Do non-teaching related characteristics, such as the physical appearance of the instructor, influence their teaching ratings?

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.

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

Let’s load the data and attach a necessary package:

```r
library(ggplot2)
load(url("http://www.openintro.org/stat/data/evals.RData"))
```

Exploring the Data

Next we explore the data set by viewing its dimensions, summary of each column, and glance at the first 6 lines.

```r
dim(evals)
summary(evals)
```

knitr::kable(head(evals[1:10]), caption = "First 6 lines of evals")
Table 1: First 6 lines of evals

<table>
<thead>
<tr>
<th>score</th>
<th>rank</th>
<th>ethnicity</th>
<th>gender</th>
<th>language</th>
<th>age</th>
<th>cls_perc_eval</th>
<th>cls_did_eval</th>
<th>cls_students</th>
<th>cls_level</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.7</td>
<td>tenure track</td>
<td>minority</td>
<td>female</td>
<td>english</td>
<td>36</td>
<td>55.81395</td>
<td>24</td>
<td>43</td>
<td>upper</td>
</tr>
<tr>
<td>4.1</td>
<td>tenure track</td>
<td>minority</td>
<td>female</td>
<td>english</td>
<td>36</td>
<td>68.80000</td>
<td>86</td>
<td>125</td>
<td>upper</td>
</tr>
<tr>
<td>3.9</td>
<td>tenure track</td>
<td>minority</td>
<td>female</td>
<td>english</td>
<td>36</td>
<td>60.80000</td>
<td>76</td>
<td>125</td>
<td>upper</td>
</tr>
<tr>
<td>4.8</td>
<td>tenure track</td>
<td>minority</td>
<td>female</td>
<td>english</td>
<td>36</td>
<td>62.60163</td>
<td>77</td>
<td>123</td>
<td>upper</td>
</tr>
<tr>
<td>4.6</td>
<td>tenured</td>
<td>not minority</td>
<td>male</td>
<td>english</td>
<td>59</td>
<td>85.00000</td>
<td>17</td>
<td>20</td>
<td>upper</td>
</tr>
<tr>
<td>4.3</td>
<td>tenured</td>
<td>not minority</td>
<td>male</td>
<td>english</td>
<td>59</td>
<td>87.50000</td>
<td>35</td>
<td>40</td>
<td>upper</td>
</tr>
</tbody>
</table>

```
knitr::kable(head(evals[11:18]), caption = "First 6 lines of evals")
```

Table 2: First 6 lines of evals

<table>
<thead>
<tr>
<th>cls_profs</th>
<th>cls_credits</th>
<th>bty_f1lower</th>
<th>bty_f1upper</th>
<th>bty_f2upper</th>
<th>bty_m1lower</th>
<th>bty_m1upper</th>
<th>bty_m2upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>single</td>
<td>multi credit</td>
<td>5</td>
<td>7</td>
<td>6</td>
<td>2</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>single</td>
<td>multi credit</td>
<td>5</td>
<td>7</td>
<td>6</td>
<td>2</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>single</td>
<td>multi credit</td>
<td>5</td>
<td>7</td>
<td>6</td>
<td>2</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>single</td>
<td>multi credit</td>
<td>5</td>
<td>7</td>
<td>6</td>
<td>2</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>multiple</td>
<td>multi credit</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>multiple</td>
<td>multi credit</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

```
knitr::kable(head(evals[19:21]), caption = "First 6 lines of evals")
```

Table 3: First 6 lines of evals

<table>
<thead>
<tr>
<th>bty_avg</th>
<th>pic_outfit</th>
<th>pic_color</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>not formal</td>
<td>color</td>
</tr>
<tr>
<td>5</td>
<td>not formal</td>
<td>color</td>
</tr>
<tr>
<td>5</td>
<td>not formal</td>
<td>color</td>
</tr>
<tr>
<td>5</td>
<td>not formal</td>
<td>color</td>
</tr>
<tr>
<td>3</td>
<td>not formal</td>
<td>color</td>
</tr>
<tr>
<td>3</td>
<td>not formal</td>
<td>color</td>
</tr>
</tbody>
</table>
Relationship between beauty scores and evaluation scores

