

Predicting College Basketball Win %

2023-01-15

```
#Imported Andrew Sundberg's College Basketball Dataset
cbb_raw = read_csv("/Users/austincicale/Downloads/archive 2/cbb.csv",
show_col_types = FALSE)

#Added a win percentage variable to the data set. Calculated by
dividing games played by games won.
cbb_raw$win_pct = cbb_raw$W / cbb_raw$G

#Removed unessential or redundant variables (TEAM G, W, BARTHAG, WAB,
POSTSEASON, SEED, YEAR)
cbb = select(cbb_raw, -c(1,3,4,7,21:24))

#Randomized the rows of the data set
set.seed(12345)
rows = sample(nrow(cbb))
cbb_shuffled = cbb[rows,]

#Split the data into training and test sets for cross validation.
cbb_train = cbb_shuffled[1:2000,]
cbb_test = cbb_shuffled[2001:2455,]

#Created model to predict win percentage using the training data and a
backward selection method.
full = lm(win_pct~., data=cbb_train)
mse = (summary(full)$sigma)^2
cbb_mod1 = step(full, scale = mse, trace = FALSE)
summary(cbb_mod1)

##
## Call:
## lm(formula = win_pct ~ CONF + ADJOE + ADJDE + EFG_O + EFG_D +
##      TOR + TORD + ORB + DRB + FTR + FTRD + `3P_D` + ADJ_T, data =
cbb_train)
##
## Residuals:
```

```
##           Min           1Q         Median           3Q           Max
## -0.227550 -0.040874 -0.000386  0.043745  0.188481
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.2730264  0.0748075   3.650 0.000269 ***
## CONFACC      -0.0567749  0.0106979  -5.307 1.24e-07 ***
## CONFAC      0.0665765  0.0124276   5.357 9.45e-08 ***
## CONFAMer     -0.0091952  0.0113341  -0.811 0.417300
## CONFASun      0.0877423  0.0120057   7.308 3.92e-13 ***
## CONFB10      -0.0528806  0.0105624  -5.006 6.04e-07 ***
## CONFB12      -0.0451137  0.0119939  -3.761 0.000174 ***
## CONFBEB      -0.0379497  0.0114651  -3.310 0.000950 ***
## CONFBESky     0.0791909  0.0116935   6.772 1.67e-11 ***
## CONFBESth     0.0947896  0.0117774   8.048 1.44e-15 ***
## CONFBW        0.0627311  0.0117048   5.359 9.34e-08 ***
## CONFCAA       0.0573240  0.0115061   4.982 6.85e-07 ***
## CONFCEUSA     0.0487792  0.0103530   4.712 2.63e-06 ***
## CONFCEWC      0.0789573  0.0301664   2.617 0.008929 **
## CONFCEHorz    0.0499373  0.0113937   4.383 1.23e-05 ***
## CONFCEInd     0.1122304  0.0378374   2.966 0.003052 **
## CONFCEInd     0.1050650  0.0642886   1.634 0.102363
## CONFCEIvy     0.0284747  0.0121247   2.348 0.018949 *
## CONFCEMAAC     0.0401269  0.0114427   3.507 0.000464 ***
## CONFCEMAC     0.0558274  0.0105382   5.298 1.31e-07 ***
## CONFCEMEAC    0.1164022  0.0121785   9.558 < 2e-16 ***
## CONFCEMVC     0.0265028  0.0111930   2.368 0.017991 *
## CONFCEMWC     0.0149370  0.0110118   1.356 0.175110
## CONFCENEC     0.0946098  0.0121526   7.785 1.12e-14 ***
## CONFCEOV      0.0705950  0.0114044   6.190 7.30e-10 ***
## CONFCEP12     -0.0204097  0.0105423  -1.936 0.053013 .
## CONFCEPat     0.0519664  0.0116070   4.477 8.00e-06 ***
## CONFCEFSB     0.0628703  0.0109675   5.732 1.14e-08 ***
## CONFCEFS      0.0854249  0.0115428   7.401 2.00e-13 ***
## CONFCESEC     -0.0334522  0.0102196  -3.273 0.001081 **
## CONFCESlnd    0.1320885  0.0116686  11.320 < 2e-16 ***
## CONFCESum     0.0704169  0.0122530   5.747 1.05e-08 ***
```

```

## CONFSWAC      0.1321378  0.0134847   9.799 < 2e-16 ***
## CONFWAC       0.0751020  0.0124261   6.044 1.80e-09 ***
## CONFWCC       0.0109336  0.0113047   0.967 0.333577
## ADJOE         0.0075244  0.0008707   8.642 < 2e-16 ***
## ADJDE        -0.0047435  0.0009141  -5.189 2.33e-07 ***
## EFG_O         0.0159435  0.0013589  11.733 < 2e-16 ***
## EFG_D        -0.0177866  0.0015237 -11.673 < 2e-16 ***
## TOR          -0.0149918  0.0013595 -11.027 < 2e-16 ***
## TORD          0.0207311  0.0013098  15.827 < 2e-16 ***
## ORB           0.0052683  0.0006610   7.970 2.68e-15 ***
## DRB          -0.0104934  0.0007521 -13.952 < 2e-16 ***
## FTR           0.0025687  0.0003206   8.013 1.90e-15 ***
## FTRD         -0.0040280  0.0002995 -13.449 < 2e-16 ***
## `3P_D`       -0.0033834  0.0008996  -3.761 0.000174 ***
## ADJ_T         0.0033358  0.0004754   7.017 3.11e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.06354 on 1953 degrees of freedom
## Multiple R-squared:  0.8762, Adjusted R-squared:  0.8733
## F-statistic: 300.5 on 46 and 1953 DF,  p-value: < 2.2e-16

#Created model to predict win percentage using the training data and a
forward selection method.
none = lm(win_pct~1, data=cbb_train)
cbb_mod2 = step(none, scope=list(upper=full), scale = mse,
direction="forward", trace = FALSE)
summary(cbb_mod2)

##
## Call:
## lm(formula = win_pct ~ ADJOE + EFG_D + CONF + ADJDE + EFG_O +
##      TOR + ORB + ADJ_T + FTR + FTRD + TORD + DRB + `3P_D`, data =
cbb_train)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.227550 -0.040874 -0.000386  0.043745  0.188481

