine quality, when the predictors fixed acidity, volatile acidity, residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, density, pH, sulphates, and alcohol are in the model.
4. Residual sugar does not have a significant association with the wine quality, when the predictors fixed acidity, volatile acidity, citric acid, chlorides, free sulfur dioxide, total sulfur dioxide, density, pH, sulphates, and alcohol are in the model.
5. Chlorides have a significant association with the wine quality, when the predictors fixed acidity, volatile acidity, citric acid, residual sugar, free sulfur dioxide, total sulfur dioxide, density, pH, sulphates, and alcohol are in the model.
6. Free sulfur dioxide does not have a significant association with the wine quality, when the predictors fixed acidity, volatile acidity, citric acid, residual sugar, chlorides, total sulfur dioxide, density, pH, sulphates, and alcohol are in the model.
7. Total sulfur dioxide has a significant association with the wine quality, when the predictors fixed acidity,

volatile acidity, citric acid, residual sugar, chlorides, free sulfur dioxide, density, pH, sulphates, and alcohol are in the model.

8. Density does not have a significant association with the wine quality, when the predictors fixed acidity, volatile acidity, citric acid, residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, pH, sulphates, and alcohol are in the model.
9. pH does not have a significant association with the wine quality, when the predictors fixed acidity, volatile acidity, citric acid, residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, density, sulphates, and alcohol are in the model.
10. Sulphates does have a significant association with the wine quality, when the predictors fixed acidity, volatile acidity, citric acid, residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, density, pH, and alcohol are in the model.
11. Alcohol does have a significant association with the wine quality, when the predictors fixed acidity, volatile acidity, citric acid, residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, density, pH, and sulphates are in the model.

4

```
confint(mlr.model1, level = 0.95)
```

```
##                              2.5 %       97.5 %
## (Intercept)          -1.946269e+01 63.878270339
## fixed.acidity        -2.632636e-02  0.075927205
## volatile.acidity     -1.318587e+00 -0.842746528
## citric.acid          -4.688618e-01  0.110356642
## residual.sugar       -1.321025e-02  0.045737496
## chlorides            -2.699427e+00 -1.051492426
## free.sulfur.dioxide   2.305476e-04  0.008811505
## total.sulfur.dioxide -4.739889e-03 -0.001867774
## density              -6.066512e+01 24.403181166
## pH                   -7.870578e-01 -0.033763209
## sulphates             6.923631e-01  1.141648686
## alcohol               2.236199e-01  0.327725225
```

fixed acidity: (-2.632636e-02, 0.075927205) volatile acidity: (-1.318587e+00, -0.842746528) citric acid: (-4.688618e-01, 0.110356642) residual sugar: (-1.321025e-02, 0.045737496) chlorides: (-2.699427e+00, -1.051492426) free sulfur dioxide: (2.305476e-04, 0.008811505) total sulfur dioxide: (-4.739889e-03, -0.001867774) density: (-6.066512e+01, 24.403181166) pH: (-7.870578e-01, -0.033763209) sulphates: (6.923631e-01, 1.141648686) alcohol: (2.236199e-01, 0.327725225)

