Developing Ordinal Regression Model to Predict NCAA Basketball Champions

Kevin Liu & Evan Zhong

```
Call:
polr(formula = as.factor(POSTSEASON) ~ SEED + G + ADJOE + ADJDE + 
    BARTHAG + EFG_O + EFG_D + TOR + TORD + ORB + DRB + FTR + 
    FTRD + X2P_O + X2P_D + X3P_O + X3P_D + ADJ_T + WAB, data = cbb)

Coefficients:
           Value Std. Error t value
SEED10  -0.77872    0.65259 -1.1933
SEED11  -1.44612    0.69545 -2.0794
SEED12   0.38092    0.72749  0.5236
SEED13   0.56780    0.81653  0.6954
SEED14   1.09899    0.87432  1.2570
SEED15   1.60809    0.95744  1.6796
SEED16   0.61143    1.11224  0.5497
SEED2   -0.26174    0.42646 -0.6137
SEED3    0.23838    0.46428  0.5135
SEED4    0.25009    0.50149  0.4987
SEED5   -0.61760    0.54695 -1.1292
SEED6   -0.66718    0.57784 -1.1546
SEED7    0.29647    0.60877  0.4870
SEED8   -0.22555    0.62342 -0.3618
SEED9   -0.20827    0.63795 -0.3265
G        0.17763    0.02709  6.5571
ADJOE    0.26490    0.06135  4.3181
ADJDE   -0.34998    0.05213 -6.7136
BARTHAG -3.73692    2.29068 -1.6314
EFG_O    0.24480    0.37013  0.6614
EFG_D   -0.95142    0.40415 -2.3541
TOR      0.02755    0.06615  0.4165
TORD    -0.01738    0.05669 -0.3065
ORB     -0.01191    0.03345 -0.3560
```

```
DRB      0.06271    0.03818  1.6423
FTR     -0.04295    0.01728 -2.4860
FTRD    -0.01786    0.01795 -0.9948
X2P_O   -0.13490    0.23451 -0.5753
X2P_D    0.70235    0.25956  2.7059
X3P_O   -0.17317    0.19604 -0.8833
X3P_D    0.54320    0.21588  2.5162
ADJ_T   -0.02015    0.02751 -0.7326
WAB     -0.00861    0.05653 -0.1523

Intercepts:
    Value   Std. Error t value
1|2 -2.5125  1.2609    -1.9927
2|3  1.9201  1.2541     1.5310
3|4  3.7441  1.2666     2.9560
4|5  5.1197  1.2870     3.9779
5|6  6.2252  1.3079     4.7597
6|7  7.2249  1.3347     5.4132
7|8  8.1635  1.3735     5.9437

Residual Deviance: 1538.581
AIC: 1618.581
```

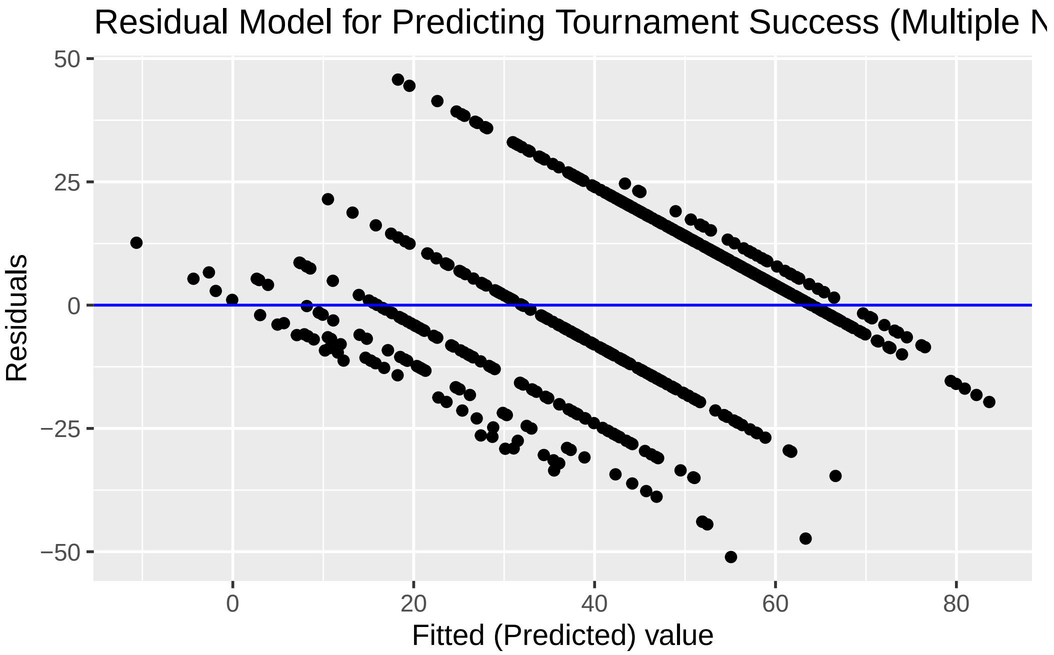
The coefficient corresponding to 8 seed is -0.22555. This means that when controlling for the other variables, if a team is an 8 seed as opposed to a 1 seed, then they are predicted to have e^-0.22555 (0.798077156) times the odds of making the next round of the NCAA tournament