```

##

Coefficients:

##	Estimate	Std. Error	t value	Pr(> t)	
## (Intercept)	0.2730264	0.0748075	3.650	0.000269	***
## ADJOE	0.0075244	0.0008707	8.642	< 2e-16	***
## EFG_D	-0.0177866	0.0015237	-11.673	< 2e-16	***
## CONFACC	-0.0567749	0.0106979	-5.307	1.24e-07	***
## CONF AE	0.0665765	0.0124276	5.357	9.45e-08	***
## CONF Amer	-0.0091952	0.0113341	-0.811	0.417300	
## CONF ASun	0.0877423	0.0120057	7.308	3.92e-13	***
## CONF B10	-0.0528806	0.0105624	-5.006	6.04e-07	***
## CONF B12	-0.0451137	0.0119939	-3.761	0.000174	***
## CONF BE	-0.0379497	0.0114651	-3.310	0.000950	***
## CONF B Sky	0.0791909	0.0116935	6.772	1.67e-11	***
## CONF B St h	0.0947896	0.0117774	8.048	1.44e-15	***
## CONF BW	0.0627311	0.0117048	5.359	9.34e-08	***
## CONF CAA	0.0573240	0.0115061	4.982	6.85e-07	***
## CONF C USA	0.0487792	0.0103530	4.712	2.63e-06	***
## CONF GWC	0.0789573	0.0301664	2.617	0.008929	**
## CONF Horz	0.0499373	0.0113937	4.383	1.23e-05	***
## CONF ind	0.1122304	0.0378374	2.966	0.003052	**
## CONF Ind	0.1050650	0.0642886	1.634	0.102363	
## CONF Ivy	0.0284747	0.0121247	2.348	0.018949	*
## CONF MAAC	0.0401269	0.0114427	3.507	0.000464	***
## CONF MAC	0.0558274	0.0105382	5.298	1.31e-07	***
## CONF MEAC	0.1164022	0.0121785	9.558	< 2e-16	***
## CONF MVC	0.0265028	0.0111930	2.368	0.017991	*
## CONF MWC	0.0149370	0.0110118	1.356	0.175110	
## CONF NEC	0.0946098	0.0121526	7.785	1.12e-14	***
## CONF OVC	0.0705950	0.0114044	6.190	7.30e-10	***
## CONF P12	-0.0204097	0.0105423	-1.936	0.053013	.
## CONF Pat	0.0519664	0.0116070	4.477	8.00e-06	***
## CONF SB	0.0628703	0.0109675	5.732	1.14e-08	***
## CONF SC	0.0854249	0.0115428	7.401	2.00e-13	***
## CONF SEC	-0.0334522	0.0102196	-3.273	0.001081	**
## CONF S lnd	0.1320885	0.0116686	11.320	< 2e-16	***
## CONF Sum	0.0704169	0.0122530	5.747	1.05e-08	***

```

## CONFSWAC      0.1321378  0.0134847   9.799 < 2e-16 ***
## CONFWAC       0.0751020  0.0124261   6.044 1.80e-09 ***
## CONFWCC       0.0109336  0.0113047   0.967 0.333577
## ADJDE        -0.0047435  0.0009141  -5.189 2.33e-07 ***
## EFG_O         0.0159435  0.0013589  11.733 < 2e-16 ***
## TOR          -0.0149918  0.0013595 -11.027 < 2e-16 ***
## ORB           0.0052683  0.0006610   7.970 2.68e-15 ***
## ADJ_T         0.0033358  0.0004754   7.017 3.11e-12 ***
## FTR           0.0025687  0.0003206   8.013 1.90e-15 ***
## FTRD         -0.0040280  0.0002995 -13.449 < 2e-16 ***
## TORD          0.0207311  0.0013098  15.827 < 2e-16 ***
## DRB          -0.0104934  0.0007521 -13.952 < 2e-16 ***
## `3P_D`       -0.0033834  0.0008996  -3.761 0.000174 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.06354 on 1953 degrees of freedom
## Multiple R-squared:  0.8762, Adjusted R-squared:  0.8733
## F-statistic: 300.5 on 46 and 1953 DF,  p-value: < 2.2e-16

#Created model to predict win percentage using the training data and a
stepwise selection method.
cbb_mod3 = step(none, scope=list(upper=full), scale = mse, trace =
FALSE)
summary(cbb_mod3)

##
## Call:
## lm(formula = win_pct ~ ADJOE + EFG_D + CONF + ADJDE + EFG_O +
##      TOR + ORB + ADJ_T + FTR + FTRD + TORD + DRB + `3P_D`, data =
cbb_train)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.227550 -0.040874 -0.000386  0.043745  0.188481
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)