5

```
fasq <- winequality$fixed.acidity ^ 2
vasq <- winequality$volatile.acidity ^ 2
casq <- winequality$citric.acid ^ 2
rssq <- winequality$residual.sugar ^ 2
chsq <- winequality$chlorides ^ 2
fsdsq <- winequality$free.sulfur.dioxide ^ 2
tsdsq <- winequality$total.sulfur.dioxide ^ 2
densq <- winequality$density ^ 2
```

```r
phsq <- winequality$pH ^ 2
sulsq <- winequality$sulphates ^ 2
alcsq <- winequality$alcohol ^ 2

mlr.model2 <- lm(quality~fixed.acidity + fasq + volatile.acidity + vasq + citric.acid + casq + residual
mlr.model2
```

```
##
## Call:
## lm(formula = quality ~ fixed.acidity + fasq + volatile.acidity +
##     vasq + citric.acid + casq + residual.sugar + rssq + chlorides +
##     chsq + free.sulfur.dioxide + fsdsq + total.sulfur.dioxide +
##     tsdsq + density + densq + pH + phsq + sulphates + sulsq +
##     alcohol + alcsq, data = winequality)
##
## Coefficients:
##          (Intercept)         fixed.acidity                  fasq
##            5.681e+03             2.801e-01            -1.384e-02
##     volatile.acidity                  vasq           citric.acid
##           -3.608e-01            -4.931e-01            -5.611e-01
##                 casq        residual.sugar                  rssq
##            6.552e-01             2.160e-02            -1.824e-04
##            chlorides                  chsq   free.sulfur.dioxide
##           -1.562e+00            -3.350e-01             1.101e-02
##                fsdsq  total.sulfur.dioxide                 tsdsq
##           -1.487e-04            -4.835e-03             1.175e-05
##              density                 densq                    pH
##           -1.136e+04             5.676e+03             7.505e-01
##                 phsq             sulphates                 sulsq
##           -1.861e-01             3.705e+00            -1.579e+00
##              alcohol                 alcsq
##            3.677e-01            -5.760e-03
```

```r
summary(mlr.model2)
```

```
##
## Call:
## lm(formula = quality ~ fixed.acidity + fasq + volatile.acidity +
##     vasq + citric.acid + casq + residual.sugar + rssq + chlorides +
##     chsq + free.sulfur.dioxide + fsdsq + total.sulfur.dioxide +
##     tsdsq + density + densq + pH + phsq + sulphates + sulsq +
##     alcohol + alcsq, data = winequality)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.80021 -0.38723 -0.03681  0.43431  1.94071
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)          5.681e+03  3.580e+03   1.587  0.11274
## fixed.acidity        2.801e-01  9.491e-02   2.951  0.00321 **
## fasq                -1.384e-02  4.816e-03  -2.873  0.00412 **
## volatile.acidity    -3.608e-01  4.278e-01  -0.843  0.39914
```

```
## vasq                -4.931e-01  3.270e-01  -1.508  0.13169
## citric.acid         -5.611e-01  3.243e-01  -1.730  0.08376 .
## casq                 6.552e-01  4.786e-01   1.369  0.17121
## residual.sugar       2.160e-02  3.908e-02   0.553  0.58051
## rssq                -1.824e-04  2.926e-03  -0.062  0.95029
## chlorides           -1.562e+00  1.118e+00  -1.397  0.16249
## chsq                -3.350e-01  2.426e+00  -0.138  0.89023
## free.sulfur.dioxide  1.101e-02  5.867e-03   1.876  0.06080 .
## fsdsq               -1.487e-04  1.077e-04  -1.381  0.16754
## total.sulfur.dioxide -4.835e-03  1.774e-03  -2.726  0.00648 **
## tsdsq                1.175e-05  9.421e-06   1.247  0.21251
## density             -1.136e+04  7.181e+03  -1.582  0.11395
## densq                5.676e+03  3.600e+03   1.576  0.11514
## pH                   7.505e-01  3.312e+00   0.227  0.82077
## phsq                -1.861e-01  4.948e-01  -0.376  0.70684
## sulphates            3.705e+00  4.096e-01   9.045  < 2e-16 ***
## sulsq               -1.579e+00  2.293e-01  -6.885 8.34e-12 ***
## alcohol              3.677e-01  2.769e-01   1.328  0.18437
## alcsq               -5.760e-03  1.282e-02  -0.449  0.65333
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6375 on 1571 degrees of freedom
## Multiple R-squared:  0.3867, Adjusted R-squared:  0.3781
## F-statistic: 45.02 on 22 and 1571 DF,  p-value: < 2.2e-16
```

```
anova(mlr.model1, mlr.model2)
```

```
## Analysis of Variance Table
##
## Model 1: quality ~ fixed.acidity + volatile.acidity + citric.acid + residual.sugar +
##     chlorides + free.sulfur.dioxide + total.sulfur.dioxide +
##     density + pH + sulphates + alcohol
## Model 2: quality ~ fixed.acidity + fasq + volatile.acidity + vasq + citric.acid +
##     casq + residual.sugar + rssq + chlorides + chsq + free.sulfur.dioxide +
##     fsdsq + total.sulfur.dioxide + tsdsq + density + densq +
##     pH + phsq + sulphates + sulsq + alcohol + alcsq
##   Res.Df    RSS Df Sum of Sq    F    Pr(>F)
## 1   1582 665.90
## 2   1571 638.45 11    27.448 6.14 6.087e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The null hypothesis is that all coefficients removed from the full model or the coefficients of the quadratic components are zero. The alternative hypothesis is that at leas tone of the coefficients removed from the full model is non-zero.

The p-value from this partial F-test is smaller than our significance level $\alpha = 0.05$. Therefore, we reject the null hypothesis and conclude that the original model and the newly constructed quadratic model are statistically different models, given that all other predictors are in the model.

6

