Regression Analysis

Haya Naviwala

5/12/2022

R Markdown

```
library(tidyverse)
```

```
## Registered S3 methods overwritten by 'tibble':
##   method     from
##   format.tbl pillar
##   print.tbl  pillar
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.5     v purrr   0.3.3
## v tibble  2.1.3     v dplyr   1.0.7
## v tidyr   1.0.0     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.4.0
```

```
## Warning: package 'ggplot2' was built under R version 3.6.2
```

```
## Warning: package 'dplyr' was built under R version 3.6.2
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(readr)

dt <- read_csv("AICPA_regressionAnalysisData (1).csv")
```

```
## Error: 'AICPA_regressionAnalysisData (1).csv' does not exist in current working directory ('/Users/ha
```

```
dt %>% glimpse()
```

```
## function (x, df, ncp, log = FALSE)
```

```
dt %>% head(2)
```

```
##
## 1 function (x, df, ncp, log = FALSE)
## 2 {
```

```
dt %>% tail(2)
```

```
##
## 5     else .Call(C_dnt, x, df, ncp, log)
## 6 }
```

```
dt %>% is.na() %>% colSums()
```

```
## Warning in is.na(.): is.na() applied to non-(list or vector) of type
## 'closure'
```

```
## Error in colSums(.): 'x' must be an array of at least two dimensions
```

```r
#splitting data into training (70-80%) vs testing(30-20%)
dt %>%
  count(type)
```

```
## Error in UseMethod("count"): no applicable method for 'count' applied to an object of class "functio
```

```r
#creating new dataset
aicpaTrain <- dt %>%
  filter(type == "dt4training")
```

```
## Error in UseMethod("filter"): no applicable method for 'filter' applied to an object of class "functi
```

```r
aicpaTrain %>% head(2)
```

```
## Error in eval(lhs, parent, parent): object 'aicpaTrain' not found
```

```r
aicpaTrain %>% tail(2)
```

```
## Error in eval(lhs, parent, parent): object 'aicpaTrain' not found
```

```r
aicpaTrain %>%
  select(-type) %>%
  summary()
```

```
## Error in eval(lhs, parent, parent): object 'aicpaTrain' not found
```

```r
#testing data

aicpaTest <- dt %>%
  filter(type == "dt4testing")
```

```
## Error in UseMethod("filter"): no applicable method for 'filter' applied to an object of class "functi
```

```r
aicpaTest %>%
  select(-type) %>%
  summary()
```

```
## Error in eval(lhs, parent, parent): object 'aicpaTest' not found
```

```r
#visualizations
dt %>%
  ggplot(aes(x=date, y=revenue))+
  geom_line() + geom_point()
```

```
## Error:   You're passing a function as global data.
##   Have you misspelled the `data` argument in `ggplot()`
```

```r
aicpaTrain <- aicpaTrain %>%
mutate(byUse =
ifelse(heatDD > coolDD, "heating", "cooling"))
```

```
## Error in eval(lhs, parent, parent): object 'aicpaTrain' not found
```

```r
aicpaTrain %>%
  ggplot(aes(x = date, y = revenue)) +
  geom_point(aes(color = byUse)) +
  geom_line()
```

```
## Error in eval(lhs, parent, parent): object 'aicpaTrain' not found
```

```
aicpaTrain %>%
  ggplot(aes(x=byUse, y=revenue)) +
  geom_boxplot(aes(fill = byUse))
```

```
## Error in eval(lhs, parent, parent): object 'aicpaTrain' not found
```

```
aicpaTrain %>%
  ggplot(aes(x = production, y = revenue)) +
  geom_point(aes(color = byUse))
```

```
## Error in eval(lhs, parent, parent): object 'aicpaTrain' not found
```

```
aicpaTrain %>%
  select(-type, -date, -byUse) %>%
  cor()
```

```
## Error in eval(lhs, parent, parent): object 'aicpaTrain' not found
```

```
options(scipen = 99)
lm1 <- lm(revenue~production, aicpaTrain)
```

```
## Error in is.data.frame(data): object 'aicpaTrain' not found
```

```
summary(lm1)
```

```
## Error in summary(lm1): object 'lm1' not found
```

```
aicpaTest %>% head(1)
```

```
## Error in eval(lhs, parent, parent): object 'aicpaTest' not found
```

```
pred_lm1 <- predict(lm1, newdata = aicpaTest)
```

```
## Error in predict(lm1, newdata = aicpaTest): object 'lm1' not found
```

```
pred_lm1
```

```
## Error in eval(expr, envir, enclos): object 'pred_lm1' not found
```

```
dt_err_lm1 <- aicpaTest %>%
  select(date, production, revenue) %>%
  mutate(pred_lm1) %>%
  mutate(pctErr = 100 * (revenue - pred_lm1)/revenue)
```

```
## Error in eval(lhs, parent, parent): object 'aicpaTest' not found
```

```
dt_err_lm1
```

```
## Error in eval(expr, envir, enclos): object 'dt_err_lm1' not found
```

```
mape_lm1 <- dt_err_lm1 %>%
  mutate(absErr = abs(pctErr)) %>%
  pull(absErr) %>%
  mean()
```

```
## Error in eval(lhs, parent, parent): object 'dt_err_lm1' not found
```

```
mape_lm1
```

```
## Error in eval(expr, envir, enclos): object 'mape_lm1' not found
```

```
dt_err_lm1 %>%
  select(date, revenue, pred_lm1) %>%
```

```
  gather("variable", "value", -date) %>%
  ggplot(aes(x=date))+
  geom_point(aes(y=value, color = variable))+
  geom_line(aes(y=value, color = variable))+
  xlab("Time")+ylab("Revenue")
```

```
## Error in eval(lhs, parent, parent): object 'dt_err_lm1' not found
```

```
xLm1 <- lm(revenue~production+heatDD, aicpaTrain)
```

```
## Error in is.data.frame(data): object 'aicpaTrain' not found
```

```
summary(xLm1)
```

```
## Error in summary(xLm1): object 'xLm1' not found
```

```
#Call:
  lm(formula = revenue ~ production + heatDD, data = aicpaTrain)
```

```
## Error in is.data.frame(data): object 'aicpaTrain' not found
```

```
pred_xLm1 <- predict(xLm1, newdata = aicpaTest)
```

```
## Error in predict(xLm1, newdata = aicpaTest): object 'xLm1' not found
```

```
dt_err_xLm1 <- aicpaTest %>%
  select(date, revenue) %>%
  mutate(pred_xLm1) %>%
  mutate(pctErr = 100 * (revenue - pred_xLm1)/revenue)
```

```
## Error in eval(lhs, parent, parent): object 'aicpaTest' not found
```

```
dt_err_xLm1
```

```
## Error in eval(expr, envir, enclos): object 'dt_err_xLm1' not found
```

```
mape_xLm1 <- dt_err_xLm1 %>%
  mutate(absErr = abs(pctErr)) %>%
  pull(absErr) %>%
  mean()
```

```
## Error in eval(lhs, parent, parent): object 'dt_err_xLm1' not found
```

```
mape_xLm1
```

```
## Error in eval(expr, envir, enclos): object 'mape_xLm1' not found
```

```
dt_err_xLm1 %>%
  select(date, revenue, pred_xLm1) %>%
  gather("variable", "value", -date) %>%
  ggplot(aes(x=date))+
  geom_point(aes(y=value, color = variable))+
  geom_line(aes(y=value, color = variable))+
  scale_x_date(date_breaks= '2 month' ) +
  xlab("Time")+ylab("Revenue")
```

```
## Error in eval(lhs, parent, parent): object 'dt_err_xLm1' not found
```

```
library(scales)
```

```
##
```

```
## Attaching package: 'scales'
```

