something about sports lol

Arthur Jakobsson, Elin Jang, Prathik Guduri

2023-09-01

```
dbd$plays = dbd$num_run + dbd$num_pass + dbd$num_no_play + dbd$num_qb_kneel + dbd$num_qb_s
dbd$runPercent = (dbd$num_run / dbd$plays) * 100
dbd$passPercent = (dbd$num_pass / dbd$plays) * 100
unique(dbd$fixed_drive_result)
```

```
[1] "Punt"              "Touchdown"         "Field goal"
[4] "Turnover"          "Opp touchdown"     "Turnover on downs"
[7] "End of half"       "Missed field goal" "Safety"
```

```
dbd <- dbd %>%
  mutate(drivePoints = case_when(
    fixed_drive_result == "Punt" ~ 0,
    fixed_drive_result == "Touchdown" ~ 7,
    fixed_drive_result == "Opp touchdown" ~ -7,
    fixed_drive_result == "Field goal" ~ -3,
    fixed_drive_result == "Turnover" ~ 0,
    fixed_drive_result == "Turnover on downs" ~ 0,
    fixed_drive_result == "End of half" ~ 0,
    fixed_drive_result == "Missed field goal" ~ 0,
    fixed_drive_result == "Safety" ~ -2,
    TRUE ~ NA_real_  # This line handles any values not explicitly mentioned
  ))
```

```
lmModel = lm(drivePoints ~ runPercent + passPercent + temperature + posteam_type, data = d
summary(lmModel)
```

```
Call:
lm(formula = drivePoints ~ runPercent + passPercent + temperature +
    posteam_type, data = dbd)

Residuals:
    Min      1Q  Median      3Q     Max
-8.8776 -1.3354 -0.8333 -0.1553  7.0744

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)      -0.831232   0.253006  -3.285  0.00102 **
runPercent        0.023455   0.002342  10.015  < 2e-16 ***
passPercent       0.009805   0.002168   4.523 6.21e-06 ***
temperature       0.004761   0.002761   1.724  0.08473 .
posteam_typehome  0.182461   0.085291   2.139  0.03245 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.301 on 5991 degrees of freedom
  (318 observations deleted due to missingness)
Multiple R-squared:  0.01885,   Adjusted R-squared:  0.01819
F-statistic: 28.77 on 4 and 5991 DF,  p-value: < 2.2e-16
```

```r
dbd_no_neg = dbd
dbd_no_neg$total_yards_gained[dbd_no_neg$total_yards_gained<0] <- 0
dbd_no_neg$drivePoints[dbd_no_neg$drivePoints<0] <- 0
pbp$yds
```

```
Warning: Unknown or uninitialised column: `yds`.
```

```
NULL
```

```r
# ZIP_model <- zeroinfl(total_yards_gained ~yardline_100+runPercent+passPercent+as.factor(
# summary(ZIP_model)

pbp$yards_gained[pbp$yards_gained<0] <- 0
pbp <- pbp %>% drop_na(drive) %>% drop_na(play_type) %>% drop_na(posteam) %>% filter(play_

ZIP_model <- zeroinfl(yards_gained ~ ydstogo+play_type+as.factor(posteam_type), data=pbp,
summary(ZIP_model)
```

```
Call:
zeroinfl(formula = yards_gained ~ ydstogo + play_type + as.factor(posteam_type),
    data = pbp, dist = "poisson", link = "logit")

Pearson residuals:
    Min      1Q  Median      3Q     Max
-1.8927 -1.0479 -0.4999  0.4617 25.8746

Count model coefficients (poisson with log link):
                            Estimate Std. Error  z value Pr(>|z|)
(Intercept)                2.3435220  0.0061952  378.280  < 2e-16 ***
ydstogo                    0.0100498  0.0005532   18.167  < 2e-16 ***
play_typerun              -0.7291020  0.0047580 -153.237  < 2e-16 ***
as.factor(posteam_type)home  0.0267698  0.0044400    6.029 1.65e-09 ***

Zero-inflation model coefficients (binomial with logit link):
                            Estimate Std. Error z value Pr(>|z|)
(Intercept)                 0.006952   0.032295   0.215 0.829557
ydstogo                    -0.028387   0.003023  -9.389  < 2e-16 ***
play_typerun               -1.318099   0.026344 -50.034  < 2e-16 ***
```

```
as.factor(posteam_type)home -0.080790   0.023697  -3.409 0.000651 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Number of iterations in BFGS optimization: 11
Log-likelihood: -1.29e+05 on 8 Df
```