```

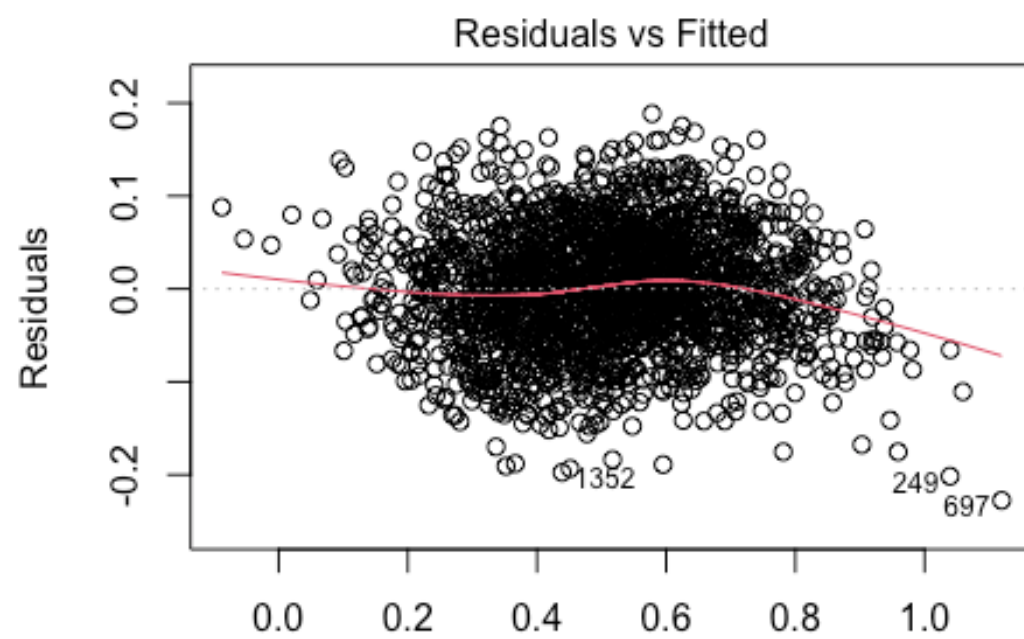
## (Intercept)	0.2730264	0.0748075	3.650	0.000269	***
## ADJOE	0.0075244	0.0008707	8.642	< 2e-16	***
## EFG_D	-0.0177866	0.0015237	-11.673	< 2e-16	***
## CONFACC	-0.0567749	0.0106979	-5.307	1.24e-07	***
## CONFAE	0.0665765	0.0124276	5.357	9.45e-08	***
## CONFAMer	-0.0091952	0.0113341	-0.811	0.417300	
## CONFASun	0.0877423	0.0120057	7.308	3.92e-13	***
## CONFB10	-0.0528806	0.0105624	-5.006	6.04e-07	***
## CONFB12	-0.0451137	0.0119939	-3.761	0.000174	***
## CONFB_E	-0.0379497	0.0114651	-3.310	0.000950	***
## CONFB_Sky	0.0791909	0.0116935	6.772	1.67e-11	***
## CONFB_Sth	0.0947896	0.0117774	8.048	1.44e-15	***
## CONFBW	0.0627311	0.0117048	5.359	9.34e-08	***
## CONFCAA	0.0573240	0.0115061	4.982	6.85e-07	***
## CONF_CUSA	0.0487792	0.0103530	4.712	2.63e-06	***
## CONFGWC	0.0789573	0.0301664	2.617	0.008929	**
## CONFHorz	0.0499373	0.0113937	4.383	1.23e-05	***
## CONF_ind	0.1122304	0.0378374	2.966	0.003052	**
## CONFInd	0.1050650	0.0642886	1.634	0.102363	
## CONF_Ivy	0.0284747	0.0121247	2.348	0.018949	*
## CONFMAAC	0.0401269	0.0114427	3.507	0.000464	***
## CONFMAC	0.0558274	0.0105382	5.298	1.31e-07	***
## CONFMEAC	0.1164022	0.0121785	9.558	< 2e-16	***
## CONF_MVC	0.0265028	0.0111930	2.368	0.017991	*
## CONF_MWC	0.0149370	0.0110118	1.356	0.175110	
## CONFNEC	0.0946098	0.0121526	7.785	1.12e-14	***
## CONF_OVC	0.0705950	0.0114044	6.190	7.30e-10	***
## CONF_P12	-0.0204097	0.0105423	-1.936	0.053013	.
## CONF_Pat	0.0519664	0.0116070	4.477	8.00e-06	***
## CONF_SB	0.0628703	0.0109675	5.732	1.14e-08	***
## CONF_SC	0.0854249	0.0115428	7.401	2.00e-13	***
## CONFSEC	-0.0334522	0.0102196	-3.273	0.001081	**
## CONF_Slnd	0.1320885	0.0116686	11.320	< 2e-16	***
## CONF_Sum	0.0704169	0.0122530	5.747	1.05e-08	***
## CONF_SWAC	0.1321378	0.0134847	9.799	< 2e-16	***
## CONF_WAC	0.0751020	0.0124261	6.044	1.80e-09	***
## CONF_WCC	0.0109336	0.0113047	0.967	0.333577	

```

## ADJDE      -0.0047435  0.0009141  -5.189  2.33e-07 ***
## EFG_O      0.0159435  0.0013589  11.733  < 2e-16 ***
## TOR       -0.0149918  0.0013595 -11.027  < 2e-16 ***
## ORB        0.0052683  0.0006610   7.970  2.68e-15 ***
## ADJ_T      0.0033358  0.0004754   7.017  3.11e-12 ***
## FTR        0.0025687  0.0003206   8.013  1.90e-15 ***
## FTRD      -0.0040280  0.0002995 -13.449  < 2e-16 ***
## TORD       0.0207311  0.0013098  15.827  < 2e-16 ***
## DRB       -0.0104934  0.0007521 -13.952  < 2e-16 ***
## `3P_D`    -0.0033834  0.0008996  -3.761  0.000174 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.06354 on 1953 degrees of freedom
## Multiple R-squared:  0.8762, Adjusted R-squared:  0.8733
## F-statistic: 300.5 on 46 and 1953 DF,  p-value: < 2.2e-16

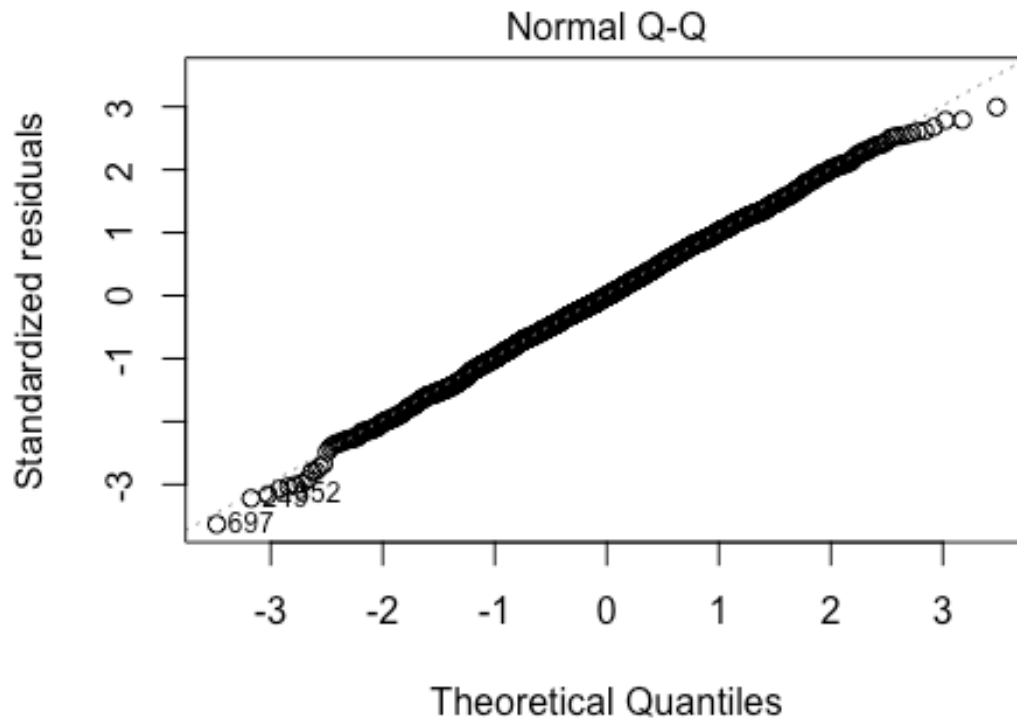
#Produced relevant plots for checking model conditions (linearity,
constant variance, and normality)
plot(cbb_mod1, 1:2)

