

analysis.rmd

2025-12-02

```
Attaching package: 'dplyr'
```

```
The following objects are masked from 'package:stats':
```

```
  filter, lag
```

```
The following objects are masked from 'package:base':
```

```
  intersect, setdiff, setequal, union
```

```
Loading required package: carData
```

```
Attaching package: 'car'
```

```
The following object is masked from 'package:dplyr':
```

```
  recode
```

```
Loading required package: lattice
```

```
Attaching package: 'olsrr'
```

```
The following object is masked from 'package:datasets':
```

```
  rivers
```

```
df <- read.csv("main.csv")
df <- na.omit(df)
head(df)
```

```
##           Player Team G  MP FG_pct   FTA   TRB   AST stocks  PTS
## 1 Shai Gilgeous-Alexander  OKC 76 34.2  0.519  8.8  5.0  6.4    2.7 32.7
## 2 Giannis Antetokounmpo  MIL 67 34.2  0.601 10.6 11.9  6.5    2.1 30.4
## 3 Nikola Jokić          DEN 70 36.7  0.576  6.4 12.7 10.2    2.4 29.6
## 5 Anthony Edwards        MIN 79 36.3  0.447  6.3  5.7  4.5    1.8 27.6
## 6 Jayson Tatum           BOS 72 36.4  0.452  6.1  8.7  6.0    1.6 26.8
## 7 Kevin Durant          PHO 62 36.5  0.527  5.8  6.0  4.2    2.0 26.6
##   Value_Billions awards_1 awards_2plus avg_salary_millions Age_22_26 Age_27_31
## 1         4.35          0             1            55.3591      1       0
## 2         4.30          0             1            58.4566      0       1
## 3         4.60          0             1            59.0331      0       1
## 5         3.60          0             1            50.6117      1       0
## 6         6.70          0             1            62.7867      1       0
## 7         5.43          1             0            54.7086      0       0
##   Age_32_34 Age_35_plus Pos_PF Pos_PG Pos_SF Pos_SG   Age     Awards
## 1          0          0      0      1      0      0 Age_1 2+ awards
## 2          0          0      1      0      0      0 Age_2 2+ awards
```

```

## 3      0      0      0      0      0 Age_2 2+ awards
## 5      0      0      0      0      1 Age_1 2+ awards
## 6      0      0      1      0      0 Age_1 2+ awards
## 7      0      1      1      0      0 Age_4 1 award

```

Initial Model Creation

```

initial_model <- lm(avg_salary_millions ~ MP + PTS + FG_pct + FTA + TRB + AST + stocks + Value_Billions
summary(initial_model)

##
## Call:
## lm(formula = avg_salary_millions ~ MP + PTS + FG_pct + FTA +
##     TRB + AST + stocks + Value_Billions + (PTS * FTA), data = df)
##
## Residuals:
##    Min      1Q  Median      3Q     Max
## -30.1308 -5.2910  0.0091  4.4900 26.9012
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.32238   4.11447 -0.564  0.57290
## MP          -0.19778   0.16255 -1.217  0.22473
## PTS          1.44802   0.28750  5.037 8.45e-07 ***
## FG_pct      -10.02353  7.82553 -1.281  0.20129
## FTA          -2.07371  1.13461 -1.828  0.06865 .
## TRB          0.79491   0.30568  2.600  0.00980 **
## AST          1.06854   0.40188  2.659  0.00829 **
## stocks       3.12002   1.13606  2.746  0.00641 **
## Value_Billions 0.25564   0.25489  1.003  0.31675
## PTS:FTA     0.11387   0.04448  2.560  0.01099 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.269 on 283 degrees of freedom
## Multiple R-squared:  0.7495, Adjusted R-squared:  0.7415
## F-statistic: 94.07 on 9 and 283 DF,  p-value: < 2.2e-16

```

VIF Analysis

```

vif_values <- vif(initial_model)

## there are higher-order terms (interactions) in this model
## consider setting type = 'predictor'; see ?vif
print(vif_values)

##           MP            PTS          FG_pct          FTA            TRB
## 8.239128 16.765183 1.521308 19.180468 2.486230
##           AST          stocks Value_Billions PTS:FTA
## 2.726504 2.127730 1.021384 21.624494

```

```

# high vifs detected, remove
vif_less <- lm(avg_salary_millions ~ MP + PTS + FG_pct + FTA + TRB + AST + stocks + Value_Billions, data)
print(vif(vif_less))

##          MP            PTS           FG_pct          FTA            TRB
## 7.020613 13.503046    1.500397  6.538584 2.477370
##          AST           stocks Value_Billions
## 2.670589   2.127717    1.013432