Let’s first study `bty_avg` and `score` individually.

```r
summary(evals$score)
```

```
# Min. 1st Qu. Median Mean 3rd Qu. Max.
## 2.300 3.800 4.300 4.175 4.600 5.000
```

```r
summary(evals$bty_avg)
```

```
# Min. 1st Qu. Median Mean 3rd Qu. Max.
## 1.667 3.167 4.333 4.418 5.500 8.167
```

Next, let’s find a correlation between them:

```r
cor(evals$bty_avg, evals$score)
```

```
# [1] 0.1871424
```

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:

```r
qplot(data = evals, x = bty_avg, y = score)
```

```r
dim(evals)
```

```
# [1] 463 21
```

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. Let’s redraw it by adding random noise to the points:

```r
qplot(data = evals, x = bty_avg, y = score, geom = "jitter")
```

Simple Linear Regression

Let’s superimpose a "regression line":

```r
qplot(data = evals, x = bty_avg, y = score, geom = "jitter") + geom_smooth(method = "lm", se = FALSE) + xlim(0, 8.5)
```
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?

Fit a linear model called \( m_{\text{bty}} \) to predict average professor score by average beauty rating. Write out the equation for the linear model and interpret the slope and the intercept.

\[
\text{LM\_model} \leftarrow \text{lm(score} \sim \text{bty\_avg, data = evals)}
\]

\[
\text{summary(LM\_model)}
\]

Does the relationship seem linear? Is it strong?

Is average beauty score a statistically significant predictor? Does it appear to be a practically significant predictor?
Prediction

One of the reasons why regression models are so useful is that they allow us to predict outcomes. Although the model that we've built is still fairly simple, we can practice using it to predict expected evaluation scores given a specific beauty score. For example, for \texttt{bty.avg}=5:

\begin{verbatim}
predict(LM_model, newdata = data.frame(bty_avg = 5), interval = "prediction")
\end{verbatim}

\begin{verbatim}
## fit lwr upr
## 1 4.213523 3.161207 5.265839
\end{verbatim}

Multiple linear regression

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:

\begin{verbatim}
m_bty_gen <- lm(score ~ bty_avg + gender, data = evals)
summary(m_bty_gen)
\end{verbatim}

\begin{verbatim}
## Call:
## lm(formula = score ~ bty_avg + gender, data = evals)
## ## Residuals:
## ## Min 1Q Median 3Q Max
## ## -1.8305 -0.3625 0.1055 0.4213 0.9314
## ## ## Coefficients:
## ## Estimate Std. Error t value Pr(>|t|)
## ## (Intercept) 3.74734 0.08466 44.266 < 2e-16 ***
## ## bty_avg 0.07416 0.01625 4.563 6.48e-06 ***
## ## gendermale 0.17239 0.05022 3.433 0.000652 ***
## ## ---
## ## Signif. codes:
## ## 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
## ## ## Residual standard error: 0.5287 on 460 degrees of freedom
## ## Multiple R-squared: 0.05912, Adjusted R-squared: 0.05503
## ## F-statistic: 14.45 on 2 and 460 DF, p-value: 8.177e-07
\end{verbatim}

Is bty-avg still a significant predictor of score? Has the addition of gender to the model changed the parameter estimate for bty-avg?

What is the effect of gender? How can we interpret it?

The interpretation of the coefficients in multiple regression is slightly different from that of simple regression. The estimate for \texttt{bty.avg} reflects how much higher a group of professors is expected to score if they have a beauty score of 5 compared to 0. The coefficient for \texttt{gendermale} indicates the effect of gender, with a positive estimate suggesting that male professors are expected to score higher than female professors, all else being equal.
rating that is one point higher while holding all other variables constant. In this case, that translates into considering only male or only female professors with bty_avg scores that are one point apart.