BSA Football Research Dist Fitting

Naren Prakash

2024-11-04

```
pbp_2018 <- read_csv("pbp_merged_2018.csv")
```

```
## New names:
## Rows: 17783 Columns: 59
## -- Column specification
## -------------------------------------------------------- Delimiter: "," chr
## (20): game_id, home_team, away_team, season_type, posteam, posteam_type... dbl
## (34): ...1, Unnamed: 0_pbp, play_id, old_game_id, week, yardline_100, q... lgl
## (3): run_location, run_gap, was_pressure date (1): game_date time (1): time
## i Use `spec()` to retrieve the full column specification for this data. i
## Specify the column types or set `show_col_types = FALSE` to quiet this message.
## * `` -> `...1`
```

```
pbp_2019 <- read_csv("pbp_merged_2019.csv")
```

```
## New names:
## Rows: 17730 Columns: 59
## -- Column specification
## -------------------------------------------------------- Delimiter: "," chr
## (20): game_id, home_team, away_team, season_type, posteam, posteam_type... dbl
## (34): ...1, Unnamed: 0_pbp, play_id, old_game_id, week, yardline_100, q... lgl
## (3): run_location, run_gap, was_pressure date (1): game_date time (1): time
## i Use `spec()` to retrieve the full column specification for this data. i
## Specify the column types or set `show_col_types = FALSE` to quiet this message.
## * `` -> `...1`
```

```
pbp_2020 <- read_csv("pbp_merged_2020.csv")
```

```
## New names:
## Rows: 18142 Columns: 59
## -- Column specification
## -------------------------------------------------------- Delimiter: "," chr
## (20): game_id, home_team, away_team, season_type, posteam, posteam_type... dbl
## (34): ...1, Unnamed: 0_pbp, play_id, old_game_id, week, yardline_100, q... lgl
## (3): run_location, run_gap, was_pressure date (1): game_date time (1): time
## i Use `spec()` to retrieve the full column specification for this data. i
## Specify the column types or set `show_col_types = FALSE` to quiet this message.
## * `` -> `...1`
```

```
pbp_2021 <- read_csv("pbp_merged_2021.csv")
```

```
## New names:
## * `` -> `...1`
```

```
## Warning: One or more parsing issues, call `problems()` on your data frame for details,
```

```
## e.g.:
##   dat <- vroom(...)
##   problems(dat)
```

```
## Rows: 18849 Columns: 59
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (20): game_id, home_team, away_team, season_type, posteam, posteam_type...
## dbl  (34): ...1, Unnamed: 0_pbp, play_id, old_game_id, week, yardline_100, q...
## lgl   (3): run_location, run_gap, was_pressure
## date  (1): game_date
## time  (1): time
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
pbp_2022 <- read_csv("pbp_merged_2022.csv")
```

```
## New names:
## Rows: 18155 Columns: 59
## -- Column specification
## -------------------------------------------------------- Delimiter: "," chr
## (20): game_id, home_team, away_team, season_type, posteam, posteam_type... dbl
## (34): ...1, Unnamed: 0_pbp, play_id, old_game_id, week, yardline_100, q... lgl
## (3): run_location, run_gap, was_pressure date (1): game_date time (1): time
## i Use `spec()` to retrieve the full column specification for this data. i
## Specify the column types or set `show_col_types = FALSE` to quiet this message.
## * `` -> `...1`
```

```
pbp_2023 <- read_csv("pbp_merged_2023.csv")
```

```
## New names:
## Rows: 16000 Columns: 59
## -- Column specification
## -------------------------------------------------------- Delimiter: "," chr
## (20): game_id, home_team, away_team, season_type, posteam, posteam_type... dbl
## (34): ...1, Unnamed: 0_pbp, play_id, old_game_id, week, yardline_100, q... lgl
## (3): run_location, run_gap, was_pressure date (1): game_date time (1): time
## i Use `spec()` to retrieve the full column specification for this data. i
## Specify the column types or set `show_col_types = FALSE` to quiet this message.
## * `` -> `...1`
```

```
pbp_models <- list(pbp_2018, pbp_2019, pbp_2020, pbp_2021, pbp_2022, pbp_2023)
```

```
sub_1 <- drop_na(data.frame(bind_rows(pbp_models[1])$epa))
sub_2 <- drop_na(data.frame(bind_rows(pbp_models[1:2])$epa))
sub_3 <- drop_na(data.frame(bind_rows(pbp_models[1:3])$epa))
sub_4 <- drop_na(data.frame(bind_rows(pbp_models[1:4])$epa))
sub_5 <- drop_na(data.frame(bind_rows(pbp_models[1:5])$epa))
```

```
print(model_select(unlist(sub_1)))
```

```
## Maximum likelihood estimates for the Skew Student-t model
##   mean      sd      nu      xi
## 0.3195  1.7125  3.2501  1.3870
```

```
print(model_select(unlist(sub_2)))
```

```
## Maximum likelihood estimates for the Skew Student-t model
##   mean      sd      nu      xi
## 0.3166  1.7511  3.1917  1.3789
```

```
print(model_select(unlist(sub_3)))
```

```
## Maximum likelihood estimates for the Skew Student-t model
##   mean      sd      nu      xi
## 0.3172  1.7249  3.2608  1.3647
```

```
print(model_select(unlist(sub_4)))
```

```
## Maximum likelihood estimates for the Skew Student-t model
##   mean      sd      nu      xi
## 0.3121  1.7248  3.2515  1.3593
```

```
print(model_select(unlist(sub_5)))
```

```
## Maximum likelihood estimates for the Skew Student-t model
##   mean      sd      nu      xi
## 0.3036  1.7283  3.2307  1.3525
```

```
print(model_select(pbp_2023$epa))
```

```
## Maximum likelihood estimates for the Skew Student-t model
##   mean      sd      nu      xi
## 0.2714  1.8010  3.0694  1.3371
```