```

Stepwise

```
ols_step_both_p(vif_less,p_ent=0.15,p_rem=0.15,details=T)
```

```

## Stepwise Selection Method
## -----
## 
## Candidate Terms:
## 
## 1. MP
## 2. PTS
## 3. FG_pct
## 4. FTA
## 5. TRB
## 6. AST
## 7. stocks
## 8. Value_Billions
## 
## 
## Step    => 0
## Model  => avg_salary_millions ~ 1
## R2     => 0
## 
## Initiating stepwise selection...
## 
## Step      => 1
## Selected  => PTS
## Model     => avg_salary_millions ~ PTS
## R2        => 0.715
## 
## Step      => 2
## Selected  => stocks
## Model     => avg_salary_millions ~ PTS + stocks
## R2        => 0.724
## 
## Step      => 3
## Selected  => AST
## Model     => avg_salary_millions ~ PTS + stocks + AST
## R2        => 0.732
## 
## Step      => 4
## Selected  => MP
## Model     => avg_salary_millions ~ PTS + stocks + AST + MP
## R2        => 0.736

```

```

## 
## Step      => 5
## Selected => TRB
## Model     => avg_salary_millions ~ PTS + stocks + AST + MP + TRB
## R2        => 0.74
##
## Step      => 6
## Selected => FG_pct
## Model     => avg_salary_millions ~ PTS + stocks + AST + MP + TRB + FG_pct
## R2        => 0.742
##
## 
## No more variables to be added or removed.

##
##
## Stepwise Summary
## -----
## Step   Variable       AIC      SBC      SBIC      R2      Adj. R2
## ----- 
## 0      Base Model    2468.817  2476.178  1634.766  0.00000  0.00000
## 1      PTS (+)       2102.890  2113.930  1271.095  0.71513  0.71415
## 2      stocks (+)   2095.460  2110.181  1263.711  0.72415  0.72225
## 3      AST (+)       2089.525  2107.925  1257.906  0.73152  0.72874
## 4      MP (+)        2086.670  2108.751  1255.187  0.73593  0.73227
## 5      TRB (+)       2083.963  2109.724  1252.676  0.74014  0.73562
## 6      FG_pct (+)   2083.664  2113.105  1252.524  0.74217  0.73677
## ----- 

## 
## Final Model Output
## -----
## 
## Model Summary
## -----
## R           0.861      RMSE          8.244
## R-Squared   0.742      MSE           67.965
## Adj. R-Squared 0.737      Coef. Var   50.835
## Pred R-Squared 0.728      AIC          2083.664
## MAE         6.429      SBC          2113.105
## ----- 

## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
## AIC: Akaike Information Criteria
## SBC: Schwarz Bayesian Criteria
## 
## ANOVA
## -----
##           Sum of
##           Squares      DF      Mean Square      F      Sig.
## ----- 
## Regression  57323.780      6      9553.963   137.213  0.00000
## Residual    19913.802     286      69.629
## Total       77237.582     292

```

```

## -----
##                               Parameter Estimates
## -----
##      model      Beta   Std. Error   Std. Beta     t    Sig   lower   upper
## 
## (Intercept) -1.223       3.979          -0.307  0.759 -9.055  6.609
##      PTS      1.844       0.164        0.781 11.221 0.000  1.520  2.167
##      stocks   3.045       1.137        0.116  2.679 0.008  0.808  5.283
##      AST      1.254       0.396        0.153  3.163 0.002  0.474  2.034
##      MP      -0.375       0.135       -0.197 -2.768 0.006 -0.641 -0.108
##      TRB      0.779       0.300        0.120  2.601 0.010  0.190  1.369
##      FG_pct   -11.749      7.827       -0.055 -1.501 0.134 -27.155 3.658
## -----

