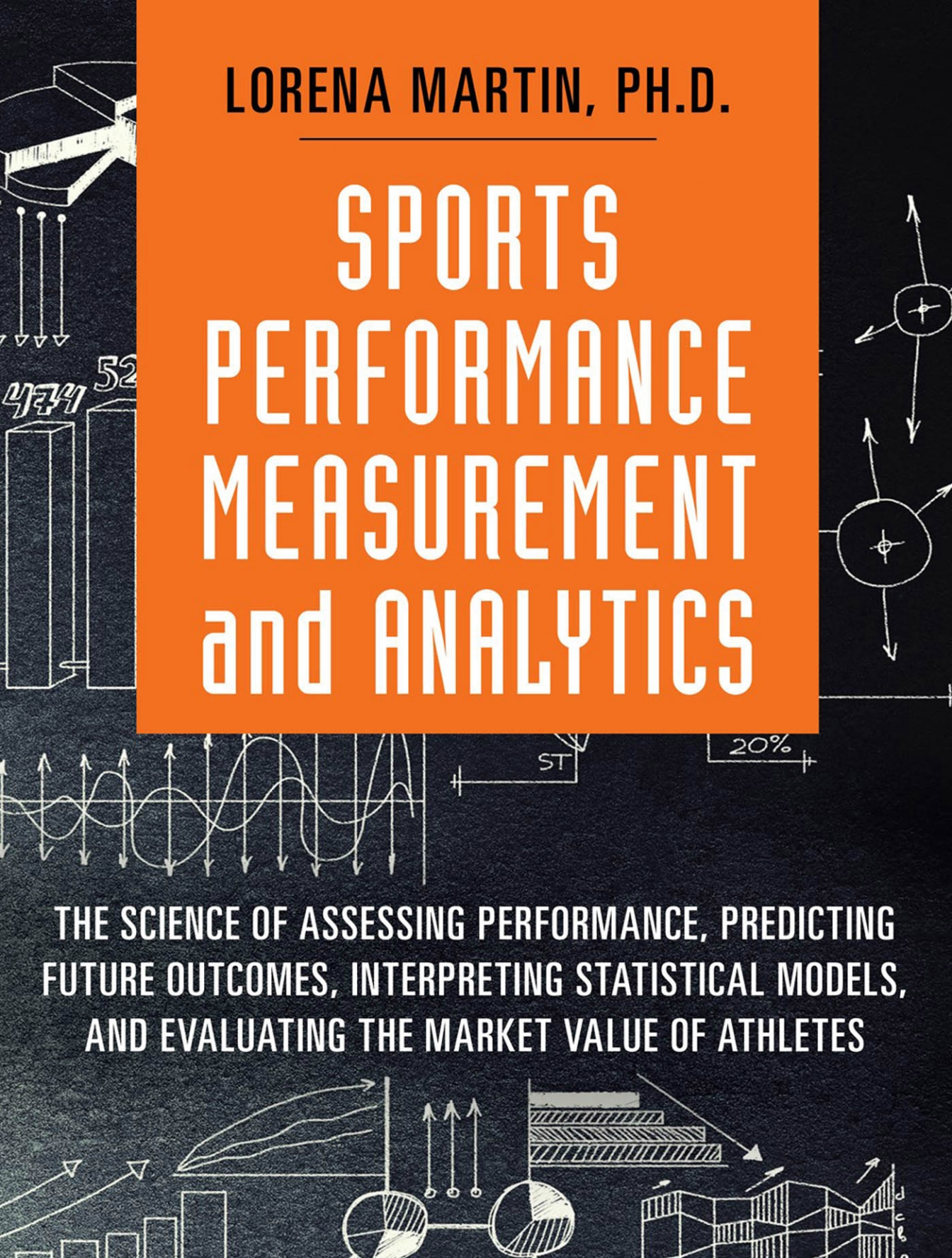
LORENA MARTIN, PH.D.