```



Fitted values

$\sim \text{CONF} + \text{ADJOE} + \text{ADJDE} + \text{EFG_O} + \text{EFG_D} + \text{TOR} + \text{TORD} +$



~ CONF + ADJOE + ADJDE + EFG_O + EFG_D + TOR + TORD +

#Residual analysis using the standardized and studentized residuals of the five largest absolute residuals.

```
indices = sort(abs(cbb_mod1$resid), decreasing = TRUE,
index.return=TRUE)$ix[1:5]
rstandard(cbb_mod1)[indices]
```

```
##          697          249          1352          949          1561
## -3.632347 -3.220039 -3.162222 -3.069980 -3.033019
```

```
rstudent(cbb_mod1)[indices]
```

```
##          697          249          1352          949          1561
## -3.643746 -3.227794 -3.169537 -3.076627 -3.039409
```

#Analyzing leverages for the five largest absolute residuals.

```
hatvalues(cbb_mod1)[indices]
```

```
##           697           249           1352           949           1561
## 0.02794625 0.02889947 0.03420749 0.01770728 0.02046238
```

```
2*2/2000
```

```
## [1] 0.002
```

```
3*2/2000
```

```
## [1] 0.003
```

```
#Calculating Cook's distance for the five largest absolute residuals,  
estimating the influence of these points.
```

```
head(sort(cooks.distance(cbb_mod1)[indices], decreasing=TRUE))
```

```
##           697           1352           249           1561           949
## 0.008070677 0.007535711 0.006565230 0.004088719 0.003614799
```

```
#Experimenting with variable transformations and interactions.
```

```
cbb_test_mod = lm(win_pct ~ ADJOE + EFG_D + CONF + ADJDE + EFG_O + TOR  
+ ORB + ADJ_T + FTR + FTRD + TORD + DRB + `3P_D` + ADJOE*TORD +  
EFG_O*ADJOE + EFG_D*ADJDE + ADJ_T*TORD + FTR*ADJDE,
```

```
data = cbb_train)
```

```
summary(cbb_test_mod)
```

```
##
```

```
## Call:
```

```
## lm(formula = win_pct ~ ADJOE + EFG_D + CONF + ADJDE + EFG_O +  
## TOR + ORB + ADJ_T + FTR + FTRD + TORD + DRB + `3P_D` + ADJOE *  
## TORD + EFG_O * ADJOE + EFG_D * ADJDE + ADJ_T * TORD + FTR *  
## ADJDE, data = cbb_train)
```

```
##
```

```
## Residuals:
```

```
##           Min           1Q       Median           3Q           Max  
## -0.205460 -0.040637 -0.000436  0.043820  0.177025
```

```
##
```

```
## Coefficients:
```

```
##           Estimate Std. Error t value Pr(>|t|)  
## (Intercept) -2.462e+00  5.895e-01  -4.177 3.09e-05 ***  
## ADJOE        2.346e-02  3.645e-03   6.435 1.55e-10 ***  
## EFG_D        6.515e-03  6.643e-03   0.981 0.326805
```

