

# 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

Attaching package: 'olsrr'

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

rivers

```
df <- read.csv("main.csv")
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
## 4 Luka Dončić 2TM 50 35.4 0.450 7.9 8.2 7.7 2.2 28.2
## 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
## 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
## 4 NA 0 0 51.8379 1 0
## 5 3.60 0 1 50.6117 1 0
## 6 6.70 0 1 62.7867 1 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 0 Age_2 2+ awards
## 4 0 0 0 1 0 0 Age_1 0 awards
## 5 0 0 0 0 0 1 Age_1 2+ awards
```

```
## 6          0          0          1          0          0          0 Age_1 2+ awards
```

## 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
## -31.3802  -5.0427   0.0288   4.2201  29.8538
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -1.86018    3.52811  -0.527  0.598364
## MP           -0.15249    0.14624  -1.043  0.297808
## PTS           1.25182    0.26460   4.731 3.27e-06 ***
## FG_pct       -6.44657    6.71992  -0.959  0.338074
## FTA          -2.55141    1.04216  -2.448  0.014859 *
## TRB           0.76480    0.29133   2.625  0.009049 **
## AST           1.05710    0.37944   2.786  0.005634 **
## stocks        1.54249    0.96492   1.599  0.110839
## Value_Billions 0.28366    0.23960   1.184  0.237292
## PTS:FTA        0.15463    0.04218   3.666  0.000286 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.28 on 342 degrees of freedom
## (218 observations deleted due to missingness)
## Multiple R-squared:  0.7187, Adjusted R-squared:  0.7113
## F-statistic: 97.09 on 9 and 342 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
##      8.192566     15.996139      1.471493      17.510374      2.692199
##              AST      stocks Value_Billions      PTS:FTA
##      2.725411      2.119833      1.019061      20.482643
# high vifs detected, remove
vif_less <- lm(avg_salary_millions ~ MP + PTS + FG_pct + FTA + TRB + AST + stocks + Value_Billions, data = df)
print(vif(vif_less))
```

##	MP	PTS	FG_pct	FTA	TRB
##	6.595433	12.115290	1.444336	6.135719	2.684134
##	AST	stocks	Value_Billions		
##	2.671925	2.116299	1.008714		

## 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      => 0.682
##
## Step    => 2
## Selected => AST
## Model   => avg_salary_millions ~ PTS + AST
## R2      => 0.689
##
## Step    => 3
## Selected => TRB
## Model   => avg_salary_millions ~ PTS + AST + TRB
## R2      => 0.694
##
## Step    => 4
## Selected => MP
## Model   => avg_salary_millions ~ PTS + AST + TRB + MP
## R2      => 0.702
##
## Step    => 5
## Selected => stocks
## Model   => avg_salary_millions ~ PTS + AST + TRB + MP + stocks
```

```

## R2          => 0.703
##
## Step        => 6
## Selected    => Value_Billions
## Model       => avg_salary_millions ~ PTS + AST + TRB + MP + stocks + Value_Billions
## R2          => 0.705
##
##
## No more variables to be added or removed.
##
##
##                               Stepwise Summary
## -----
## Step   Variable                AIC          SBC          SBIC          R2          Adj. R2
## -----
## 0      Base Model                2927.395    2935.123    1926.083    0.00000    0.00000
## 1      PTS (+)                   2526.552    2538.143    1527.381    0.68160    0.68069
## 2      AST (+)                   2520.300    2535.755    1521.153    0.68898    0.68719
## 3      TRB (+)                   2516.367    2535.685    1517.282    0.69417    0.69154
## 4      MP (+)                    2509.800    2532.982    1510.919    0.70153    0.69809
## 5      stocks (+)                2509.509    2536.554    1510.717    0.70346    0.69918
## 6      Value_Billions (+)        2509.349    2540.258    1510.667    0.70528    0.70015
## -----
##
## Final Model Output
## -----
##
##                               Model Summary
## -----
## R                0.840          RMSE                8.354
## R-Squared        0.705          MSE                69.791
## Adj. R-Squared   0.700          Coef. Var         57.484
## Pred R-Squared   0.688          AIC                2509.349
## MAE              6.410          SBC                2540.258
## -----
## 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    58787.907        6      9797.984    137.599    0.0000
## Residual      24566.324       345        71.207
## Total         83354.230       351
## -----
##
##                               Parameter Estimates
## -----