SPORTS PERFORMANCE MEASUREMENT and ANALYTICS

THE SCIENCE OF ASSESSING PERFORMANCE, PREDICTING FUTURE OUTCOMES, INTERPRETING STATISTICAL MODELS, AND EVALUATING THE MARKET VALUE OF ATHLETES



Sports Performance Measurement and Analytics

The Science of Assessing Performance,

Predicting Future Outcomes,

Interpreting Statistical Models,

and Evaluating the Market Value of Athletes

LORENA MARTIN

Exhibit 5.1. *Analyzing NFL Combine Measures (R)*

```
# Analyzing NFL Combine Measures (R)

##########################################
##      Touchdown Analytics R code      ##
##########################################

NFLcombine<-read.csv(file.choose())

#########################################################
##   NFL Combine Performance and Player Positions   ##
#########################################################

summary(NFLcombine)
# NAME     Year...4..2014      POS        X
# Aaron Corp     :  1   Min.   :0.00   QB     :75   Mode:logical
# Aaron Murray   :  1   1st Qu.:2.00   DT     :26   NA's:166
# Andrew Luck    :  1   Median :4.00   OT     :19
# Andy Dalton    :  1   Mean   :2.97   OG     :15
# Anthony Johnson:  1   3rd Qu.:4.00   C      :12
# Anthony Steen  :  1   Max.   :4.00   DE     : 7
# (Other)        :160                  (Other):12

#   X40.SPEED          X3CONE         SHUTTLE         VERTICAL
# Min.   :4.410   Min.   :6.640   Min.   :3.990   Min.   :20.50
# 1st Qu.:4.900   1st Qu.:7.043   1st Qu.:4.280   1st Qu.:27.00
# Median :5.010   Median :7.335   Median :4.440   Median :29.00
# Mean   :5.021   Mean   :7.406   Mean   :4.458   Mean   :29.59
# 3rd Qu.:5.190   3rd Qu.:7.800   3rd Qu.:4.620   3rd Qu.:32.50
# Max.   :5.590   Max.   :8.290   Max.   :5.160   Max.   :40.00
# NA's   :1       NA's   :52      NA's   :71      NA's   :34

#    BROAD           BENCH
# Min.   : 88.0   Min.   :14.00
# 1st Qu.:102.0   1st Qu.:21.00
# Median :108.0   Median :25.00
# Mean   :107.7   Mean   :25.23
# 3rd Qu.:113.0   3rd Qu.:28.00
# Max.   :126.0   Max.   :42.00
# NA's   :33      NA's   :87

# Let us examine shuttle speed by different player positions
# We will use ANOVA as the independent variable is categorical
# and the dependent variable is continuous

SHUTTLEbyPOSANOVAModel<-aov(NFLcombine$SHUTTLE~NFLcombine$POS)
summary(SHUTTLEbyPOSANOVAModel)
# Df Sum Sq Mean Sq F value   Pr(>F)
# NFLcombine$POS  6  3.418  0.5697   21.36 2.52e-15 ***
# Residuals      88  2.347  0.0267
# Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

```
# After running an ANOVA if there are differences between groups
# as was evident from our p value
# It is customary to run post hoc analyses
# Tukey and Games Howell are preferred
TukeyHSD(SHUTTLEbyPOSANOVAModel)

# Here is the output of comparisons obtained from the post hoc
# Tukey multiple comparisons of means
# 95% family-wise confidence level

# Fit: aov(formula = NFLcombine$SHUTTLE ~ NFLcombine$POS)

# $`NFLcombine$POS`
#             diff         lwr         upr     p adj
# FB-C   -0.0422222222 -0.56170510  0.47726065 0.9999809
# OG-C    0.1032323232 -0.11827592  0.32474057 0.7970296
# OLB-C  -0.4722222222 -0.99170510  0.04726065 0.0997784
# OT-C    0.1037777778 -0.10401537  0.31157093 0.7398832
# QB-C   -0.3094152047 -0.48618407 -0.13264634 0.0000186
# TE-C   -0.1522222222 -0.67170510  0.36726065 0.9740431
# OG-FB   0.1454545455 -0.36928409  0.66019318 0.9783760
# OLB-FB -0.4300000000 -1.12695941  0.26695941 0.5108420
# OT-FB   0.1460000000 -0.36298719  0.65498719 0.9766767
# QB-FB  -0.2671929825 -0.76432194  0.22993597 0.6686972
# TE-FB  -0.1100000000 -0.80695941  0.58695941 0.9990874
# OLB-OG -0.5754545455 -1.09019318 -0.06071591 0.0183342
# OT-OG   0.0005454545 -0.19508533  0.19617624 1.0000000
# QB-OG  -0.4126475279 -0.57494553 -0.25034953 0.0000000
# TE-OG  -0.2554545455 -0.77019318  0.25928409 0.7455144
# OT-OLB  0.5760000000  0.06701281  1.08498719 0.0161997
# QB-OLB  0.1628070175 -0.33432194  0.65993597 0.9553029
# TE-OLB  0.3200000000 -0.37695941  1.01695941 0.8080715
# QB-OT  -0.4131929825 -0.55620604 -0.27017993 0.0000000
# TE-OT  -0.2560000000 -0.76498719  0.25298719 0.7335129
# TE-QB   0.1571929825 -0.33993597  0.65432194 0.9622422