## CONFACC	-5.398e-02	1.062e-02	-5.082	4.09e-07	***
## CONF AE	7.177e-02	1.236e-02	5.809	7.33e-09	***
## CONF Amer	-6.126e-03	1.124e-02	-0.545	0.585667	
## CONF ASun	9.348e-02	1.194e-02	7.832	7.85e-15	***
## CONF B10	-4.877e-02	1.055e-02	-4.623	4.04e-06	***
## CONF B12	-4.207e-02	1.189e-02	-3.539	0.000412	***
## CONF BE	-3.364e-02	1.137e-02	-2.959	0.003126	**
## CONF B Sky	8.843e-02	1.168e-02	7.573	5.61e-14	***
## CONF B Sth	9.780e-02	1.168e-02	8.370	< 2e-16	***
## CONF BW	6.505e-02	1.161e-02	5.602	2.42e-08	***
## CONF CAA	6.102e-02	1.142e-02	5.343	1.02e-07	***
## CONF C USA	5.271e-02	1.027e-02	5.134	3.12e-07	***
## CONF GWC	8.149e-02	3.017e-02	2.701	0.006965	**
## CONF Horz	5.323e-02	1.130e-02	4.713	2.61e-06	***
## CONF ind	1.213e-01	3.760e-02	3.225	0.001280	**
## CONF Ind	1.040e-01	6.363e-02	1.635	0.102194	
## CONF Ivy	3.100e-02	1.201e-02	2.581	0.009925	**
## CONF MAAC	4.391e-02	1.137e-02	3.863	0.000115	***
## CONF MAC	5.892e-02	1.047e-02	5.628	2.08e-08	***
## CONF MEAC	1.276e-01	1.243e-02	10.265	< 2e-16	***
## CONF MVC	2.756e-02	1.108e-02	2.487	0.012965	*
## CONF MWC	1.640e-02	1.091e-02	1.504	0.132824	
## CONF NEC	9.792e-02	1.209e-02	8.100	9.62e-16	***
## CONF OVC	7.752e-02	1.136e-02	6.824	1.18e-11	***
## CONF P12	-2.010e-02	1.045e-02	-1.924	0.054487	.
## CONF Pat	5.296e-02	1.150e-02	4.606	4.38e-06	***
## CONF SB	6.516e-02	1.087e-02	5.992	2.47e-09	***
## CONF SC	9.116e-02	1.148e-02	7.940	3.37e-15	***
## CONF SEC	-3.374e-02	1.013e-02	-3.330	0.000884	***
## CONF Slnd	1.402e-01	1.167e-02	12.010	< 2e-16	***
## CONF Sum	7.329e-02	1.215e-02	6.030	1.95e-09	***
## CONF SWAC	1.437e-01	1.373e-02	10.463	< 2e-16	***
## CONF WAC	8.125e-02	1.238e-02	6.565	6.65e-11	***
## CONF WCC	1.549e-02	1.125e-02	1.377	0.168557	
## ADJDE	1.167e-02	3.884e-03	3.004	0.002695	**
## EFG_O	3.763e-02	5.458e-03	6.894	7.33e-12	***
## TOR	-1.412e-02	1.365e-03	-10.342	< 2e-16	***

```

## ORB          4.899e-03  6.594e-04   7.430 1.62e-13 ***
## ADJ_T        -4.867e-03  3.821e-03  -1.274 0.202948
## FTR          1.609e-02  4.349e-03   3.699 0.000222 ***
## FTRD         -4.135e-03  2.975e-04 -13.900 < 2e-16 ***
## TORD         1.549e-02  1.577e-02   0.982 0.326147
## DRB          -1.042e-02  7.460e-04 -13.973 < 2e-16 ***
## `3P_D`       -3.484e-03  8.916e-04  -3.908 9.63e-05 ***
## ADJOE:TORD   -2.282e-04  9.068e-05  -2.517 0.011910 *
## ADJOE:EFG_O  -2.193e-04  5.294e-05  -4.143 3.57e-05 ***
## EFG_D:ADJDE  -2.343e-04  6.259e-05  -3.744 0.000187 ***
## ADJ_T:TORD   4.279e-04  2.000e-04   2.139 0.032541 *
## ADJDE:FTR    -1.329e-04  4.186e-05  -3.175 0.001519 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.06289 on 1948 degrees of freedom
## Multiple R-squared:  0.879, Adjusted R-squared:  0.8759
## F-statistic: 277.6 on 51 and 1948 DF,  p-value: < 2.2e-16

#Finalizing model using a backward selection method, including
potential variable interactions selected in cbb_test_mod.
full_2 = lm(win_pct ~ ADJOE + EFG_D + CONF + ADJDE + EFG_O +
            TOR + ORB + ADJ_T + FTR + FTRD + TORD + DRB + `3P_D` + ADJOE*TORD
+ EFG_O*ADJOE + EFG_D*ADJDE + ADJ_T*TORD + FTR*ADJDE, data=cbb_train)
mse_2 = (summary(full_2)$sigma)^2
cbb_final_mod = step(full_2, scale = mse_2, trace = FALSE)
summary(cbb_final_mod)

##
## Call:
## lm(formula = win_pct ~ ADJOE + EFG_D + CONF + ADJDE + EFG_O +
##     TOR + ORB + ADJ_T + FTR + FTRD + TORD + DRB + `3P_D` + ADJOE *
##     TORD + EFG_O * ADJOE + EFG_D * ADJDE + ADJ_T * TORD + FTR *
##     ADJDE, data = cbb_train)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.205460 -0.040637 -0.000436  0.043820  0.177025