```

	model	Beta	Std. Error	Std. Beta	t	Sig	lower	upper
##	(Intercept)	-6.546	1.927		-3.398	0.001	-10.336	-2.757
##	PTS	1.853	0.153	0.803	12.144	0.000	1.553	2.153
##	AST	1.362	0.375	0.170	3.637	0.000	0.626	2.099
##	TRB	0.585	0.258	0.094	2.266	0.024	0.077	1.092
##	MP	-0.392	0.118	-0.220	-3.311	0.001	-0.625	-0.159
##	stocks	1.551	0.975	0.067	1.592	0.112	-0.365	3.468
##	Value_Billions	0.353	0.243	0.043	1.457	0.146	-0.124	0.831

## 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
## -31.5085  -5.4221  -0.0979   4.6283  28.9494
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -2.2846     3.5896  -0.636  0.52491
## MP            -0.3892     0.1336  -2.914  0.00381 **
## PTS            1.7296     0.2344   7.378 1.22e-12 ***
## FG_pct        -9.7931     6.7773  -1.445  0.14937
## FTA            0.5277     0.6280   0.840  0.40132
## TRB            0.7063     0.2961   2.385  0.01761 *
## AST            1.2520     0.3824   3.274  0.00117 **
## stocks         1.6869     0.9814   1.719  0.08655 .
## Value_Billions 0.3722     0.2427   1.534  0.12604
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.429 on 343 degrees of freedom
## (218 observations deleted due to missingness)
## Multiple R-squared:  0.7076, Adjusted R-squared:  0.7008
## F-statistic: 103.8 on 8 and 343 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
## -31.1269  -4.9182  -0.0684   4.5662  29.5636
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -4.8411     1.3315  -3.636 0.000314 ***
## MP           -0.3215     0.1051  -3.059 0.002373 **
## PTS           1.8412     0.1419  12.971 < 2e-16 ***
## TRB           0.7630     0.2182   3.497 0.000524 ***
## AST           1.3912     0.3527   3.944 9.48e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.362 on 393 degrees of freedom
## (172 observations deleted due to missingness)
## Multiple R-squared:  0.7087, Adjusted R-squared:  0.7058
## F-statistic: 239.1 on 4 and 393 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
## -25.9536  -4.6282  -0.5886   4.5199  26.8865
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -5.16538     1.93231  -2.673  0.00784 **
## MP           -0.18030     0.10929  -1.650  0.09981 .
## PTS           1.57116     0.14250  11.026 < 2e-16 ***
## TRB           0.05635     0.31127   0.181  0.85644
## AST           1.62076     0.41008   3.952 9.22e-05 ***
## awards_1      2.98982     1.97163   1.516  0.13024
## awards_2plus 11.49110     2.35847   4.872 1.62e-06 ***
## Age_22_26     2.17957     1.16495   1.871  0.06211 .
## Age_27_31     6.36656     1.27120   5.008 8.40e-07 ***
## Age_32_34     5.69093     1.74793   3.256  0.00123 **
## Age_35_plus   3.86745     2.15448   1.795  0.07343 .
## Pos_PF        1.18073     1.40622   0.840  0.40163
## Pos_PG       -3.63427     2.04020  -1.781  0.07565 .
## Pos_SF       -0.94499     1.62190  -0.583  0.56048
## Pos_SG       -3.49451     1.71234  -2.041  0.04196 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 7.742 on 383 degrees of freedom
## (172 observations deleted due to missingness)
## Multiple R-squared: 0.7566, Adjusted R-squared: 0.7477
## F-statistic: 85.06 on 14 and 383 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)
## 1 385 24397
## 2 383 22958 2 1439 12.003 8.792e-06 ***
## ---
## 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)
## 1 387 24947
## 2 383 22958 4 1989.3 8.2965 1.997e-06 ***
## ---
## 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)
anova(reduced_pos, quant_and_qual)
```

