Multiple Linear Regression

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 might reflect the influence of nonteaching-related characteristics, such as the physical appearance of the instructor. An article titled "Beauty in the classroom: instructors' pulchritude and putative pedagogical productivity" (Hamermesh and Parker2005) found that instructors who are viewed to be better looking receive higher instructional ratings.[1]

In this lab, we analyze the data from this study in order to learn what goes into a positive professor evaluation.[2]

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.[3] The result is a data set where each row contains a different course and columns represent variables about the courses and professors.

- Note: If you are using SAS University Edition, you need to ensure that interactive mode is turned off. To do this, click the button to the right of **Sign Out** in the upper right corner of the window and then click Preferences. In the **Preferences** window, on the **General** tab, the bottom check box (located next to the text **Start new programs** in interactive mode) should not be selected. If the box is selected, you need to clear it and save your change.

---

[1] Daniel S. Hamermesh and 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).

[2] This is a product of OpenIntro that is released under a Creative Commons Attribution-ShareAlike 3.0 Unported (http://creativecommons.org/licenses/by-sa/3.0/). This lab was written for OpenIntro by Mine C¸etinkaya-Rundel and Andrew Bray and modified by SAS Institute Inc. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries (® indicates USA registration) and are not included under the CC-BY-SA license.

[3] 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).

```
filename evals url
         'http://www.openintro.org/stat/data/evals_sas.csv';

proc import datafile=evals
            out=evals
            dbms=csv
            replace;
   getnames=yes;
run;
```

score	Average professor evaluation score: (1) very unsatisfactory through (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, and so on), multi- credit
bty_f1lower	Beauty rating of professor from lower-level female: (1) lowest to (10) highest
bty_f1upper	Beauty rating of professor from upper-level female: (1) lowest to (10) highest
bty_f2upper	Beauty rating of professor from second upper-level female: (1) lowest to (10) highest
bty_m1lower	Beauty rating of professor from lower-level male: (1) lowest to (10) highest
bty_m1upper	Beauty rating of professor from upper-level male: (1) lowest to (10) highest

bty_m2upper	Beauty rating of professor from second upper-level male: (1) lowest to (10) highest
bty_avg	Average beauty rating of professor
pic_outfit	Outfit of professor in picture: not formal, formal
pic_color	Color of professor's picture: color, black & white

Exercise 1: Is this an observational study or an experiment? The original research question posed in the paper is whether beauty leads directly to the differences in course evaluations. Given the study design, is it possible to answer this question as it is phrased? If not, rephrase the question.

Exercise 2: Describe the distribution of **score**. Is the distribution skewed? What does that tell you about how students rate courses? Is this what you expected to see? Why or why not?

Exercise 3: Excluding **score**, select two other variables and describe their relationship using an appropriate visualization (scatter plot, side-by-side box plots, 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 scatter plot to see whether this appears to be the case:

```
proc sgplot data=evals;
   scatter y=score x=bty_avg;
run;
```

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

Exercise 4: Re-plot the scatter plot, but this time use the JITTER option in the SCATTER statement. What was misleading about the initial scatter plot?

Exercise 5: Let's see whether the apparent trend in the plot is something more than natural variation. Fit a linear model to predict average professor score by average beauty rating and redo the scatter plot with the regression line added to your plot. Write out the equation for the linear model and interpret the slope. Is average beauty score a statistically significant predictor? Does it appear to be a practically significant predictor?

Exercise 6: Use residual plots to evaluate whether the conditions of least squares regression are reasonable. Provide plots and comments for each one.

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 quick look at the relationship between one of these scores and the average beauty score: proc corr

```
proc corr data=evals plots=scatter(ellipse=none);
   var bty_avg bty_f1lower;
run;
```

The option ELLIPSE=NONE is specified to remove prediction ellipses from the scatter plot, which are added by default.