```

T-Test Quant Only

```

summary(vif_less)

##
## Call:
## lm(formula = avg_salary_millions ~ MP + PTS + FG_pct + FTA +
##     TRB + AST + stocks + Value_Billions, data = df)
##
## Residuals:
##      Min      1Q      Median      3Q      Max
## -30.0240 -5.5008 -0.2633  4.8354 26.2877
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.6726    4.1522 -0.644  0.52032
## MP         -0.3578    0.1515 -2.362  0.01887 *
## PTS         1.7727    0.2605  6.804 6.02e-11 ***
## FG_pct     -12.3722   7.8472 -1.577  0.11599
## FTA         0.2843    0.6689  0.425  0.67111
## TRB         0.7482    0.3081  2.428  0.01579 *
## AST         1.2159    0.4016  3.028  0.00269 **
## stocks      3.1129    1.1471  2.714  0.00706 **
## Value_Billions 0.3132    0.2564  1.222  0.22282
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.349 on 284 degrees of freedom
## Multiple R-squared:  0.7437, Adjusted R-squared:  0.7364
## F-statistic: 103 on 8 and 284 DF, p-value: < 2.2e-16

```

New Model as a Result of Test

```

quant <- lm(avg_salary_millions ~ MP + PTS + TRB + AST, data = df)
summary(quant)

##

```

```

## Call:
## lm(formula = avg_salary_millions ~ MP + PTS + TRB + AST, data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -31.2415  -5.1619   0.0729  4.5430  28.4028
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -6.3780    1.6228  -3.930 0.000106 ***
## MP          -0.2091    0.1235  -1.693 0.091484 .
## PTS          1.7512    0.1627  10.761 < 2e-16 ***
## TRB          0.8646    0.2454   3.524 0.000495 ***
## AST          1.3825    0.3963   3.488 0.000562 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.434 on 288 degrees of freedom
## Multiple R-squared:  0.7347, Adjusted R-squared:  0.7311
## F-statistic: 199.4 on 4 and 288 DF,  p-value: < 2.2e-16

```

Adding Qualitative Predictors

```

quant_and_qual <- lm(avg_salary_millions ~ MP + PTS + TRB + AST + awards_1 + awards_2plus + Age_22_26 +
summary(quant_and_qual)

##
## Call:
## lm(formula = avg_salary_millions ~ MP + PTS + TRB + AST + awards_1 +
##     awards_2plus + Age_22_26 + Age_27_31 + Age_32_34 + Age_35_plus +
##     Pos_PF + Pos_PG + Pos_SF + Pos_SG, data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -25.5348  -4.1983  -0.6067   5.0558  24.9099
##
## Coefficients: (1 not defined because of singularities)
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -4.114614  2.672790  -1.539  0.12483
## MP          0.009948  0.131089   0.076  0.93956
## PTS          1.453234  0.165701   8.770 < 2e-16 ***
## TRB         -0.084596  0.346047  -0.244  0.80705
## AST          1.544745  0.459850   3.359  0.00089 ***
## awards_1     4.356863  2.255561   1.932  0.05442 .
## awards_2plus 11.834128  2.472409   4.786 2.76e-06 ***
## Age_22_26    0.099894  2.080365   0.048  0.96174
## Age_27_31    4.190384  2.162765   1.938  0.05369 .
## Age_32_34    2.994712  2.505735   1.195  0.23305
## Age_35_plus    NA        NA        NA        NA
## Pos_PF       0.276624  1.592812   0.174  0.86225
## Pos_PG      -4.665921  2.340322  -1.994  0.04716 *
## Pos_SF      -3.381229  1.836950  -1.841  0.06673 .
## Pos_SG      -5.236117  1.980071  -2.644  0.00865 **

```

```

## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.832 on 279 degrees of freedom
## Multiple R-squared: 0.7784, Adjusted R-squared: 0.7681
## F-statistic: 75.39 on 13 and 279 DF, p-value: < 2.2e-16

```

ANOVA Test for reduced model (Awards)

```

reduced_awards <- lm(avg_salary_millions ~ MP + PTS + TRB + AST + Age_22_26 + Age_27_31 + Age_32_34 + Age_35_plus + Pos_PF + Pos_PG + Pos_SF + Pos_SG)
anova(reduced_awards, quant_and_qual)

```

```

## Analysis of Variance Table
##
## Model 1: avg_salary_millions ~ MP + PTS + TRB + AST + Age_22_26 + Age_27_31 + Age_32_34 + Age_35_plus + Pos_PF + Pos_PG + Pos_SF + Pos_SG
## Model 2: avg_salary_millions ~ MP + PTS + TRB + AST + awards_1 + awards_2plus + Age_22_26 + Age_27_31 + Age_32_34 + Age_35_plus + Pos_PF + Pos_PG + Pos_SF + Pos_SG
##   Res.Df   RSS Df Sum of Sq      F    Pr(>F)
## 1     281 18582
## 2     279 17116  2     1466.4 11.952 1.047e-05 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

ANOVA Test for reduced model (Age)