# Let us run a regression model to examine if our results would differ
# from the ANOVA model
SHUTTLEregressionmodel<-lm(SHUTTLE~POS, data=NFLcombine)
summary(SHUTTLEregressionmodel)

# Call:
# lm(formula = SHUTTLE ~ POS, data = NFLcombine)

# Residuals:
#    Min       1Q   Median       3Q      Max
# -0.32600 -0.13281  0.00778  0.10927  0.43455

# Coefficients:
#              Estimate Std. Error t value Pr(>|t|)
# (Intercept)  4.62222    0.05444  84.907  < 2e-16 ***
```

```
# POSFB       -0.04222    0.17215  -0.245  0.80682
# POSOG        0.10323    0.07340   1.406  0.16314
# POSOLB      -0.47222    0.17215  -2.743  0.00737 **
# POSOT        0.10378    0.06886   1.507  0.13537
# POSQB       -0.30942    0.05858  -5.282 9.15e-07 ***
# POSTE       -0.15222    0.17215  -0.884  0.37898

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

# Residual standard error: 0.1633 on 88 degrees of freedom
# Multiple R-squared:  0.5929,  Adjusted R-squared:  0.5651
# F-statistic: 21.36 on 6 and 88 DF,  p-value: 2.516e-15
# The results reveal significant differences as did the ANOVA model
# Differences between the models include the number of comparisons
# As well as the values that are obtained from the output
# If interested in examining group mean differences go with ANOVA
# If interested in examining slopes and intercepts regression is a better model

# Run analyses to examine differences by player position on vertical jump

VERTICALregressionmodel<-lm(VERTICAL~POS, data=NFLcombine)
summary(VERTICALregressionmodel)

# Call:
# lm(formula = VERTICAL ~ POS, data = NFLcombine)

# Residuals:
#    Min      1Q  Median      3Q     Max
# -7.6500 -1.9841  0.0159  1.8165  8.5159

# Coefficients:
#             Estimate Std. Error t value Pr(>|t|)
# (Intercept)  26.6500     1.0037  26.553  < 2e-16 ***
# POSDE         4.0167     1.6390   2.451  0.01567 *
# POSDT         1.5000     1.2292   1.220  0.22471
# POSFB         1.8500     3.3287   0.556  0.57939
# POSILB        8.1000     2.4584   3.295  0.00129 **
# POSOG        -0.3167     1.3590  -0.233  0.81613
# POSOLB        5.3500     3.3287   1.607  0.11059
# POSOT         0.2250     1.2794   0.176  0.86069
# POSQB         4.8341     1.0804   4.474 1.73e-05 ***
# POSTE         5.8500     3.3287   1.757  0.08135 .

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

# Residual standard error: 3.174 on 122 degrees of freedom
# Multiple R-squared:  0.3545,  Adjusted R-squared:  0.3068
# F-statistic: 7.443 on 9 and 122 DF,  p-value: 1.294e-08

# Run an ANOVA model on vertical jump by player position
VERTICALANOVAmodel<-aov(VERTICAL ~ POS, data = NFLcombine)
```

```
# Get the summary of the ANOVA model
summary(VERTICALANOVAmodel)

# Df Sum Sq Mean Sq F value   Pr(>F)
# POS           9  674.8   74.98   7.443 1.29e-08 ***
# Residuals   122 1228.9   10.07

