Course evaluations and linear regression

Kaitlyn Group 9

M6 ICA2

Introduction

This is adapted from Lab 6 in Duke's Introduction to Data Science course.

We will analyze what goes into course evaluations and how certain variables effect the overall score.

To get started, load packages `tidyverse` and `broom`. Install any packages with code `install.packages("package_name")`.

```
library(tidyverse)
library(broom)
```

Data

The data were gathered from end of semester student evaluations for a large sample of professors from the University of Texas at Austin. In addition, six students rated the professors' physical appearance. Each row in `evals` contains a different course and the columns represent variables about the courses and professors.

Use `read_csv()` to read in the data and save it as an object named `evals`. The data is available on D2L.

```
evals <- read_csv("evals-mod.csv")
```

Data dictionary

Variable	**Description**
score	Average professor evaluation score: (1) very unsatisfactory - (5) excellent
rank	Rank of professor: teaching, tenure track, tenure
ethnicity	Ethnicity of professor: not minority, minority
gender	Gender of professor: female, male
language	Language of school where professor received education: english or non-english
age	Age of professor
cls_perc_eval	Percent of students in class who completed evaluation
cls_did_eval	Number of students in class who completed evaluation
cls_students	Total number of students in class
cls_level	Class level: lower, upper
cls_profs	Number of professors teaching sections in course in sample: single, multiple
cls_credits	Number of credits of class: one credit (lab, PE, etc.), multi credit
bty_f1lower	Beauty rating of professor from lower level female: (1) lowest - (10) highest
bty_f1upper	Beauty rating of professor from upper level female: (1) lowest - (10) highest

Variable	Description
bty_f2upper	Beauty rating of professor from upper level female: (1) lowest - (10) highest
bty_m1lower	Beauty rating of professor from lower level male: (1) lowest - (10) highest
bty_m1upper	Beauty rating of professor from upper level male: (1) lowest - (10) highest
bty_m2upper	Beauty rating of professor from upper level male: (1) lowest - (10) highest

Before you get started, add the `avg_bty` variable.

```
evals <- evals %>%
  rowwise() %>%
  mutate(avg_bty = mean(bty_f1lower:bty_m2upper)) %>%
  ungroup()
```

Part 1

Categorical predictors

Task 1 Fit a linear model with `score` as the response and `language` as a single predictor. Write out the model output.

```
m.evals<- lm(score~factor(language), data=evals)
m.evals
```

```
Call:
lm(formula = score ~ factor(language), data = evals)

Coefficients:
                (Intercept)  factor(language)non-english
                     4.1897                      -0.2468
```

score= 4.1897- 0.2468 x language

Task 2 What is the baseline level in Task 1? Interpret the meaning of coefficient b_1.