```
## Analysis of Variance Table
##
## Model 1: avg_salary_millions ~ MP + PTS + TRB + AST + awards_1 + awards_2plus +
## 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      387 23808
## 2      383 22958   4      849.47 3.5428 0.007451 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

## Confidence/Prediction Intervals

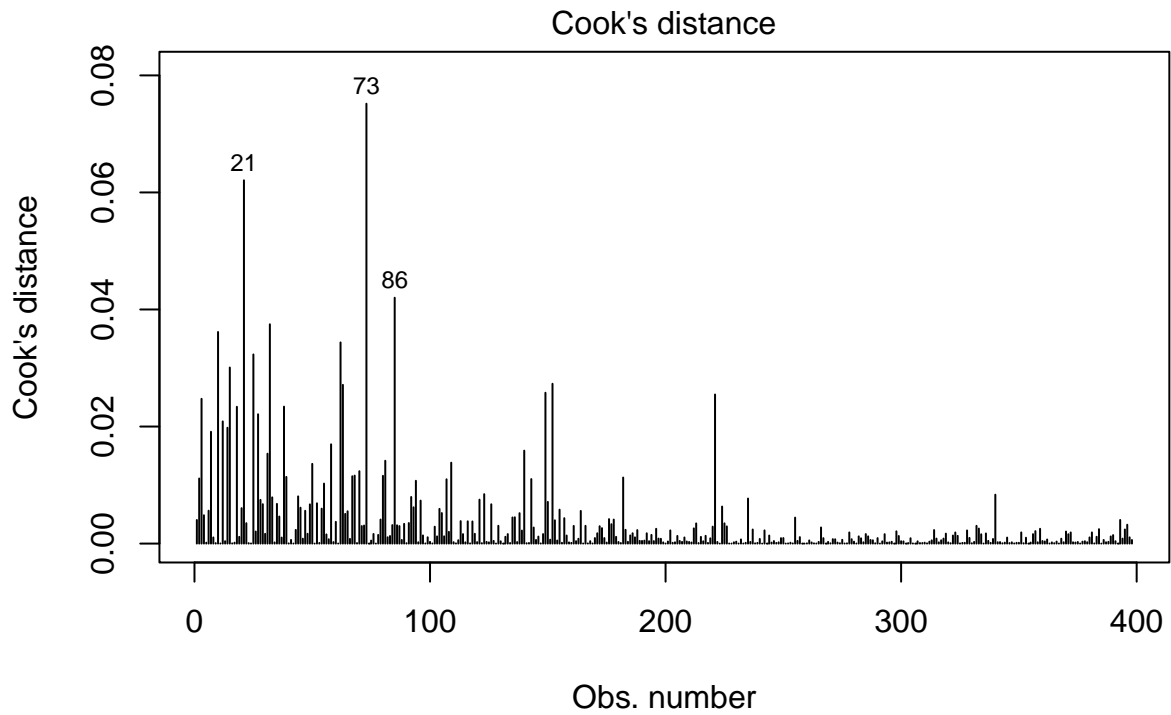
```
conf <- confint(quant_and_qual, level = 0.95)
conf
```

```
##              2.5 %      97.5 %
## (Intercept) -8.9646519 -1.3661091
## MP          -0.3951891  0.0345813
## PTS          1.2909795  1.8513318
## TRB          -0.5556541  0.6683504
## AST          0.8144629  2.4270608
## awards_1     -0.8867470  6.8663866
## awards_2plus  6.8539179 16.1282741
## Age_22_26     -0.1109316  4.4700691
## Age_27_31      3.8671633  8.8659545
## Age_32_34      2.2541889  9.1276762
## Age_35_plus   -0.3686395  8.1035371
## Pos_PF        -1.5841382  3.9456016
## Pos_PG        -7.6456599  0.3771299
## Pos_SF        -4.1339343  2.2439509
## Pos_SG        -6.8612683 -0.1277516
```

```
# add prediction
```