```

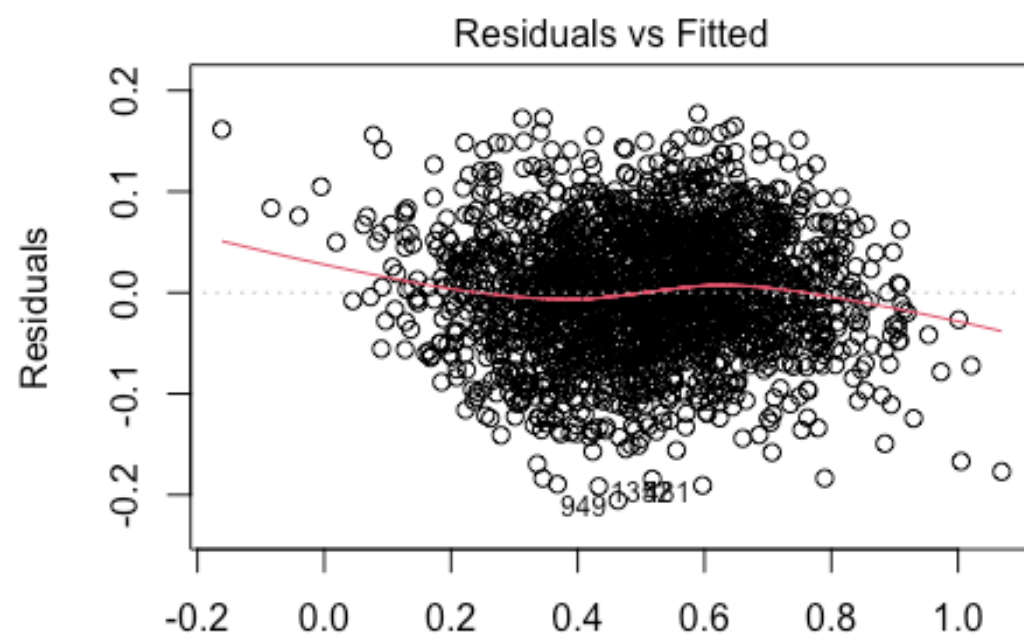
```
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.462e+00  5.895e-01  -4.177 3.09e-05 ***
## ADJOE       2.346e-02  3.645e-03   6.435 1.55e-10 ***
## EFG_D       6.515e-03  6.643e-03   0.981 0.326805
## CONFACC     -5.398e-02  1.062e-02  -5.082 4.09e-07 ***
## CONFAC      7.177e-02  1.236e-02   5.809 7.33e-09 ***
## CONFAMer    -6.126e-03  1.124e-02  -0.545 0.585667
## CONFASun     9.348e-02  1.194e-02   7.832 7.85e-15 ***
## CONFB10     -4.877e-02  1.055e-02  -4.623 4.04e-06 ***
## CONFB12     -4.207e-02  1.189e-02  -3.539 0.000412 ***
## CONFBE      -3.364e-02  1.137e-02  -2.959 0.003126 **
## CONFBSky     8.843e-02  1.168e-02   7.573 5.61e-14 ***
## CONFBSth     9.780e-02  1.168e-02   8.370 < 2e-16 ***
## CONFBW      6.505e-02  1.161e-02   5.602 2.42e-08 ***
## CONFCAA     6.102e-02  1.142e-02   5.343 1.02e-07 ***
## CONFUSA     5.271e-02  1.027e-02   5.134 3.12e-07 ***
## CONFGWC     8.149e-02  3.017e-02   2.701 0.006965 **
## CONFHorz    5.323e-02  1.130e-02   4.713 2.61e-06 ***
## CONFInd     1.213e-01  3.760e-02   3.225 0.001280 **
## CONFInd     1.040e-01  6.363e-02   1.635 0.102194
## CONFIVy     3.100e-02  1.201e-02   2.581 0.009925 **
## CONFMAAC     4.391e-02  1.137e-02   3.863 0.000115 ***
## CONFMAC     5.892e-02  1.047e-02   5.628 2.08e-08 ***
## CONFMEAC    1.276e-01  1.243e-02  10.265 < 2e-16 ***
## CONF MVC    2.756e-02  1.108e-02   2.487 0.012965 *
## CONF MWC    1.640e-02  1.091e-02   1.504 0.132824
## CONFNEC     9.792e-02  1.209e-02   8.100 9.62e-16 ***
## CONF OVC    7.752e-02  1.136e-02   6.824 1.18e-11 ***
## CONF P12    -2.010e-02  1.045e-02  -1.924 0.054487 .