```

reduced_age <- lm(avg_salary_millions ~ MP + PTS + TRB + AST + awards_1 + awards_2plus + Pos_PF + Pos_PG + Pos_SF + Pos_SG)
anova(reduced_age, quant_and_qual)

```

```

## Analysis of Variance Table
##
## Model 1: avg_salary_millions ~ MP + PTS + TRB + AST + awards_1 + awards_2plus + Pos_PF + Pos_PG + Pos_SF + Pos_SG
## Model 2: avg_salary_millions ~ MP + PTS + TRB + AST + awards_1 + awards_2plus + Age_22_26 + Age_27_31 + Age_32_34 + Age_35_plus + Pos_PF + Pos_PG + Pos_SF + Pos_SG
##   Res.Df   RSS Df Sum of Sq      F    Pr(>F)
## 1     282 18105
## 2     279 17116  3     988.97 5.3736 0.001306 **
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

ANOVA Test for reduced model (Pos)

```

reduced_pos <- lm(avg_salary_millions ~ MP + PTS + TRB + AST + awards_1 + awards_2plus + Age_22_26 + Age_27_31 + Age_32_34 + Age_35_plus + Pos_PF + Pos_PG + Pos_SF + Pos_SG)
anova(reduced_pos, quant_and_qual)

```

```

## Analysis of Variance Table
##
## Model 1: avg_salary_millions ~ MP + PTS + TRB + AST + awards_1 + awards_2plus +
## Model 2: avg_salary_millions ~ MP + PTS + TRB + AST + awards_1 + awards_2plus + Age_22_26 + Age_27_31 + Age_32_34 + Age_35_plus + Pos_PF + Pos_PG + Pos_SF + Pos_SG
##   Res.Df   RSS Df Sum of Sq      F    Pr(>F)
## 1     282 18105
## 2     279 17116  3     988.97 5.3736 0.001306 **
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

##      Age_22_26 + Age_27_31 + Age_32_34 + Age_35_plus
## Model 2: avg_salary_millions ~ MP + PTS + TRB + AST + awards_1 + awards_2plus +
##      Age_22_26 + Age_27_31 + Age_32_34 + Age_35_plus + Pos_PF +
##      Pos_PG + Pos_SF + Pos_SG
##   Res.Df   RSS Df Sum of Sq    F   Pr(>F)
## 1     283 17960
## 2     279 17116  4     844.26 3.4405 0.009149 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Confidence Intervals

```

conf <- confint(quant_and_qual, level = 0.95)
conf

```

	2.5 %	97.5 %
## (Intercept)	-9.37600976	1.14678214
## MP	-0.24810170	0.26799800
## PTS	1.12705068	1.77941713
## TRB	-0.76579043	0.59659941
## AST	0.63952979	2.44996070
## awards_1	-0.08321464	8.79694141
## awards_2plus	6.96718419	16.70107227
## Age_22_26	-3.99531150	4.19509932
## Age_27_31	-0.06702517	8.44779358
## Age_32_34	-1.93783492	7.92725821
## Age_35_plus	NA	NA
## Pos_PF	-2.85883188	3.41207906
## Pos_PG	-9.27285233	-0.05899061
## Pos_SF	-6.99727148	0.23481320
## Pos_SG	-9.13389344	-1.33833995

K-Fold Cross Validation

```

cv_model <- train(
  avg_salary_millions ~ MP + PTS + TRB + AST + awards_1 + awards_2plus + Age_22_26 + Age_27_31 + Age_32_34 + Age_35_plus,
  method = "lm",
  trControl = trainControl(method = "cv", number = 5)
)

# display results
print(cv_model)

## Linear Regression
##
## 293 samples
## 14 predictor
##
## No pre-processing
## Resampling: Cross-Validated (5 fold)
## Summary of sample sizes: 235, 236, 234, 233, 234
## Resampling results:
##

```

```

##      RMSE      Rsquared     MAE
## 7.978641  0.7593793 6.164269
##
## Tuning parameter 'intercept' was held constant at a value of TRUE

