Regression: In Class Work

Garcia

```r
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.0.5

## -- Attaching packages --------------------------------------- tidyverse
1.3.1 --

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.4     v dplyr   1.0.7
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   2.0.1     v forcats 0.5.1

## Warning: package 'ggplot2' was built under R version 4.0.5

## Warning: package 'tibble' was built under R version 4.0.5

## Warning: package 'tidyr' was built under R version 4.0.5

## Warning: package 'readr' was built under R version 4.0.5

## Warning: package 'purrr' was built under R version 4.0.5

## Warning: package 'dplyr' was built under R version 4.0.5

## Warning: package 'stringr' was built under R version 4.0.5

## Warning: package 'forcats' was built under R version 4.0.5

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

```r
library(tidymodels)
```

```
## Warning: package 'tidymodels' was built under R version 4.0.5

## Registered S3 method overwritten by 'tune':
##   method                   from
##   required_pkgs.model_spec parsnip

## -- Attaching packages -------------------------------------- tidymodels
0.1.3 --

## v broom        0.7.9      v rsample      0.1.0
## v dials        0.0.10     v tune         0.1.6
## v infer        1.0.0      v workflows    0.2.3
## v modeldata    0.1.1      v workflowsets 0.1.0
## v parsnip      0.1.7      v yardstick    0.0.8
## v recipes      0.1.17

## Warning: package 'broom' was built under R version 4.0.5

## Warning: package 'dials' was built under R version 4.0.5

## Warning: package 'scales' was built under R version 4.0.5

## Warning: package 'infer' was built under R version 4.0.5

## Warning: package 'modeldata' was built under R version 4.0.5

## Warning: package 'parsnip' was built under R version 4.0.5

## Warning: package 'rsample' was built under R version 4.0.5

## Warning: package 'tune' was built under R version 4.0.5

## Warning: package 'workflows' was built under R version 4.0.5

## Warning: package 'workflowsets' was built under R version 4.0.5

## Warning: package 'yardstick' was built under R version 4.0.5

## -- Conflicts -----------------------------------------
tidymodels_conflicts() --
## x scales::discard() masks purrr::discard()
## x dplyr::filter()   masks stats::filter()
## x recipes::fixed()  masks stringr::fixed()
## x dplyr::lag()      masks stats::lag()
## x yardstick::spec() masks readr::spec()
## x recipes::step()   masks stats::step()
## * Use tidymodels_prefer() to resolve common conflicts.
```

```r
library(plotly)
```

```
## Warning: package 'plotly' was built under R version 4.0.5

##
## Attaching package: 'plotly'

## The following object is masked from 'package:ggplot2':
##
##     last_plot

## The following object is masked from 'package:stats':
##
##     filter

## The following object is masked from 'package:graphics':
##
##     layout
```

```r
ad<-read_rds("area_data.Rds")
```

1. Estimate a model that includes the census division (`division`) of the area as the sole independent variable, and mobility `perc_moved  in` as the dependent variable. Provide an interpretation of the results. "perc_moved_in~division"

```r
set.seed(35202)

split_data<-ad%>%initial_split(prop=.5)

ad_train<-training(split_data)

ad_test<-testing(split_data)


lm_fit <-
  linear_reg() %>%
  set_engine("lm")%>%
  set_mode("regression")


move_wf<-workflow()%>%
  add_model(lm_fit)


move_formula<-as.formula("perc_moved_in~division")


move_rec<-recipe(move_formula,data=ad)%>%
  step_dummy(division)

move_wf<-move_wf%>%
  add_recipe(move_rec)


lm_results<-fit(move_wf,ad_train)

lm_results%>%
  tidy()

## # A tibble: 9 x 5
##   term                          estimate std.error statistic  p.value
##   <chr>                            <dbl>     <dbl>     <dbl>    <dbl>
## 1 (Intercept)                      1.62      0.188     8.64  9.47e-17
## 2 division_West.North.Central      1.44      0.280     5.16  3.77e- 7
## 3 division_Mid.Atlantic            0.219     0.354     0.620 5.35e- 1
## 4 division_New.England             1.16      0.576     2.02  4.45e- 2
## 5 division_East.South.Central      0.724     0.314     2.31  2.15e- 2
## 6 division_South.Atlantic          1.16      0.272     4.28  2.27e- 5
## 7 division_West.South.Central    0.570     0.280     2.04  4.23e- 2
## 8 division_Mountain               2.39      0.320     7.47  4.15e-13
## 9 division_Pacific                1.37      0.366     3.75  2.03e- 4
```