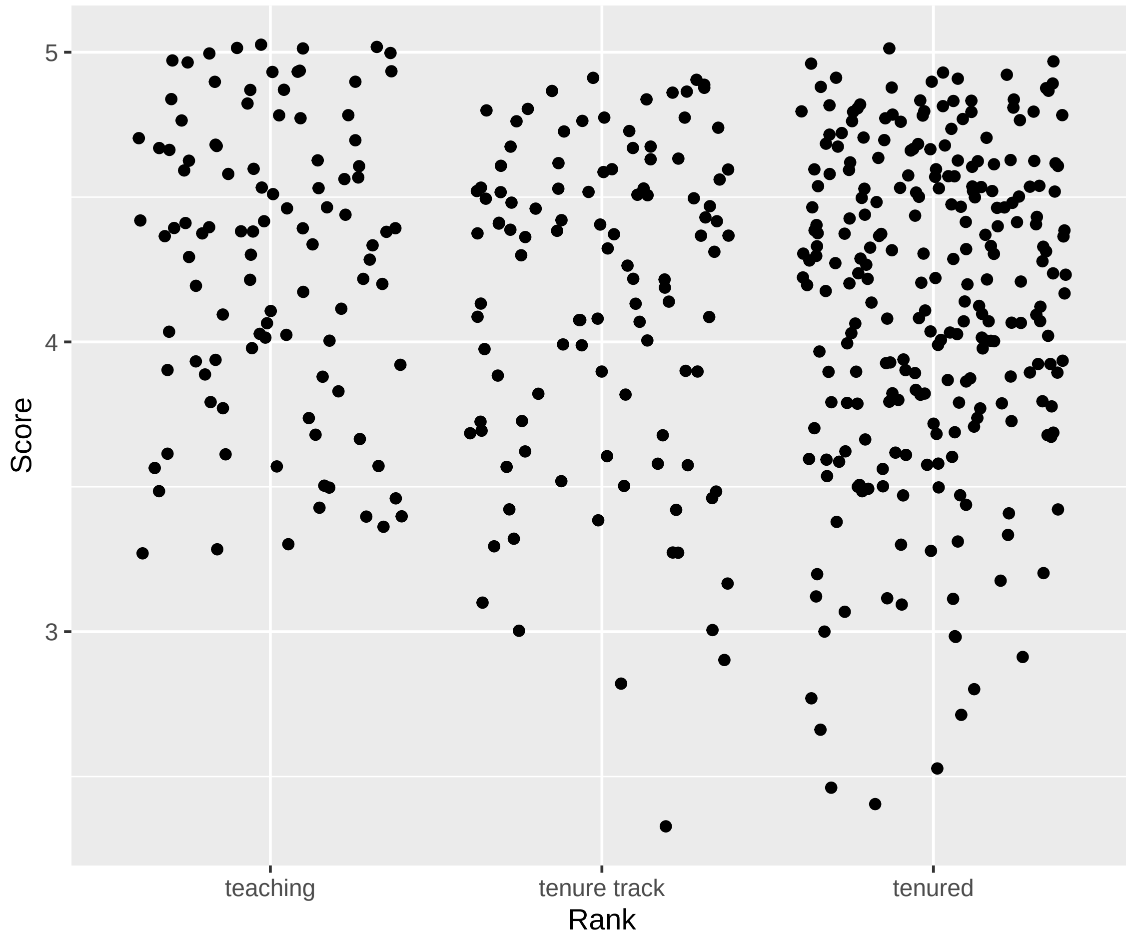
The baseline level is 4.1897 represented in the equation above represents if English language was spoken. #### Task 3

Based on Task 1, what is the equation of the line for English speaking professors? What about non-English speaking professors?

English Speaking Professor: score= 4.1897 Non-English Speaking Professor: score= 4.1897- 0.2468 x language

Task 4 Create a scatter plot of `score` versus `rank` with `ggplot()`. Use `geom_jitter()`.

```
evals %>%
  ggplot(mapping=aes(x=rank, y=score))+
  geom_jitter()+
  labs(x="Rank", y="Score")
```



Task 5 Fit a linear model with `score` as the response and `rank` as a single predictor. What is the baseline? Write out the model output.

```
m2.evals<- lm(score~factor(rank), data=evals)
m2.evals
```

```
Call:
lm(formula = score ~ factor(rank), data = evals)

Coefficients:
             (Intercept)  factor(rank)tenure track        factor(rank)tenured
                  4.2843                   -0.1297                    -0.1452
```