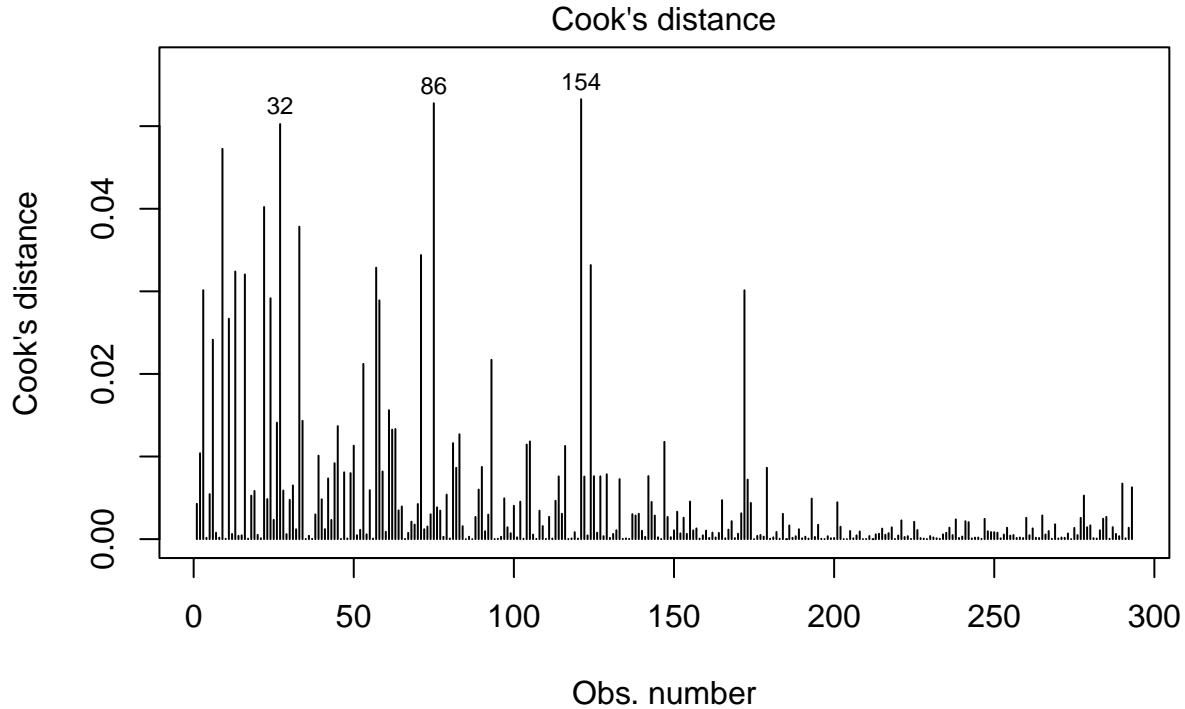
```

Residual Analysis

```

# Cooks Distance Thresholds
plot(quant_and_qual,which=4)

```

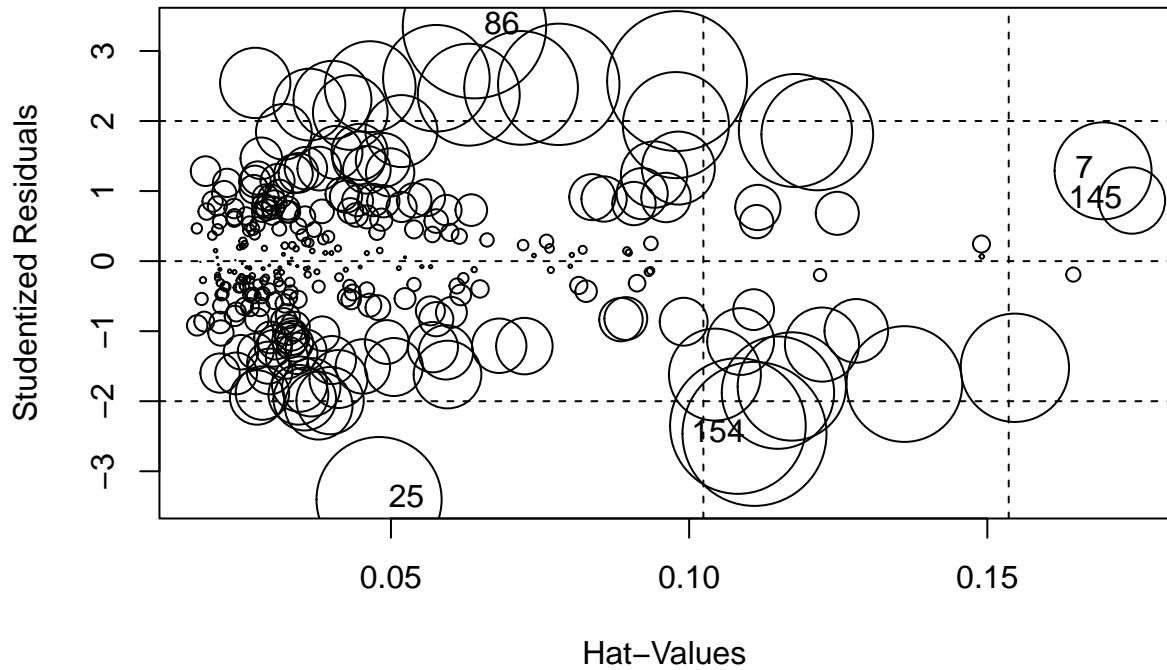


lm(avg_salary_millions ~ MP + PTS + TRB + AST + awards_1 + awards_2plus + A .