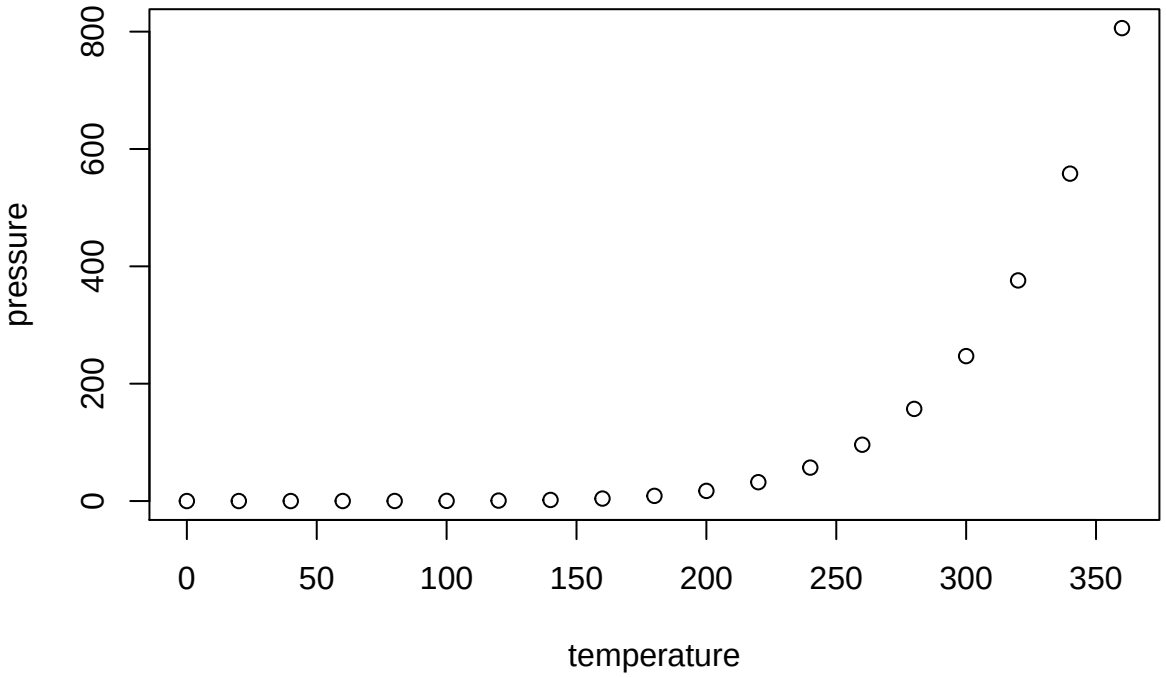
```
## The following object is masked from 'package:purrr':
##
##     discard
```

```
## The following object is masked from 'package:readr':
##
##     col_factor
```

```
dt %>%
  ggplot(aes(x=date, y=revenue))+
  geom_line() +
  geom_point() +
  scale_x_date(date_breaks= '1 year',
    labels = date_format(format =  '%b-%Y')) +
    scale_y_continuous(labels = unit_format(scale = 1e-6,
                                            unit =  MM ),
                       breaks = extended_breaks(n = 7),
                       limits = c(0, 30*1e6)) +
xlab("Time") + ylab("Revenue")
```

```
## Error:   You're passing a function as global data.
##   Have you misspelled the `data` argument in `ggplot()`
```

Including Plots



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.