Lab 7: Multiple linear regression

Complete all **Exercises**, and submit answers to **Questions** on the Coursera platform. Note that the order of the choices in multiple choice questions may be different on the Duke Coursera platform than the order in this document.

Grading the professor

Many college courses conclude by giving students the opportunity to evaluate the course and the instructor anonymously. However, the use of these student evaluations as an indicator of course quality and teaching effectiveness is often criticized because these measures may reflect the influence of non-teaching related characteristics, such as the physical appearance of the instructor. The article titled, ``Beauty in the classroom: instructors' pulchritude and putative pedagogical productivity'' (Hamermesh and Parker, 2005) found that instructors who are viewed to be better looking receive higher instructional ratings.

Daniel S. Hamermesh, Amy Parker, Beauty in the classroom: instructors' pulchritude and putative pedagogical productivity, **Economics of Education Review**, Volume 24, Issue 4, August 2005, Pages 369-376, ISSN 0272-7757, 10.1016/j.econedurev.2004.07.013.
http://www.sciencedirect.com/science/article/pii/S0272775704001165
(http://www.sciencedirect.com/science/article/pii/S0272775704001165)

In this lab we will analyze the data from this study in order to learn what goes into a positive professor evaluation.

The 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.

This is a slightly modified version of the original data set that was released as part of the replication data for *Data Analysis Using Regression and Multilevel/Hierarchical Models* (Gelman and Hill, 2007).} The result is a data frame where each row contains a different course and columns represent variables about the courses and professors.

```
load(url("http://www.openintro.org/stat/data/evals.RData"))
```

- `score` - average professor evaluation score: (1) very unsatisfactory - (5) excellent.
- `rank` - rank of professor: teaching, tenure track, tenured.
- `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.
- `bty_f2upper` - beauty rating of professor from second 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 second upper level male: (1) lowest - (10) highest.
- `bty_avg` - average beauty rating of professor.
- `pic_outfit` - outfit of professor in picture: not formal, formal. (not used in this analysis)
- `pic_color` - color of professor's picture: color, black & white. (not used in this analysis)

Exploring the data

Question 1

These data come from an experiment.

A) True
B) False

Question 2

The original research question posed in the paper is whether beauty leads directly to the differences in course evaluations. Given the study design, should the question be rephrased? If so, how?

A) No, the question is worded accurately.
B) Yes, revise wording to "Is there an association between beauty and course evaluations?"
C) Yes, revise wording to "Does beauty score increase the professor's course evaluations?"
D) Yes, revise wording to "Does beauty score decrease the professor's course evaluations?"

Question 3

Which of the following statements is **false** about the distribution of `score` ?

A) The median of the distribution is 4.3.
B) 25% of the students gave their professors a score of over 4.6.
C) 11 of students gave a professor a score below 3.
C) The left skewness of the data suggests that the students are less likely to rate the professors highly.

Exercise: Excluding `score`, select two other variables and describe their relationship using an appropriate visualization (scatterplot, side-by-side boxplots, or mosaic plot).

Simple linear regression

The fundamental phenomenon suggested by the study is that better looking teachers are evaluated more favorably. Let's create a scatterplot to see if this appears to be the case:

```
plot(evals$score ~ evals$bty_avg)
```

Before we draw conclusions about the trend, compare the number of observations in the data frame with the approximate number of points on the scatterplot. Is anything awry?

Exercise: Replot the scatterplot, but this time use the function `jitter()` on the y- or the x-coordinate. (Use `?jitter` to learn more.) What was misleading about the initial scatterplot?

Exercise: Let's see if the apparent trend in the plot is something more than natural variation. Fit a linear model called `m_bty` to predict average professor score by average beauty rating and add the line to your plot using `abline(m_bty)`. Write out the equation for the linear model and interpret the slope.

Question 4   Average beauty score seems to be a statistically and practically significant predictor.

A) True
B) False

Question 5   Use residual plots to evaluate whether the conditions of least squares regression are reasonable. Which of the following statements is an incorrect analysis of the residual plots and conditions?

A) Linear association: The residuals plot shows a random scatter.
B) Constant variance of residuals: No fan shape in residuals plot.
C) Nearly normal residuals: Residuals are right skewed, but the sample size is large, so this may not be an important violation of conditions.
D) Independent observations: Classes sampled randomly, no order effect.

Multiple linear regression

The data set contains several variables on the beauty score of the professor: individual ratings from each of the six students who were asked to score the physical appearance of the professors and the average of these six scores. Let's take a look at the relationship between one of these scores and the average beauty score.

```
plot(evals$bty_avg ~ evals$bty_f1lower)
cor(evals$bty_avg, evals$bty_f1lower)
```

As expected the relationship is quite strong; after all, the average score is calculated using the individual scores. We can actually take a look at the relationships between all beauty variables (columns 13 through 19) using the following command:

```
plot(evals[,13:19])
```

These variables are collinear (correlated), and adding more than one of these variables to the model would not add much value to the model. In this application and with these highly-correlated predictors, it is reasonable to use the average beauty score as the single representative of these variables.

In order to see if beauty is still a significant predictor of professor score after we've accounted for the gender of the professor, we can add the gender term into the model.

```
m_bty_gen <- lm(score ~ bty_avg + gender, data = evals)
summary(m_bty_gen)
```

Question 6   P-values and parameter estimates should only be trusted if the conditions for the regression are reasonable. Using diagnostic plots, we can conclude that the conditions for this model are reasonable.