```

# Leverage vs Studentized Residuals
influencePlot(quant_and_qual,fill=F)

```



```

##          StudRes      Hat      CookD
## 7     1.2889804 0.16940188 0.02414702
## 25    -3.4041183 0.04800574 0.04021274
## 86     3.3468310 0.06400243 0.05277950
## 145    0.8648404 0.17424756 0.01128375
## 154   -2.4664804 0.11095420 0.05326048

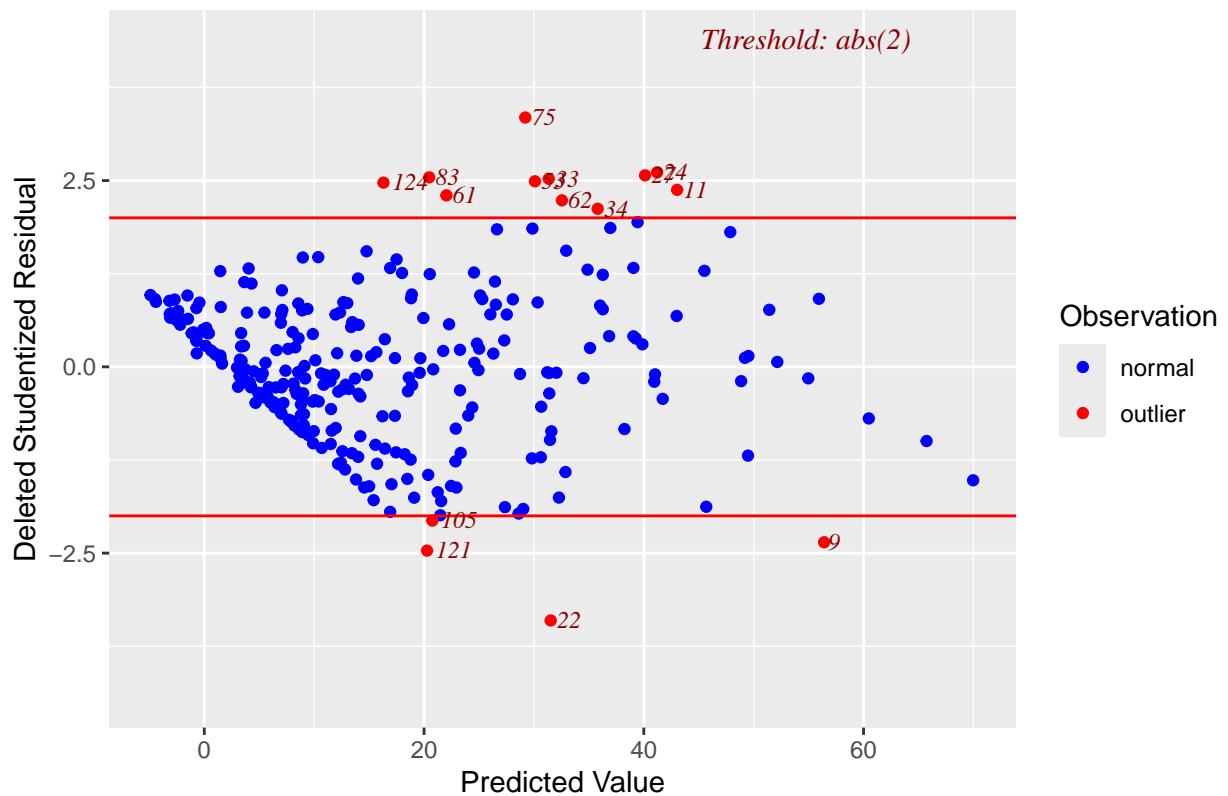
```

```

# Deleted Studentized Residuals vs Predicted values
ols_plot_resid_stud_fit(quant_and_qual)

```

Deleted Studentized Residual vs Predicted Values



remove obs 86 and 154