As expected the relationship is quite strong. After all, the average score is calculated using the individual scores. We can actually examine the relationships between all beauty variables (variables **bty_f1lower** through **bty_avg**) using the following code:

```
ods select matrixplot;
proc corr data=evals plots(maxpoints=20000)=matrix(nvar=7);
   var bty_f1lower--bty_avg;
run;
```

In the code above, the ODS SELECT MATRIXPLOT statement is included to tell SAS to output only the matrix plot. That statement could be excluded, but that would result in extra output that we do not want. The MAXPOINTS=20000 option is included because SAS, by default, suppresses plots with more than 5000 points. NVAR=7 was specified because, by default, the CORR procedure limits matrix plots to five variables.

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 a single representative of these variables.

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

```
proc glm data=evals;
   class gender / ref=first;
   model score=bty_avg gender / solution;
run;
quit;
```

Note that we use the GLM procedure here, rather than the REG procedure. The GLM procedure enables us to add categorical variables (for example, **gender**) to a CLASS statement. This tells

SAS to create one or more dummy variables for each categorical variable. (The REG procedure would require us to create the dummy variables ourselves.) By default, the last category is the reference category. REF=FIRST specifies that the first category is the reference category. The SOLUTION option tells SAS to output estimates of the regression coefficients.

Exercise 7: p-values and parameter estimates should be trusted only if the conditions for the regression are reasonable. Verify that the conditions for this model are reasonable using diagnostic plots.

Exercise 8: 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 labeled **gender male**. The reason is that SAS recodes **gender** from having the values of female and male to being an indicator variable that takes a value of 0 for females and a value of 1 for males. (Such variables are often referred to as dummy variables.) The output also contains a row for **gender female** with an estimate of 0 and missing values for the rest of the row, indicating that the value female is the reference category for **gender**.

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

$$\begin{aligned}\widehat{score} &= \hat{\beta}_0 + \hat{\beta}_1 \times bty\_avg + \hat{\beta}_2 \times (0) \\ &= \hat{\beta}_0 + \hat{\beta}_1 \times bty\_avg\end{aligned}$$

This line and the line corresponding to males were already plotted by our previous call to PROC GLM.

Exercise 9: What is the equation of the line corresponding to males? (Hint: For males, the parameter estimate is multiplied by 1.) For two professors who received the same beauty rating, which gender tends to have the higher course evaluation score?

The decision to use female as the reference category for **gender** has no deeper meaning. SAS simply codes the category that comes last alphabetically as the reference category. (You can change the reference level of a categorical variable by using the REF= option in SAS.)

Exercise 10: Estimate a new model with **gender** removed and **rank** added in. How does SAS appear to handle categorical variables that have more than two levels? Note that the **rank** variable has three levels: teaching, tenure track, tenured.

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

Let's run the model.

```
proc glm data=evals;
   class rank ethnicity gender language cls_level cls_profs
         cls_credits pic_outfit pic_color / ref=first;
   model score=rank ethnicity gender language age cls_perc_eval
         cls_students cls_level cls_profs cls_credits bty_avg
         pic_outfit pic_color / solution;
run;
quit;
```

Exercise 12: Check your suspicions from the previous exercise. Include the model output in your response.

Exercise 13: Interpret the coefficient associated with the ethnicity variable.

Exercise 14: Drop the variable with the highest p-value and refit the model. Did the coefficients and significance of the other explanatory variables change? (One thing 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 the dropped variable was collinear with the other explanatory variables?

Exercise 15: Using backward-elimination and significance level as the selection criterion, determine the best model. (Note that you will need to switch to GLMSELECT to use backward-elimination.) You do not need to show all steps in your answer, just the output for the final model. Also, write out the linear model for predicting score based on the final model that you settle on.

Exercise 16: Verify that the conditions for this model are reasonable using diagnostic plots.

Exercise 17: 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 18: 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 19: Would you be comfortable generalizing your conclusions to apply to professors generally (at any university)? Why or why not?

Notes:

This lab was adapted for *OpenIntro* by Andrew Bray and mine Cetinkaya-Rundel from a lab written by Mark Hansen of UCLA Statistics.