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

# Conduct post hoc analyses to examine which player positions
# differ on vertical jump
TukeyHSD(VERTICALANOVAmodel)

# Tukey multiple comparisons of means
# 95% family-wise confidence level

# Fit: aov(formula = VERTICAL ~ POS, data = NFLcombine)

# $POS
# diff         lwr        upr     p adj
# DE-C     4.0166667  -1.2658413  9.2991746 0.3059528
# DT-C     1.5000000  -2.4618809  5.4618809 0.9677442
# FB-C     1.8500000  -8.8788243 12.5788243 0.9999217
# ILB-C    8.1000000   0.1762381 16.0237619 0.0407549
# OG-C    -0.3166667  -4.6966908  4.0633575 1.0000000
# OLB-C    5.3500000  -5.3788243 16.0788243 0.8422035
# OT-C     0.2250000  -3.8986564  4.3486564 1.0000000
# QB-C     4.8341270   1.3519812  8.3162727 0.0007077
# TE-C     5.8500000  -4.8788243 16.5788243 0.7600416
# DT-DE   -2.5166667  -7.2782550  2.2449216 0.7914253
# FB-DE   -2.1666667 -13.2158247  8.8824914 0.9997712
# ILB-DE   4.0833333  -4.2690451 12.4357117 0.8571960
# OG-DE   -4.3333333  -9.4480996  0.7814330 0.1733884
# OLB-DE   1.3333333  -9.7158247 12.3824914 0.9999963
# OT-DE   -3.7916667  -8.6886826  1.1053492 0.2814152
# QB-DE    0.8174603  -3.5530730  5.1879936 0.9998454
# TE-DE    1.8333333  -9.2158247 12.8824914 0.9999434
# FB-DT    0.3500000 -10.1321517 10.8321517 1.0000000
# ILB-DT   6.6000000  -0.9864244 14.1864244 0.1468270
# OG-DT   -1.8166667  -5.5519638  1.9186305 0.8609423
# OLB-DT   3.8500000  -6.6321517 14.3321517 0.9735410
# OT-DT   -1.2750000  -4.7060895  2.1560895 0.9714333
# QB-DT    3.3341270   0.7086418  5.9596121 0.0029812
# TE-DT    4.3500000  -6.1321517 14.8321517 0.9427739
# ILB-FB   6.2500000  -6.2785676 18.7785676 0.8418830
# OG-FB   -2.1666667 -12.8139018  8.4805684 0.9996896
# OLB-FB   3.5000000 -10.9667437 17.9667437 0.9987506
# OT-FB   -1.6250000 -12.1693609  8.9193609 0.9999698
# QB-FB    2.9841270  -7.3262728 13.2945267 0.9950577
# TE-FB    4.0000000 -10.4667437 18.4667437 0.9964925
# OG-ILB  -8.4166667 -16.2296012 -0.6037321 0.0240101
# OLB-ILB -2.7500000 -15.2785676  9.7785676 0.9994254
```

```
# OT-ILB  -7.8750000 -15.5471495 -0.2028505 0.0391878
# QB-ILB  -3.2658730 -10.6131633  4.0814172 0.9144899
# TE-ILB  -2.2500000 -14.7785676 10.2785676 0.9998898
# OLB-OG   5.6666667  -4.9805684 16.3139018 0.7846814
# OT-OG    0.5416667  -3.3648006  4.4481340 0.9999878
# QB-OG    5.1507937   1.9287937  8.3727936 0.0000431
# TE-OG    6.1666667  -4.4805684 16.8139018 0.6911095
# OT-OLB  -5.1250000 -15.6693609  5.4193609 0.8613964
# QB-OLB  -0.5158730 -10.8262728  9.7945267 1.0000000
# TE-OLB   0.5000000 -13.9667437 14.9667437 1.0000000
# QB-OT    4.6091270   1.7453508  7.4729032 0.0000371
# TE-OT    5.6250000  -4.9193609 16.1693609 0.7823877
# TE-QB    1.0158730  -9.2945267 11.3262728 0.9999994

# Let us run a predictive model of regression on speed
# as measured by the 40-yard dash
# with player position as the predictor variable
X40.SPEEDregressionmodel<-lm(X40.SPEED~POS, data=NFLcombine)

# Get a summary of the data
summary(X40.SPEEDregressionmodel)

# Call:
# lm(formula = X40.SPEED ~ POS, data = NFLcombine)

# Residuals:
# Min      1Q  Median      3Q     Max
# -0.4332 -0.1000  0.0000  0.1090  0.4568

# Coefficients:
#              Estimate Std.Error t value Pr(>|t|)
# (Intercept)  5.21417    0.04881 106.830  < 2e-16 ***
# POSDE       -0.21845    0.08041  -2.717  0.00735 **
# POSDT       -0.07378    0.05901  -1.250  0.21306
# POSFB       -0.26417    0.17598  -1.501  0.13538
# POSILB      -0.19417    0.10914  -1.779  0.07721 .
# POSOG        0.06717    0.06548   1.026  0.30664
# POSOLB      -0.26417    0.17598  -1.501  0.13538
# POSOT        0.01689    0.06234   0.271  0.78687
# POSPK       -0.01417    0.12913  -0.110  0.91279
# POSPT       -0.21417    0.10914  -1.962  0.05154 .
# POSQB       -0.37092    0.05262  -7.050 5.77e-11 ***
# POSTE       -0.25417    0.12913  -1.968  0.05085 .