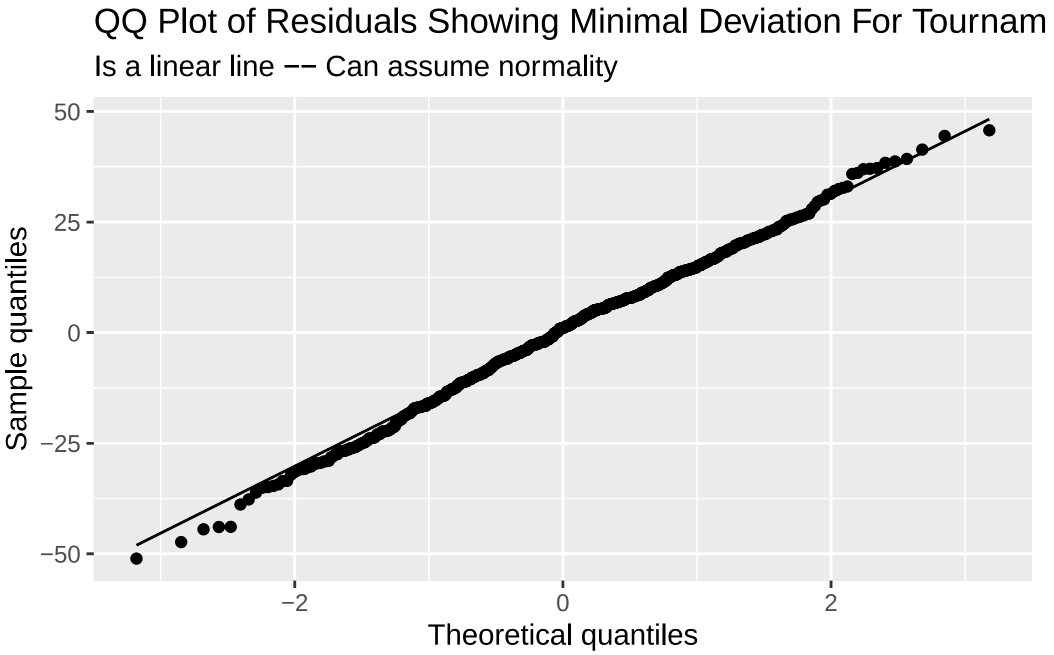
The coefficient corresponding to adjusted offensive efficiency is 0.26490. This means that when controlling for the other variables, for team with one more adjusted offensive efficiency rating point compared to a team without that extra point, the team with the extra point is predicted to have e^-0.26490 (1.303) times the odds of making the next round of the NCAA tournament

```
Call:
lm(formula = POSTSEASON ~ CONF + G + ADJOE + ADJDE + BARTHAG +
    EFG_O + EFG_D + TOR + TORD + ORB + DRB + FTR + FTRD + X2P_O +
    X2P_D + X3P_O + X3P_D + ADJ_T + WAB + SEED, data = cbbAug)

Residuals:
    Min      1Q  Median      3Q     Max
-51.078 -10.115   1.081  10.312  45.735
```

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  42.63047   42.59056   1.001   0.3173
CONFACC       0.64398    4.11381   0.157   0.8757
CONFAE       -2.17712    7.39514  -0.294   0.7686
CONFAmer      2.08825    4.74817   0.440   0.6602
CONFASun     -7.91012    7.56494  -1.046   0.2961
CONFB10       1.32210    4.04554   0.327   0.7439
CONFB12       3.91908    4.15344   0.944   0.3458
CONFBE        4.41670    4.07879   1.083   0.2793
CONFBSky     -3.86337    7.49384  -0.516   0.6064
CONFBSth     -7.08763    7.15941  -0.990   0.3226
CONFBW       -6.51771    7.12542  -0.915   0.3607
CONFCAA       2.84082    6.98531   0.407   0.6844
CONFCUSA    -17.04497    6.66582  -2.557   0.0108 *
CONFHorz      2.49330    7.26323   0.343   0.7315
CONFIvy     -13.53415    7.18654  -1.883   0.0601 .
CONFMAAC     -3.81782    7.17787  -0.532   0.5950
CONFMAC      -3.84109    6.83948  -0.562   0.5746
CONFMEAC     -9.64269    8.26950  -1.166   0.2440
CONFMVC     -10.13881    5.57819  -1.818   0.0696 .
CONFMWC       4.65357    4.57508   1.017   0.3095
CONFNEC      -6.97788    8.11885  -0.859   0.3904
CONFOVC      -1.98486    6.64980  -0.298   0.7654
CONFP12      -5.21460    4.12764  -1.263   0.2069
CONFPat       1.08669    7.27052   0.149   0.8812
CONFSB       -2.53574    6.74266  -0.376   0.7070
CONFSC       -0.41678    6.95959  -0.060   0.9523
CONFSEC       1.06662    4.10962   0.260   0.7953
CONFSlnd    -10.16717    7.58226  -1.341   0.1804
CONFSum      -8.87361    7.04127  -1.260   0.2081
CONFSWAC     -7.28974    8.20248  -0.889   0.3745
CONFWAC      -0.24855    7.15668  -0.035   0.9723
CONFWCC       1.15853    5.27947   0.219   0.8264
G            -1.19944    0.20587  -5.826 9.15e-09 ***
ADJOE        -2.53446    0.56246  -4.506 7.91e-06 ***
ADJDE         3.13359    0.67600   4.635 4.35e-06 ***
BARTHAG      51.11117   29.12980   1.755   0.0798 .
EFG_O        -0.90335    3.12862  -0.289   0.7729
EFG_D         6.78706    3.63701   1.866   0.0625 .
TOR          -0.33699    0.64765  -0.520   0.6030
TORD          0.36945    0.58650   0.630   0.5290
ORB           0.13453    0.30883   0.436   0.6633
```