The baseline is 4.2843 if the rank is teaching. #### Task 6

Add a new variable to `evals` called `rank_new` where the baseline level is set to "tenured". *Hint*: `relevel()`

```
evals <- evals %>%
  mutate(rank_new = relevel(factor(rank), ref="tenured"))
glimpse(evals)
```

```
Rows: 463
```

```
Columns: 20
$ score         <dbl> 4.7, 4.1, 3.9, 4.8, 4.6, 4.3, 2.8, 4.1, 3.4, 4.5, 3.8, 4~
$ rank          <chr> "tenure track", "tenure track", "tenure track", "tenure ~
$ ethnicity     <chr> "minority", "minority", "minority", "minority", "not min~
$ gender        <chr> "female", "female", "female", "female", "male", "male", ~
$ language      <chr> "english", "english", "english", "english", "english", "~
$ age           <dbl> 36, 36, 36, 36, 59, 59, 59, 51, 51, 40, 40, 40, 40, 40, ~
$ cls_perc_eval <dbl> 55.81395, 68.80000, 60.80000, 62.60163, 85.00000, 87.500~
$ cls_did_eval  <dbl> 24, 86, 76, 77, 17, 35, 39, 55, 111, 40, 24, 24, 17, 14,~
$ cls_students  <dbl> 43, 125, 125, 123, 20, 40, 44, 55, 195, 46, 27, 25, 20, ~
$ cls_level     <chr> "upper", "upper", "upper", "upper", "upper", "upper", "u~
$ cls_profs     <chr> "single", "single", "single", "single", "multiple", "mul~
$ cls_credits   <chr> "multi credit", "multi credit", "multi credit", "multi c~
$ bty_f1lower   <dbl> 5, 5, 5, 5, 4, 4, 4, 5, 5, 2, 2, 2, 2, 2, 2, 2, 2, 7, 7,~
$ bty_f1upper   <dbl> 7, 7, 7, 7, 4, 4, 4, 2, 2, 5, 5, 5, 5, 5, 5, 5, 5, 9, 9,~
$ bty_f2upper   <dbl> 6, 6, 6, 6, 2, 2, 2, 5, 5, 4, 4, 4, 4, 4, 4, 4, 4, 9, 9,~
$ bty_m1lower   <dbl> 2, 2, 2, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 3, 3, 3, 7, 7,~
$ bty_m1upper   <dbl> 4, 4, 4, 4, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 6, 6,~
$ bty_m2upper   <dbl> 6, 6, 6, 6, 3, 3, 3, 3, 3, 2, 2, 2, 2, 2, 2, 2, 2, 6, 6,~
$ avg_bty       <dbl> 5.5, 5.5, 5.5, 5.5, 3.5, 3.5, 3.5, 4.0, 4.0, 2.0, 2.0, 2~
$ rank_new      <fct> tenure track, tenure track, tenure track, tenure track, ~
```

Task 7 Fit a linear model with `score` as the response and `rank_new` as a single predictor. Is the baseline now different from the baseline in Task 5?

```
m3.evals<- lm(score~factor(rank_new), data=evals)
m3.evals
```

```

Call:
lm(formula = score ~ factor(rank_new), data = evals)

Coefficients:
                 (Intercept)      factor(rank_new)teaching
                      4.1391                        0.1452
factor(rank_new)tenure track
                      0.0155
```

The baseline is slightly different than the one above. It is now 4.1391 for a tenured professor.

Part 2

Multiple regression

Task 8 Fit a linear model with `score` as the response and `gender`, `rank`, and `avg_bty` as predictors. Write out the model. Give an interpretation for the coefficient of `avg_bty`.

```
m4.evals<- lm(score~factor(rank) + factor(gender)+ avg_bty, data=evals)
m4.evals %>%
tidy()
```

```
# A tibble: 5 x 5
  term                     estimate std.error statistic   p.value
  <chr>                       <dbl>     <dbl>     <dbl>     <dbl>
1 (Intercept)                3.96      0.0957     41.4  7.21e-157
2 factor(rank)tenure track  -0.0940    0.0739     -1.27 2.04e-  1
3 factor(rank)tenured       -0.156     0.0634     -2.47 1.40e-  2
4 factor(gender)male         0.171     0.0522      3.28 1.13e-  3
5 avg_bty                    0.0509    0.0161      3.15 1.73e-  3
```

score= 3.96 -0.094 x rank(tenure track)- 0.16 x rank(tenured) + 0.17 x gender(male) + 0.051 x avg_bty

The coefficient 0.051 represents the difference from baseline in regards to professor's beauty on any level.

Task 9 What are the R^2 and adjusted R^2 values from your model in Task 8?

```
m4.evals %>%
 glance() %>%
 select(r.squared, adj.r.squared)
```

```
# A tibble: 1 x 2
  r.squared adj.r.squared
      <dbl>         <dbl>
1    0.0534        0.0452
```

Task 10 Fit a linear model with `score` as the response and only `gender` and `avg_bty` as predictors. How did the R^2 and adjusted R^2 values change compared to Task 9?

```
m5.evals<- lm(score~factor(gender)+ avg_bty, data=evals)
m5.evals %>%
tidy()
```

```
# A tibble: 3 x 5
  term               estimate std.error statistic   p.value
  <chr>                 <dbl>     <dbl>     <dbl>     <dbl>
1 (Intercept)          3.85      0.0811     47.4  2.76e-179
2 factor(gender)male   0.149     0.0503      2.96 3.22e-  3
3 avg_bty              0.0550    0.0161      3.41 7.05e-  4
```

```
m5.evals %>%
 glance() %>%
 select(r.squared, adj.r.squared)
```

```
# A tibble: 1 x 2
  r.squared adj.r.squared
      <dbl>         <dbl>
1    0.0408        0.0366
```

The new r squared and adjusted r squared values without rank as a predictor are about .1 lower than with rank included in the predictors.