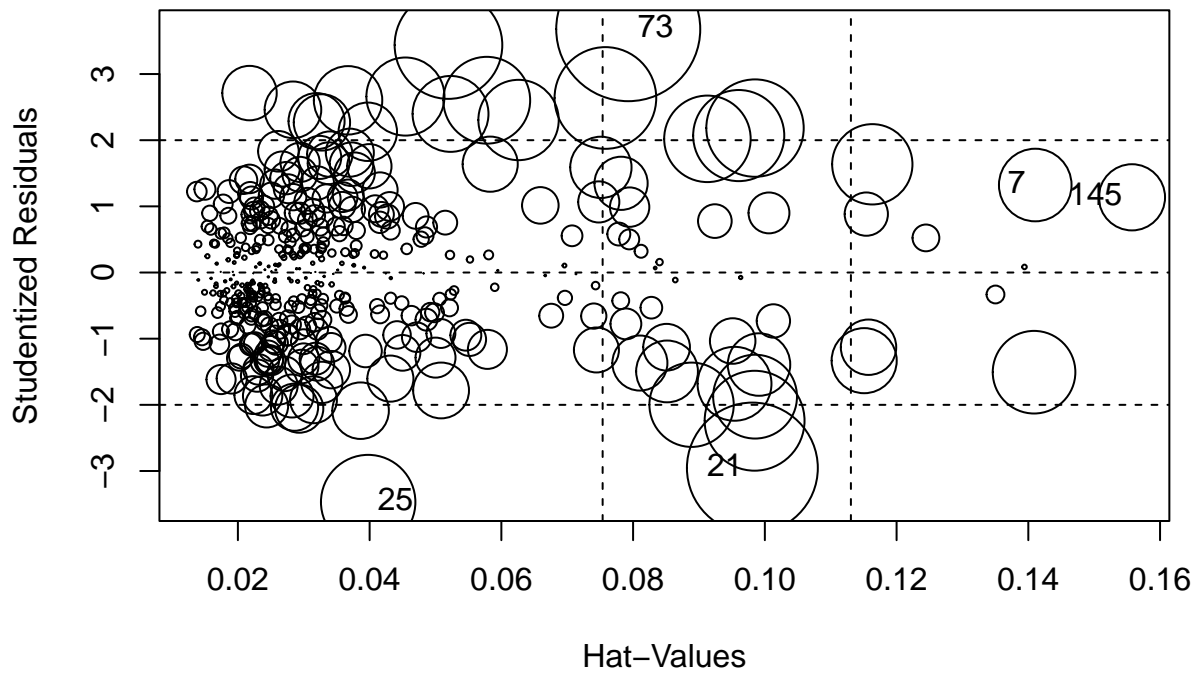
## Residual Analysis

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



$\text{lm}(\text{avg\_salary\_millions} \sim \text{MP} + \text{PTS} + \text{TRB} + \text{AST} + \text{awards\_1} + \text{awards\_2plus} + \text{A} .$

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



##	StudRes	Hat	CookD
## 7	1.322322	0.14105368	0.01910530
## 21	-2.955050	0.09810717	0.06207313
## 25	-3.469876	0.03977828	0.03231987
## 73	3.677788	0.07926145	0.07516773
## 145	1.136801	0.15573349	0.01587993

```
# Deleted Studentized Residuals vs Predicted values
ols_plot_resid_stud_fit(quant_and_qual)
```