A) True
B) False

Exercise: Is `bty_avg` still a significant predictor of `score`? Has the addition of `gender` to the model changed the parameter estimate for `bty_avg`?

Note that the estimate for `gender` is now called `gendermale`. You'll see this name change whenever you introduce a categorical variable. The reason is that R recodes `gender` from having the values of `female` and `male` to being an indicator variable called `gendermale` that takes a value of 0 for females and a value of 1 for males. (Such variables are often referred to as "dummy" variables.)

As a result, for females, the parameter estimate is multiplied by zero, leaving the intercept and slope form familiar from simple regression.

$$\widehat{score} = b_0 + b_1 \times bty_avg + b_2 \times (0)$$
$$= b_0 + b_1 \times bty_avg$$

We can plot this line and the line corresponding to males with the following custom function.

```
multiLines(m_bty_gen)
```

Exercise: What is the equation of the line corresponding to males? (*Hint:* For males, the parameter estimate is multiplied by 1.)

Question 7   For two professors (one male and one female) who received the same
beauty rating, the male professor is predicted to have the higher course
evaluation score than the female.

A) True
B) False

The decision to call the indicator variable `gendermale` instead of `genderfemale` has no deeper
meaning. R simply codes the category that comes first alphabetically as a 0.

You can change the reference level of a categorical variable, which is the level that is coded as a 0, using
the `relevel` function. Use `?relevel` to learn more.

Exercise: Create a new model called `m_bty_rank` with `gender` removed and `rank` added in. How does
R appear to handle categorical variables that have more than two levels? Note that the rank variable has
three levels: teaching, tenure track, tenured.

Question 8   Which of the following is the correct order of the three levels of rank if we
were to order them from lowest predicted course evaluation score to
highest predicted course evaluation score?

A) Teaching, Tenure Track, Tenured
B) Tenure track, Tenured
C) Tenure Track, Tenured, Teaching
D) Teaching, Tenured, Tenure Track

The interpretation of the coefficients in multiple regression is slightly different from that of simple
regression. The estimate for `bty_avg` reflects how much higher a group of professors is expected to
score if they have a beauty rating that is one point higher *while holding all other variables constant*. In this
case, that translates into considering only professors of the same rank with `bty_avg` scores that are one
point apart.

The search for the best model

We will start with a full model that predicts professor score based on rank, ethnicity, gender, language of
the university where they got their degree, age, proportion of students that filled out evaluations, class
size, course level, number of professors, number of credits, average beauty rating, outfit, and picture
color.

Exercise: Which variable would you expect to have the highest p-value in this model? Why? *Hint:* Think
about which variable would you expect to not have any association with the professor score.

Let's run the model…

```
m_full <- lm(score ~ rank + ethnicity + gender + language + age + cls_perc_eval
            + cls_students + cls_level + cls_profs + cls_credits + bty_avg, data
= evals)
summary(m_full)
```

Note that we have not included the `pic_outfit` or `pic_color` variables in the full model because the original study states that these variables were used in a different analysis evaluating whether they're related to how highly the six students inolved in the study score the professors' beauty (not related to how the students evaluate their professors in class).

Exercise: Check your suspicions from the previous exercise. Include the model output in your response.

Question 9   Which of the following is the correct intrepetation of the coefficient associated with the ethnicity variable.
Non-minority professors are expected on average to score …

A) 0.19 points lower than minority professors, all else held constant.
B) 0.19 points higher than minority professors, all else held constant.
C) 0.02 points lower than minority professors, all else held constant.
D) 0.02 points higher than minority professors, all else held constant.

Exercise: Drop the variable with the highest p-value and re-fit the model. Did the coefficients and significance of the other explanatory variables change? (One of the things that makes multiple regression interesting is that coefficient estimates depend on the other variables that are included in the model.) If not, what does this say about whether or not the dropped variable was collinear with the other explanatory variables?

Now we try a different model selection method: adjusted R^2. Create a new model, `m1`, where you remove `rank` from the list of explanatory variables. Check out the adjusted R^2 of this new model and compare it to the adjusted R^2 of the full model.

```
m1 <- lm(score ~ ethnicity + gender + language + age + cls_perc_eval
            + cls_students + cls_level + cls_profs + cls_credits + bty_avg, data
= evals)
summary(m1)$adj.r.squared
```

Then, try dropping the next variable from the full model (`ethnicity`):

```
m2 = lm(score ~ rank + gender + language + age + cls_perc_eval +
    cls_students + cls_level + cls_profs + cls_credits + bty_avg, data = evals)
summary(m2)$adj.r.squared
```

Exercise: Repeat this process until you have tried removing each variable from the full model at a time, and determine removal of which variable yields the highest improvement in the adjusted R^2.

Question 10   Elimination of which variable from the full model yielded the highest adjusted R-squared?

A) `bty_avg`
B) `cls_profs`
C) `cls_students`
D) `rank`

To complete the model selection we would continue removing variables one at a time until removal of another variable did not increase adjusted R^2.

Exercise: The original paper describes how these data were gathered by taking a sample of professors from the University of Texas at Austin and including all courses that they have taught. Considering that each row represents a course, could this new information have an impact on any of the conditions of linear regression?

Exercise: Based on your final model, describe the characteristics of a professor and course at University of Texas at Austin that would be associated with a high evaluation score.

Exercise: Would you be comfortable generalizing your conclusions to apply to professors generally (at any university)? Why or why not?