Model Selection

Task 11 Fit a full model with `score` as the response and predictors: `rank`, `ethnicity`, `gender`, `language`, `age`, `cls_perc_eval`, `cls_students`, `cls_level`, `cls_profs`, `cls_credits`, `bty_avg`.

```
full.model <- lm(score ~ factor(rank) + factor(ethnicity) + factor(gender) +factor(language) + age+ cls_

step(object = full.model, direction = "backward", trace = FALSE)
```

```
Call:
lm(formula = score ~ factor(ethnicity) + factor(gender) + age +
    cls_perc_eval + cls_students + factor(cls_credits) + avg_bty,
    data = evals)

Coefficients:
                  (Intercept)  factor(ethnicity)not minority
                    3.4647978                      0.2242588
           factor(gender)male                            age
                    0.1651058                     -0.0067325
                cls_perc_eval                   cls_students
                    0.0062801                      0.0005732
factor(cls_credits)one credit                        avg_bty
                    0.5376738                      0.0496457
```

Task 12 Why did we not consider `cls_did_eval` and the individual beauty scores?

We did not include "cls_did_eval" and the individual beauty scores because they were too highly correlated with the other predictors being used. Therefore, they were ignored.

Task 13 Use the fitted full model in Task 11 and backward selection to determine the "best"" model. What are the R^2 and adjusted R^2 values from this "best" model?

```
m.best<-full.model %>%
  step(object = full.model, direction = "backward", trace = FALSE)
m.best
```

```
Call:
lm(formula = score ~ factor(ethnicity) + factor(gender) + age +
    cls_perc_eval + cls_students + factor(cls_credits) + avg_bty,
    data = evals)

Coefficients:
                  (Intercept)  factor(ethnicity)not minority
                    3.4647978                      0.2242588
           factor(gender)male                            age
                    0.1651058                     -0.0067325
                cls_perc_eval                   cls_students
                    0.0062801                      0.0005732
factor(cls_credits)one credit                        avg_bty
                    0.5376738                      0.0496457
```

```
m.best %>%
  glance() %>%
  select(r.squared, adj.r.squared)
```

```
# A tibble: 1 x 2
  r.squared adj.r.squared
      <dbl>         <dbl>
1     0.150         0.137
```

The r squared value for the best fit model is 0.1496 while the adjusted r squared value is 0.13653

Inference

Task 14 Create a 95% prediction interval based on new predictor values of your choosing. Use your "best" model from Task 13.

```
new.evals <- data.frame(age = c(75))
new.predict<-predict(m.best, new.data = new.evals, interval="prediction")
glimpse(new.predict)
```

```
 num [1:463, 1:3] 3.87 4 3.95 3.96 4.18 ...
 - attr(*, "dimnames")=List of 2
  ..$ : chr [1:463] "1" "2" "3" "4" ...
  ..$ : chr [1:3] "fit" "lwr" "upr"
```

Task 15 Create 95% confidence intervals for the coeffients of your "best"" model from Task 13.

```
confint(m.best)
```

```
                                    2.5 %       97.5 %
(Intercept)                  3.071728e+00  3.857867512
factor(ethnicity)not minority  8.386718e-02  0.364650454
factor(gender)male            6.596259e-02  0.264249072
age                          -1.172166e-02 -0.001743328
cls_perc_eval                 3.266998e-03  0.009293137
cls_students                 -9.484521e-05  0.001241185
factor(cls_credits)one credit  3.312716e-01  0.744075951
avg_bty                       1.904555e-02  0.080245852
```

Task 16 Can we use this model to make valid predictions about professors from any University?

Although these are factors of professors at majority of universities, the R squared values are extremely low indicating very little correlation. Therefore, I do not believe this model makes valid predictions for any university.

References

1. http://www2.stat.duke.edu/courses/Spring18/Sta199/labs/lab-06-modelling-course-evals.html