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

# Residual standard error: 0.1691 on 153 degrees of freedom
# Multiple R-squared:  0.5383,  Adjusted R-squared:  0.5051
# F-statistic: 16.22 on 11 and 153 DF,  p-value: < 2.2e-16

# The output from the regression yields the
# Estimate, standard error, t value and the p value
# To examine how a regression model differs from the ANOVA model
```

```
# Let us run an ANOVA model on speed using the 40-yard dash by player positions
X40.SPEEDANOVAmodel<-aov(X40.SPEED~POS, data=NFLcombine)
summary(X40.SPEEDANOVAmodel)

# Df Sum Sq Mean Sq F value Pr(>F)
# POS          11  5.099  0.4635   16.21 <2e-16 ***
# Residuals   153  4.374  0.0286

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

# The results of ANOVA model indicate that there are significant differences
# on 40-yard dash speed by player positions
# however additional analyses will need to be run to determine the location
# Let us run Tukey post hoc analyses

TukeyHSD(X40.SPEEDANOVAmodel)
# Tukey multiple comparisons of means
# 95% family-wise confidence level

# Fit: aov(formula = X40.SPEED ~ POS, data = NFLcombine)

# $POS
# diff         lwr         upr     p adj
# DE-C    -2.184524e-01 -0.48536657  0.04846181 0.2283696
# DT-C    -7.378205e-02 -0.26964352  0.12207942 0.9837284
# FB-C    -2.641667e-01 -0.84830480  0.31997147 0.9382295
# ILB-C   -1.941667e-01 -0.55643376  0.16810043 0.8267502
# OG-C     6.716667e-02 -0.15019359  0.28452692 0.9968830
# OLB-C   -2.641667e-01 -0.84830480  0.31997147 0.9382295
# OT-C     1.688596e-02 -0.19005570  0.22382763 1.0000000
# PK-C    -1.416667e-02 -0.44280687  0.41447354 1.0000000
# PT-C    -2.141667e-01 -0.57643376  0.14810043 0.7177360
# QB-C    -3.709234e-01 -0.54557691 -0.19626994 0.0000000
# TE-C    -2.541667e-01 -0.68280687  0.17447354 0.7138677
# DT-DE    1.446703e-01 -0.09430649  0.38364715 0.6863498
# FB-DE   -4.571429e-02 -0.64568559  0.55425702 1.0000000
# ILB-DE   2.428571e-02 -0.36299410  0.41156553 1.0000000
# OG-DE    2.856190e-01  0.02872668  0.54251141 0.0157542
# OLB-DE  -4.571429e-02 -0.64568559  0.55425702 1.0000000
# OT-DE    2.353383e-01 -0.01280084  0.48347754 0.0807664
# PK-DE    2.042857e-01 -0.24569277  0.65426419 0.9366058
# PT-DE    4.285714e-03 -0.38299410  0.39156553 1.0000000
# QB-DE   -1.524710e-01 -0.37439906  0.06945698 0.4949289
# TE-DE   -3.571429e-02 -0.48569277  0.41426419 1.0000000
# FB-DT   -1.903846e-01 -0.76229728  0.38152805 0.9940736
# ILB-DT  -1.203846e-01 -0.46258950  0.22182027 0.9906072
# OG-DT    1.409487e-01 -0.04101891  0.32291635 0.3055013
# OLB-DT  -1.903846e-01 -0.76229728  0.38152805 0.9940736
# OT-DT    9.066802e-02 -0.07871790  0.26005393 0.8279721
# PK-DT    5.961538e-02 -0.35220880  0.47143957 0.9999982
# PT-DT   -1.403846e-01 -0.48258950  0.20182027 0.9688814
# QB-DT   -2.971414e-01 -0.42508892 -0.16919382 0.0000000
```