## CONF Pat    5.296e-02  1.150e-02   4.606 4.38e-06 ***
## CONF SB     6.516e-02  1.087e-02   5.992 2.47e-09 ***
## CONF SC     9.116e-02  1.148e-02   7.940 3.37e-15 ***
## CONF SEC   -3.374e-02  1.013e-02  -3.330 0.000884 ***
## CONF SInd   1.402e-01  1.167e-02  12.010 < 2e-16 ***
## CONF Sum    7.329e-02  1.215e-02   6.030 1.95e-09 ***
```

```

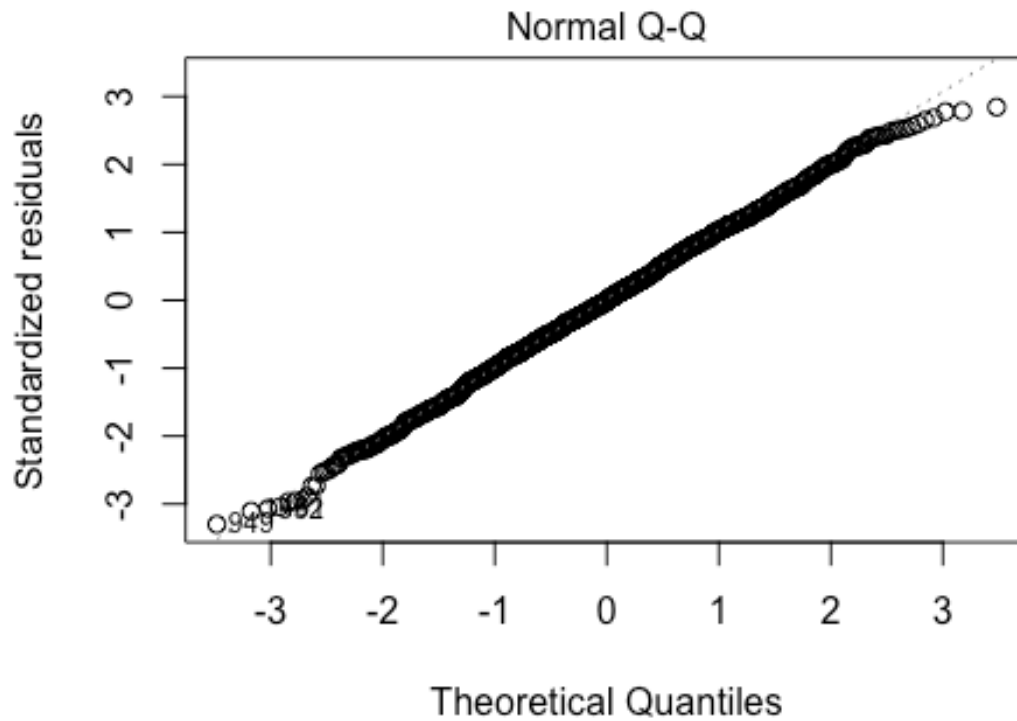
## CONFSWAC      1.437e-01  1.373e-02  10.463  < 2e-16 ***
## CONFWAC       8.125e-02  1.238e-02   6.565  6.65e-11 ***
## CONFWCC       1.549e-02  1.125e-02   1.377  0.168557
## ADJDE         1.167e-02  3.884e-03   3.004  0.002695 **
## EFG_O         3.763e-02  5.458e-03   6.894  7.33e-12 ***
## TOR          -1.412e-02  1.365e-03 -10.342  < 2e-16 ***
## ORB           4.899e-03  6.594e-04   7.430  1.62e-13 ***
## ADJ_T        -4.867e-03  3.821e-03  -1.274  0.202948
## FTR           1.609e-02  4.349e-03   3.699  0.000222 ***
## FTRD         -4.135e-03  2.975e-04 -13.900  < 2e-16 ***
## TORD          1.549e-02  1.577e-02   0.982  0.326147
## DRB          -1.042e-02  7.460e-04 -13.973  < 2e-16 ***
## `3P_D`       -3.484e-03  8.916e-04  -3.908  9.63e-05 ***
## ADJOE:TORD   -2.282e-04  9.068e-05  -2.517  0.011910 *
## ADJOE:EFG_O  -2.193e-04  5.294e-05  -4.143  3.57e-05 ***
## EFG_D:ADJDE  -2.343e-04  6.259e-05  -3.744  0.000187 ***
## ADJ_T:TORD   4.279e-04  2.000e-04   2.139  0.032541 *
## ADJDE:FTR    -1.329e-04  4.186e-05  -3.175  0.001519 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.06289 on 1948 degrees of freedom
## Multiple R-squared:  0.879, Adjusted R-squared:  0.8759
## F-statistic: 277.6 on 51 and 1948 DF,  p-value: < 2.2e-16

#Produced relevant plots for checking conditions (linearity, constant
variance, and normality) of the final model.
plot(cbb_final_mod, 1:2)