```
lm_results%>%
  pull_workflow_fit()%>%
  glance()
```

```
## Warning: `pull_workflow_fit()` was deprecated in workflows 0.2.3.
## Please use `extract_fit_parsnip()` instead.
```

```
## # A tibble: 1 x 12
##   r.squared adj.r.squared sigma statistic  p.value    df logLik   AIC
BIC
##       <dbl>         <dbl> <dbl>     <dbl>    <dbl> <dbl>  <dbl> <dbl>
<dbl>
## 1     0.142         0.127  1.72      9.41 4.68e-12     8  -904. 1828.
1869.
## # ... with 3 more variables: deviance <dbl>, df.residual <int>, nobs <int>
```

```
ad_test<-
  predict(lm_results,ad_test)%>%
  rename(pred1=.pred)%>%
  bind_cols(ad_test)

rmse_1<-ad_test%>%rmse(truth=perc_moved_in,estimate=pred1)
rmse_1
```

```
## # A tibble: 1 x 3
##   .metric .estimator .estimate
##   <chr>   <chr>          <dbl>
## 1 rmse    standard        1.37
```

2. Add both income (income_75) and commute times (perc_commute_30p) to the above model and describe the coefficients for all three of the variables.

```
move_formula<-as.formula("perc_moved_in~division+income_75+perc_commute_30p")



move_rec<-recipe(move_formula,data=ad)%>%
  step_dummy(division)

move_wf<-move_wf%>%
  update_recipe(move_rec)



lm_results<-fit(move_wf,ad_train)

lm_results%>%
  tidy()
```

```
## # A tibble: 11 x 5
##    term                          estimate std.error statistic  p.value
##    <chr>                            <dbl>     <dbl>     <dbl>    <dbl>
##  1 (Intercept)                     1.58     0.491       3.22  1.38e- 3
##  2 income_75                       0.0362   0.0109      3.31  9.92e- 4
##  3 perc_commute_30p               -0.0368   0.00916    -4.01  7.02e- 5
##  4 division_West.North.Central     1.12     0.281       3.99  7.80e- 5
##  5 division_Mid.Atlantic           0.201    0.348       0.577 5.64e- 1
##  6 division_New.England            0.851    0.585       1.46  1.46e- 1
##  7 division_East.South.Central     1.11     0.316       3.52  4.71e- 4
##  8 division_South.Atlantic         1.43     0.271       5.28  2.06e- 7
##  9 division_West.South.Central     0.734    0.275       2.67  7.91e- 3
## 10 division_Mountain               2.29     0.314       7.31  1.26e-12
## 11 division_Pacific                1.30     0.366       3.56  4.16e- 4
```

```
lm_results%>%
  pull_workflow_fit()%>%
  glance()
```

```
## Warning: `pull_workflow_fit()` was deprecated in workflows 0.2.3.
## Please use `extract_fit_parsnip()` instead.
```

```
## # A tibble: 1 x 12
##   r.squared adj.r.squared sigma statistic  p.value    df logLik   AIC
BIC
##       <dbl>         <dbl> <dbl>     <dbl>    <dbl> <dbl>  <dbl> <dbl>
<dbl>
## 1     0.188         0.170  1.68      10.4 5.98e-16    10  -891. 1807.
1857.
## # ... with 3 more variables: deviance <dbl>, df.residual <int>, nobs <int>
```

```
ad_test<-
  predict(lm_results,ad_test)%>%
  #PLEASE NOTE: new models being fit need new pred# names, so your first is
pred1, the second is pred2, the third is pred3, etc.
  rename(pred2=.pred)%>%
  bind_cols(ad_test)

rmse_2<-ad_test%>%rmse(truth=perc_moved_in,estimate=pred2)
rmse_2
```

```
## # A tibble: 1 x 3
##   .metric .estimator .estimate
##   <chr>   <chr>          <dbl>
## 1 rmse    standard        1.30
```

3. Which of the two models above fit the data better? How do you know?

```
#This is a way of comparing the models in the same table.
rmse_comp<-rbind(rmse_1,rmse_2)
rmse_comp
```

```
## # A tibble: 2 x 3
##   .metric .estimator .estimate
##   <chr>   <chr>          <dbl>
## 1 rmse    standard        1.37
## 2 rmse    standard        1.30
```