

lm_analysis

Waveley Qiu (wq2162)

2022-05-07

```
source("shared_code/data_cleaning.R")

## Warning: package 'tidyverse' was built under R version 4.1.2
## -- Attaching packages ----- tidyverse 1.3.1 --
## v ggplot2 3.3.5      v purrr 0.3.4
## v tibble 3.1.6       v dplyr 1.0.8
## v tidyr 1.2.0        v stringr 1.4.0
## v readr 2.0.1        v forcats 0.5.1
## Warning: package 'ggplot2' was built under R version 4.1.2
## Warning: package 'tibble' was built under R version 4.1.2
## Warning: package 'tidyr' was built under R version 4.1.2
## Warning: package 'dplyr' was built under R version 4.1.2
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag() masks stats::lag()
## Warning: package 'lubridate' was built under R version 4.1.3
##
## Attaching package: 'lubridate'
##
## The following objects are masked from 'package:base':
##
##   date, intersect, setdiff, union
##
## Rows: 22038 Columns: 8
## -- Column specification -----
## Delimiter: ","
## chr (4): ID, Month, Nature, time
## dbl (4): Season, Latitude, Longitude, Wind.kt
##
## 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.
## New names:
## * id -> id...1
## * season -> season...2
## * nature -> nature...3
## * time -> time...4
## * latitude -> latitude...5
## * ...
##
```

```
## New names:
## * id -> id...1
## * wkt_cur -> wkt_cur...2
## * id -> id...10
## * wkt_cur -> wkt_cur...11

## Warning: `add_rownames()` was deprecated in dplyr 1.0.0.
## Please use `tibble::rownames_to_column()` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was generated.

## Joining, by = "id"

load("generated_data/B_final.RData")
load("generated_data/mu_final.RData")
```

LMs for all hurricanes

```
new_train <- train %>% group_by(i) %>% mutate(new_i = cur_group_id())
new_test <- test %>% group_by(i) %>% mutate(new_i = cur_group_id())
new_full_dat <- dt %>% group_by(i) %>% mutate(new_i = cur_group_id())

mu_ws_coefs <- lm(wkt_new ~ wkt_cur + d_lat + d_log + d_wkt, data = new_full_dat) %>% coef()

examined_dt <- new_full_dat %>% filter(new_i == 602)

all_coefs <- rep(NA, 6)

for (j in 1:max(new_full_dat$new_i)) {
  cur_hur <- new_full_dat %>% filter(new_i == j)

  cur_lm <- lm(wkt_new ~ wkt_cur + d_lat + d_log + d_wkt, data = cur_hur)
  cur_lm_coefs <- c(cur_hur$new_i %>% mean(), cur_lm %>% coef())

  all_coefs <- rbind(all_coefs, cur_lm_coefs)
}

B_ws_coefs <- all_coefs[-1,]

par(mfrow = c(2, 5))
B_ws_coefs[,2] %>% hist(breaks = 100, main = "LM Beta0")
abline(v = mu_ws_coefs[1], col = "red")
B_ws_coefs[,3] %>% hist(breaks = 100, main = "LM Beta1")
abline(v = mu_ws_coefs[2], col = "red")
B_ws_coefs[,4] %>% hist(breaks = 100, main = "LM Beta2")
abline(v = mu_ws_coefs[3], col = "red")
B_ws_coefs[,5] %>% hist(breaks = 100, main = "LM Beta3")
abline(v = mu_ws_coefs[4], col = "red")
B_ws_coefs[,6] %>% hist(breaks = 100, main = "LM Beta4")
abline(v = mu_ws_coefs[5], col = "red")

B_final[,1] %>% hist(breaks = 100, main = "MCMC Beta0")
abline(v = mu_final[1], col = "red")
B_final[,2] %>% hist(breaks = 100, main = "MCMC Beta1")
```

```

abline(v = mu_final[2], col = "red")
B_final[,3] %>% hist(breaks = 100, main = "MCMC Beta2")
abline(v = mu_final[3], col = "red")
B_final[,4] %>% hist(breaks = 100, main = "MCMC Beta3")
abline(v = mu_final[4], col = "red")
B_final[,5] %>% hist(breaks = 100, main = "MCMC Beta4")
abline(v = mu_final[5], col = "red")

```