```
summary(mlr.model1)
```

```
##
## Call:
## lm(formula = quality ~ fixed.acidity + volatile.acidity + citric.acid +
##     residual.sugar + chlorides + free.sulfur.dioxide + total.sulfur.dioxide +
##     density + pH + sulphates + alcohol, data = winequality)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.69261 -0.36656 -0.04891  0.45239  2.02508
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)           2.221e+01  2.124e+01   1.045   0.2960
## fixed.acidity         2.480e-02  2.607e-02   0.951   0.3415
## volatile.acidity     -1.081e+00  1.213e-01  -8.909  < 2e-16 ***
## citric.acid          -1.793e-01  1.476e-01  -1.214   0.2249
## residual.sugar        1.626e-02  1.503e-02   1.082   0.2793
## chlorides            -1.875e+00  4.201e-01  -4.465 8.59e-06 ***
## free.sulfur.dioxide   4.521e-03  2.187e-03   2.067   0.0389 *
## total.sulfur.dioxide -3.304e-03  7.321e-04  -4.513 6.88e-06 ***
## density              -1.813e+01  2.168e+01  -0.836   0.4032
## pH                   -4.104e-01  1.920e-01  -2.137   0.0327 *
## sulphates             9.170e-01  1.145e-01   8.007 2.26e-15 ***
## alcohol               2.757e-01  2.654e-02  10.388  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6488 on 1582 degrees of freedom
## Multiple R-squared:  0.3603, Adjusted R-squared:  0.3559
## F-statistic:    81 on 11 and 1582 DF,  p-value: < 2.2e-16
```

In this case for the original model (model 1), the R^2 and adjusted R^2 values are 0.3602 and 0.3559, respectively. Recall that high values are preferred. Ideally, these values are larger than, say, 70%. Thus, we can say that the R squared values are relatively low for the original model.

```
summary(mlr.model2)
```

```
##
## Call:
## lm(formula = quality ~ fixed.acidity + fasq + volatile.acidity +
##     vasq + citric.acid + casq + residual.sugar + rssq + chlorides +
##     chsq + free.sulfur.dioxide + fsdsq + total.sulfur.dioxide +
##     tsdsq + density + densq + pH + phsq + sulphates + sulsq +
##     alcohol + alcsq, data = winequality)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.80021 -0.38723 -0.03681  0.43431  1.94071
```

