Tidying Linear Models for Company Facilities

Tyler Revering

This code is based off of client work that I did for a company that offers services at various facilities and locations. They wanted to have a linear model for each one of their facilities based on how many visitors they brought in per week. Using tidy and nest functions, I aimed to create linear model objects based off of each facilities weekly visitors

```
# Loading the required libraries
library(tidyr)
library(purrr)
library(broom)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(nestr)
library(tidyverse)
```

```
## ── Attaching packages
## ─────────────────────────────────────── tidyverse 1.3.2 ──

## ✓ ggplot2 3.4.0     ✓ stringr 1.5.0
## ✓ tibble  3.1.8     ✓ forcats 0.5.2
## ✓ readr   2.1.3
```

```
## ── Conflicts ──────────────────────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
```

```
library(scales)
```

```
##
## Attaching package: 'scales'
##
## The following object is masked from 'package:readr':
##
##     col_factor
##
## The following object is masked from 'package:purrr':
##
##     discard
```

Off of data that was engineered prior to this analysis, there is data available for this physical therapy company's individual facilities and how many visitors they brought in per week.

```
# Loading the required data
load("TrendAnalysis.RData")

# For privacy purposes, not selecting the company and name of facility
facilitiesweek %>%
  select(WeekEndDate, TotalPatientVisits)
```

```
## # A tibble: 16,541 × 2
##    WeekEndDate TotalPatientVisits
##    <date>                   <int>
##  1 2020-01-04                  19
##  2 2020-01-11                  44
##  3 2020-01-18                  41
##  4 2020-01-25                  40
##  5 2020-02-01                  47
##  6 2020-02-08                  39
##  7 2020-02-15                  46
##  8 2020-02-22                  42
##  9 2020-02-29                  50
```

```
## 10 2020-03-07                  54
## # … with 16,531 more rows
```

Using the week dates, and total patient visits, I aimed to make linear model objects based off of each facilities weekly visitors.

```
# Using the tidyr and nest packages to make objects of linear models
facilitiesweek_lm <- facilitiesweek %>%
  nest(.by = c(CompanyName, FacilityName)) %>%
  mutate(model=map(data, ~lm(TotalPatientVisits~WeekEndDate, data = .)),
         tidied = map(model, tidy))

# Rows of linear model and tidied model objects
facilitiesweek_lm %>%
  select(model, tidied)
```

```
## # A tibble: 137 × 2
##    model  tidied
##    <list> <list>
##  1 <lm>   <tibble [2 × 5]>
##  2 <lm>   <tibble [2 × 5]>
##  3 <lm>   <tibble [2 × 5]>
##  4 <lm>   <tibble [2 × 5]>
##  5 <lm>   <tibble [2 × 5]>
##  6 <lm>   <tibble [2 × 5]>
##  7 <lm>   <tibble [2 × 5]>
##  8 <lm>   <tibble [2 × 5]>
##  9 <lm>   <tibble [2 × 5]>
## 10 <lm>   <tibble [2 × 5]>
## # … with 127 more rows
```

To analyze the trends of some, we can analyze one of the linear models, which will portray the trends of their weekly visitors depending on the date, and their estimate and statistical significance.

```
# Using unnesting of the tidied objects which contains coefficients and intercepts of
slopes <- facilitiesweek_lm %>%
  unnest(tidied) %>%
  filter(term=="WeekEndDate") %>%
  mutate(p.adjusted = p.adjust(p.value)) %>%
  filter(p.adjusted < 0.5) %>% # Filtering for statistically significant estimates.
  select(estimate, p.adjusted)
```

Now to view the results of the varying facilities, we will view our model's results, still without facility names, by descending estimates

```
slopes %>%
  arrange(desc(estimate))
```

```
## # A tibble: 108 × 2
##    estimate p.adjusted
##       <dbl>      <dbl>
##  1    1.09    4.13e- 1
##  2    0.861   3.42e- 1
##  3    0.853   1.22e- 1
##  4    0.735   6.98e- 8
##  5    0.623   1.18e- 1
##  6    0.568   8.32e- 2
##  7    0.419   1.85e- 1
##  8    0.414   2.57e-14
##  9    0.337   1.18e-11
## 10    0.325   8.53e- 7
## # … with 98 more rows
```

Considering that the client is aware of the facility names as they relate to the linear model objects, they can see the trends of weekly visitors, how their visitors change week on week throughout longer time horizons, and how statistically significant the models are.

Future steps as it pertains to this project is using the model objects to forecast the visitors that each facility might bring in for the next year