```



Fitted values
 $t \sim \text{ADJOE} + \text{EFG_D} + \text{CONF} + \text{ADJDE} + \text{EFG_O} + \text{TOR} + \text{ORB} + /$



$\text{win_pct} \sim \text{ADJOE} + \text{EFG_D} + \text{CONF} + \text{ADJDE} + \text{EFG_O} + \text{TOR} + \text{ORB} + \text{FTR} + \text{FTRD} + \text{TORD} + \text{DRB} + \text{`3P_D`}$

#Reassure model improvement using ANOVA testing.

```
anova(cbb_mod1, cbb_final_mod)
```

```
## Analysis of Variance Table
```

```
##
```

```
## Model 1: win_pct ~ CONF + ADJOE + ADJDE + EFG_O + EFG_D + TOR +  
TORD +
```

```
##      ORB + DRB + FTR + FTRD + `3P_D` + ADJ_T
```

```
## Model 2: win_pct ~ ADJOE + EFG_D + CONF + ADJDE + EFG_O + TOR + ORB  
+
```

```
##      ADJ_T + FTR + FTRD + TORD + DRB + `3P_D` + ADJOE * TORD +
```

```
##      EFG_O * ADJOE + EFG_D * ADJDE + ADJ_T * TORD + FTR * ADJDE
```

```
##   Res.Df    RSS Df Sum of Sq      F     Pr(>F)
```

```
## 1    1953  7.8848
```

```
## 2    1948  7.7036   5      0.1812  9.1639 1.247e-08 ***
```



```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

#Compute residuals for the testing data, using the model created with
the training data.
fit_win_pct = predict(cbb_final_mod, newdata = cbb_test)
cbb_test_resid = cbb_test$win_pct - fit_win_pct

#Compute the mean value of the testing data residuals to assure the
zero mean condition is being upheld.
mean(cbb_test_resid)

## [1] -0.001951512

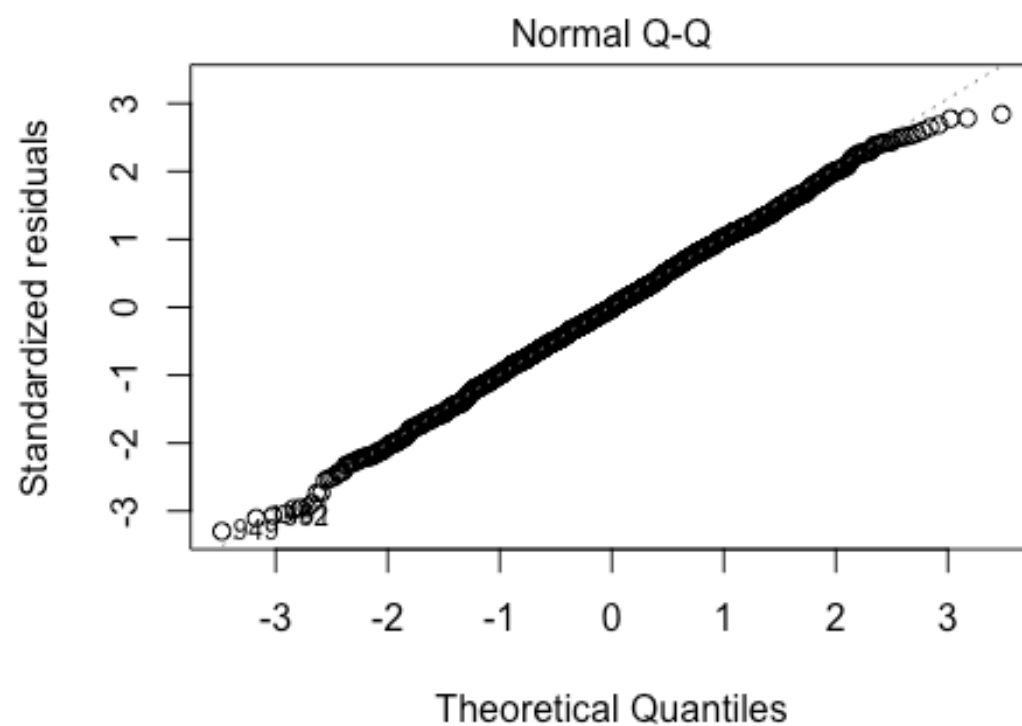
#Compute the standard deviation of the testing and trained data
residuals to check variability condition.
summary(cbb_final_mod)$sigma

## [1] 0.06288574

sd(cbb_test_resid)

## [1] 0.05838225

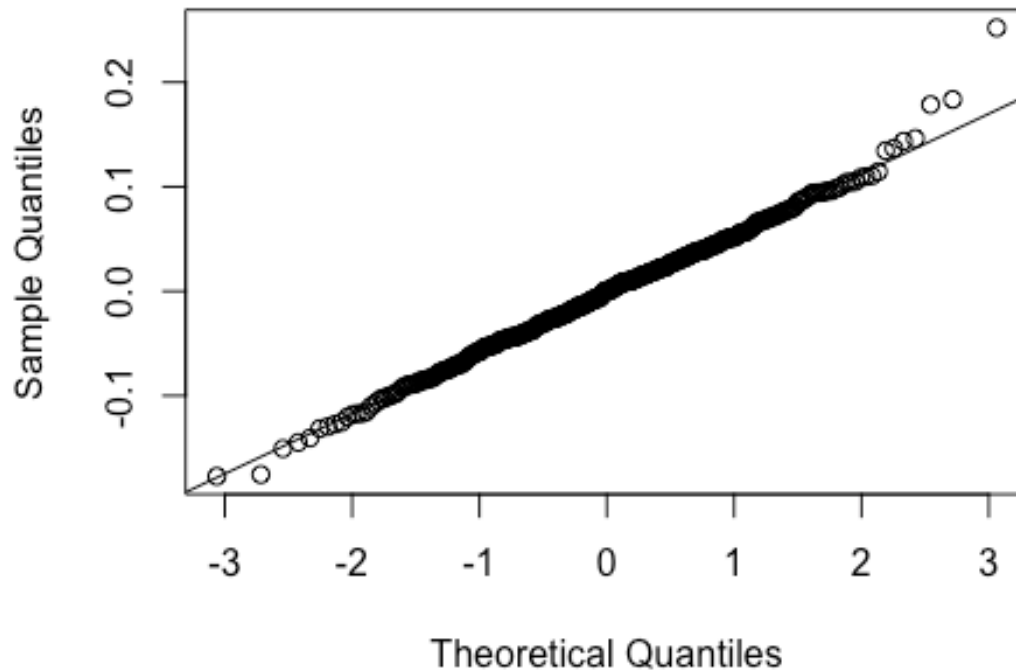
#Construct plots of the testing and trained data residuals to check
normality condition.
plot(cbb_final_mod, 2)
```



$t \sim \text{ADJOE} + \text{EFG_D} + \text{CONF} + \text{ADJDE} + \text{EFG_O} + \text{TOR} + \text{ORB} + /$

```
qqnorm(cbb_test_resid)  
qqline(cbb_test_resid)
```

Normal Q-Q Plot



```
#Compute shrinkage; square the cross-validation correlation and  
subtract it from the multiple R squared of the training sample.  
shrinkage = summary(cbb_mod1)$r.squared - cor(cbb_test$win_pct,  
fit_win_pct)^2  
shrinkage
```

```
## [1] -0.02446831
```

```
#Produce a plot displaying relationship between actual win percentage  
and predicted win percentage.
```

```
yhat = predict(cbb_final_mod, newdata = cbb_test)  
win.pct.test = cbb_test$win_pct  
plot(win.pct.test, yhat, ylab = "Predicted Win Percentage", xlab =  
"Actual Win Percentage")  
abline(0,1, col = 'red')
```

