AE 11: MLR Inference + conditions
Trail riders

Oct 26, 2022

> ❗ **Important**
>
> The AE is due on GitHub by Saturday, October 29 at 11:59pm.

Packages

```r
library(tidyverse)
library(tidymodels)
library(knitr)
```

Data

```r
rail_trail <- read_csv("data/rail_trail.csv")
```

Exercise 1

Below is the model predicting `volume` from `hightemp` and `season`.

```r
rt_mlr_main_fit <- linear_reg() %>%
  set_engine("lm") %>%
  fit(volume ~ hightemp + season, data = rail_trail)

tidy(rt_mlr_main_fit) |>
  kable(digits = 2)
```

term	estimate	std.error	statistic	p.value
(Intercept)	-125.23	71.66	-1.75	0.08
hightemp	7.54	1.17	6.43	0.00
seasonSpring	5.13	34.32	0.15	0.88
seasonSummer	-76.84	47.71	-1.61	0.11

Add an interaction effect between `hightemp` and `season` to the model. Do the data provide evidence of a significant interaction effect? Comment on the significance of the interaction terms.

```
interact <- linear_reg() %>%
  set_engine("lm") %>%
  fit(volume ~ hightemp + season + hightemp*season, data = rail_trail)

tidy(interact) %>%
  kable(digits = 3)
```

term	estimate	std.error	statistic	p.value
(Intercept)	-10.530	166.796	-0.063	0.950
hightemp	5.482	2.950	1.858	0.067
seasonSpring	-293.949	190.331	-1.544	0.126
seasonSummer	354.179	255.078	1.389	0.169
hightemp:seasonSpring	4.884	3.262	1.497	0.138
hightemp:seasonSummer	-4.537	3.754	-1.209	0.230

These data, surprisingly, do not provide evidence of a significant interaction effect. The p-value of hightemp * Spring and Summer are .138 and .230, respectively.

```
glance(rt_mlr_main_fit) %>%
  select(adj.r.squared, AIC, BIC)
```

```
# A tibble: 1 x 3
  adj.r.squared   AIC   BIC
          <dbl> <dbl> <dbl>
1         0.365 1093. 1106.
```

```
glance(interact) %>%
  select(adj.r.squared, AIC, BIC)
```

```
# A tibble: 1 x 3
  adj.r.squared   AIC   BIC
          <dbl> <dbl> <dbl>
1         0.435 1084. 1102.
```

Exercise 2

Below is the model predicting `volume` from all available predictors.

```r
rt_full_fit <- linear_reg() %>%
  set_engine("lm") %>%
  fit(volume ~ ., data = rail_trail)

tidy(rt_full_fit) |>
  kable(digits = 2)
```

term	estimate	std.error	statistic	p.value
(Intercept)	17.62	76.58	0.23	0.82
hightemp	7.07	2.42	2.92	0.00
avgtemp	-2.04	3.14	-0.65	0.52
seasonSpring	35.91	32.99	1.09	0.28
seasonSummer	24.15	52.81	0.46	0.65
cloudcover	-7.25	3.84	-1.89	0.06
precip	-95.70	42.57	-2.25	0.03
day_typeWeekend	35.90	22.43	1.60	0.11

Fill in the code to plot the histogram of residuals with an overlay of the normal distribution based on the results of the model.
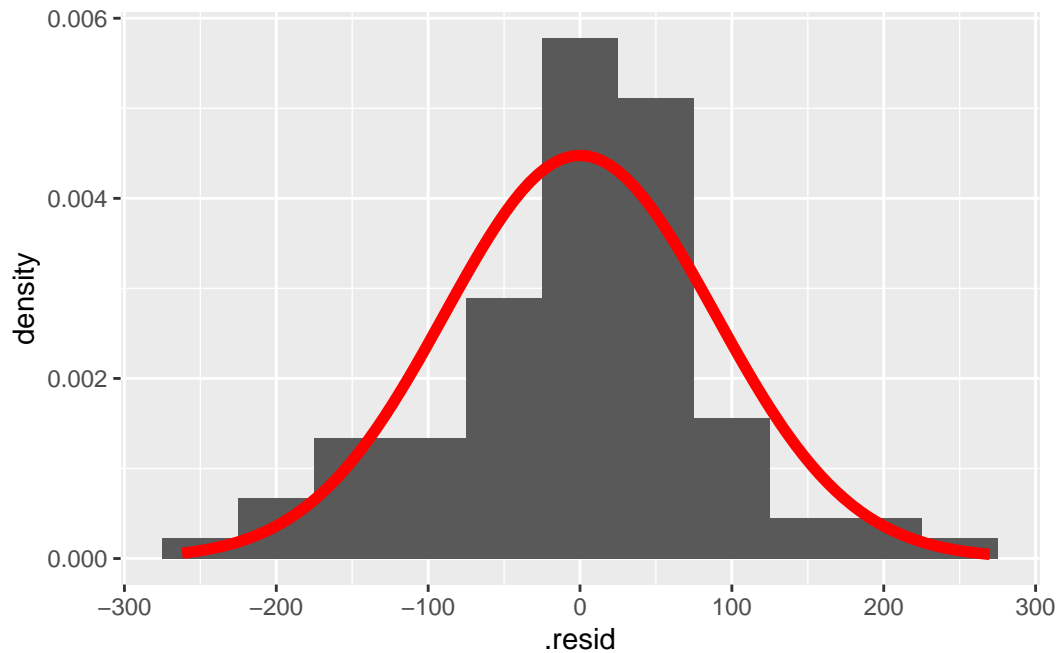
> **ℹ Note**
>
> Update to `eval: true` once the code is updated.

```r
rt_full_aug <- augment(rt_full_fit$fit)

ggplot(rt_full_aug, aes(.resid)) +
  geom_histogram(aes(y = after_stat(density)), binwidth = 50) +
  stat_function(
    fun = dnorm,
    args = list(mean = mean(rt_full_aug$.resid), sd = sd(rt_full_aug$.resid)),
```

```
    lwd = 2,
    color = "red"
  )
```


Submission

> **!** Important
>
> To submit the AE:
>
> - Render the document to produce the PDF with all of your work from today's class.
> - Push all your work to your `ae-11-` repo on GitHub. (You do not submit AEs on Gradescope).