```
DRB          -0.77167    0.36505  -2.114   0.0349 *
FTR           0.35529    0.15069   2.358   0.0187 *
FTRD          0.10691    0.15559   0.687   0.4923
X2P_O         0.43196    1.97413   0.219   0.8269
X2P_D        -5.41850    2.29896  -2.357   0.0187 *
X3P_O         1.21432    1.66637   0.729   0.4664
X3P_D        -3.74556    1.93066  -1.940   0.0528 .
ADJ_T         0.18139    0.24090   0.753   0.4517
WAB           0.36522    0.49147   0.743   0.4577
SEED10       11.20725    5.89458   1.901   0.0577 .
SEED11        8.00954    5.96829   1.342   0.1801
SEED12        5.71098    6.77083   0.843   0.3993
SEED13        5.73418    7.73757   0.741   0.4589
SEED14        4.32954    8.25426   0.525   0.6001
SEED15        0.30975    8.77817   0.035   0.9719
SEED16        2.93267   10.10204   0.290   0.7717
SEED2         0.43149    3.97865   0.108   0.9137
SEED3        -4.39253    4.38892  -1.001   0.3173
SEED4        -2.80058    4.67726  -0.599   0.5495
SEED5         5.82495    4.97454   1.171   0.2421
SEED6         8.21792    5.26494   1.561   0.1191
SEED7         0.06631    5.57294   0.012   0.9905
SEED8         5.63351    5.69805   0.989   0.3232
SEED9         5.39142    5.83444   0.924   0.3558
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 16.53 on 615 degrees of freedom
Multiple R-squared:  0.5289,    Adjusted R-squared:  0.4799
F-statistic: 10.79 on 64 and 615 DF,  p-value: < 2.2e-16
```





Assumptions needed for the linear model: Independence, Linearity, Constant Variance, and Normal Distribution

1. Independence The dataset includes records of NCAA March Madness Tournament team

performances (playoff runs) from 2013 to 2021. These observations–tournament runs—are NOT independent. Several factors contribute to this–for example, teams retaining players across seasons, powerhouse teams that show dominance throughout the years, the trading of players between teams and more. Thus, a team's performance in one tournament could influence its performance in future tournaments as well as other teams in the current tournament. Therefore, the assumption of independence is not met.

2. Linearity Linearity is not satisfied because the residuals are not symmetrically distributed along y = 0. The data points follow several very strong negative trends. This could be for many reasons, most likely because there are interaction effects. All in all though, the residual plot is not symmetric and therefore does not satisfy linearity.
3. Constant Variance Constant variance is not satisfied because the data set is not consistent/spreadout across all the predicted values. Although there is no signicant clumps, there are clear trends. Based on these features, we can conclude that the variance is not independent from the predictors
4. Normal Distribution Because the QQ plot follows the linear line, therefore we cannot assume normality.

Interpretations

Holding all other variables constant, if a team is in Conference USA we'd expect them to place 17 higher than a team in the West Coast Conference

Holding all other variables constant, then, on average, for every additional point of adjusted offensive efficiency, we'd expect a team to place 2.53446 ranks higher

Holding all other variables constant, then, on average, for every additional point of adjusted defensive efficiency, we'd expect a team to place 3.13359 ranks higher

Hypothesis Test: Null Hypothesis, H0: there is not a linear relationship between 2-pt field goal percentage and how far a team makes it in the NCAA tournament. Alternative Hypothesis, Ha: there is a linear relationship between 2-pt field goal percentage and how far a team makes it in the NCAA tournament.

Given that there is no relationship between 2-pt field goal percentage and how far a team makes it in the NCAA tournament, the probability of getting a slope of -5.41850 or less is 0.0187. Taking a 0.05 significance level, since the p-value is smaller than alpha (0.0187<0.05), there is convincing and sufficient evidence that there is a linear relationship between 2-pt field goal percentage and how far a team makes it in the NCAA tournament.

```
[1] 0.249834
```

Moderate Explanatory Power: A McFadden's R-squared value of about 0.25 suggests that the model has moderate explanatory power. While it's not particularly high (such as values closer to 0.5 or above), it's substantial enough to suggest that the model does provide valuable insights into the relationship between the conference (CONF) and the postseason outcomes (POSTSEASON). ˆ