```
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)           5.681e+03  3.580e+03   1.587  0.11274
## fixed.acidity         2.801e-01  9.491e-02   2.951  0.00321 **
## fasq                 -1.384e-02  4.816e-03  -2.873  0.00412 **
## volatile.acidity     -3.608e-01  4.278e-01  -0.843  0.39914
## vasq                 -4.931e-01  3.270e-01  -1.508  0.13169
## citric.acid          -5.611e-01  3.243e-01  -1.730  0.08376 .
## casq                  6.552e-01  4.786e-01   1.369  0.17121
## residual.sugar        2.160e-02  3.908e-02   0.553  0.58051
## rssq                 -1.824e-04  2.926e-03  -0.062  0.95029
## chlorides            -1.562e+00  1.118e+00  -1.397  0.16249
## chsq                 -3.350e-01  2.426e+00  -0.138  0.89023
## free.sulfur.dioxide   1.101e-02  5.867e-03   1.876  0.06080 .
## fsdsq                -1.487e-04  1.077e-04  -1.381  0.16754
## total.sulfur.dioxide -4.835e-03  1.774e-03  -2.726  0.00648 **
## tsdsq                 1.175e-05  9.421e-06   1.247  0.21251
## density              -1.136e+04  7.181e+03  -1.582  0.11395
## densq                 5.676e+03  3.600e+03   1.576  0.11514
## pH                    7.505e-01  3.312e+00   0.227  0.82077
## phsq                 -1.861e-01  4.948e-01  -0.376  0.70684
## sulphates             3.705e+00  4.096e-01   9.045  < 2e-16 ***
## sulsq                -1.579e+00  2.293e-01  -6.885 8.34e-12 ***
## alcohol               3.677e-01  2.769e-01   1.328  0.18437
## alcsq                -5.760e-03  1.282e-02  -0.449  0.65333
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6375 on 1571 degrees of freedom
## Multiple R-squared:  0.3867, Adjusted R-squared:  0.3781
## F-statistic: 45.02 on 22 and 1571 DF,  p-value: < 2.2e-16
```

In this case for the original model (model 1), the R^2 and adjusted R^2 values are 0.3867 and 0.3781, respectively. Again, the adjusted R squared value is not very high.

Although both models does not have very high adjusted R squared models, model 2 provides a better model according to this criterion.

7

```
data.for.prediction1 <- data.frame(fixed.acidity = c(6.2, 5.9, 6.3, 5.9, 6), volatile.acidity = c(0.6,
```

```
predict(mlr.model1, data.for.prediction1, interval = "prediction", level = 0.95)
```

```
##        fit      lwr      upr
## 1 5.533859 4.258408 6.809310
## 2 5.966468 4.688943 7.243993
## 3 5.946234 4.670331 7.222137
## 4 5.475676 4.199044 6.752308
## 5 6.009979 4.732074 7.287884
```

Using predictor values, we have these predicted scores of wines

1. 5.533859
2. 5.966468
3. 5.946234
4. 5.475676
5. 6.009979

8

```
mlr.model1mod <- lm(quality~volatile.acidity + chlorides + total.sulfur.dioxide + sulphates + alcohol, d
summary(mlr.model1mod)
```

```
##
## Call:
## lm(formula = quality ~ volatile.acidity + chlorides + total.sulfur.dioxide +
##     sulphates + alcohol, data = winequality)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.67484 -0.38310 -0.06225  0.45350  2.07268
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)           3.0052728  0.2040557  14.728  < 2e-16 ***
## volatile.acidity     -1.1403304  0.0971251 -11.741  < 2e-16 ***
## chlorides            -1.7069660  0.3922863  -4.351 1.44e-05 ***
## total.sulfur.dioxide -0.0023119  0.0005088  -4.544 5.95e-06 ***
## sulphates             0.9150247  0.1104649   8.283 2.51e-16 ***
## alcohol               0.2770586  0.0165063  16.785  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6521 on 1588 degrees of freedom
## Multiple R-squared:  0.3513, Adjusted R-squared:  0.3492
## F-statistic:   172 on 5 and 1588 DF,  p-value: < 2.2e-16
```

The reduced model that only contains significant predictors is shown above