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 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.)

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 aslightly 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("more/evals.RData")
```

variable	description
`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.
`pic_color`	color of professor's picture: color, black & white.

Exploring the data

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.

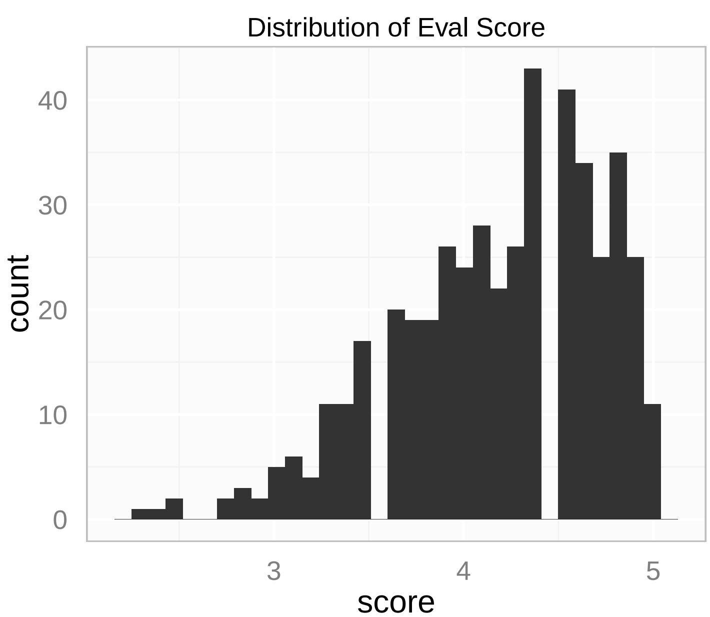
   This is an experiment but it uses observational data in both the the treatment (beauty measure), and with the course evaluation (as the response variable). Although the reseach is designed as an experiment, all aspects are extremely subjective any doesn't seem to be a good design for causal inference. A better research question would be whether beauty is correlated with differences in course evaluations.

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?

   The distribution of `score` is skewed left. This suggests students tend to rate fairly high typically. I had not developed any expectation before running the analysis, but I guess I would have assumed a bit more normalness and less skew in the sense that not all professors are as good as others and this should be reflected in the scores.

```
g1 <- ggplot(data=evals) +
  geom_histogram(aes(x=score)) +
  labs(title="Distribution of Eval Score") +
  myTheme
g1
```

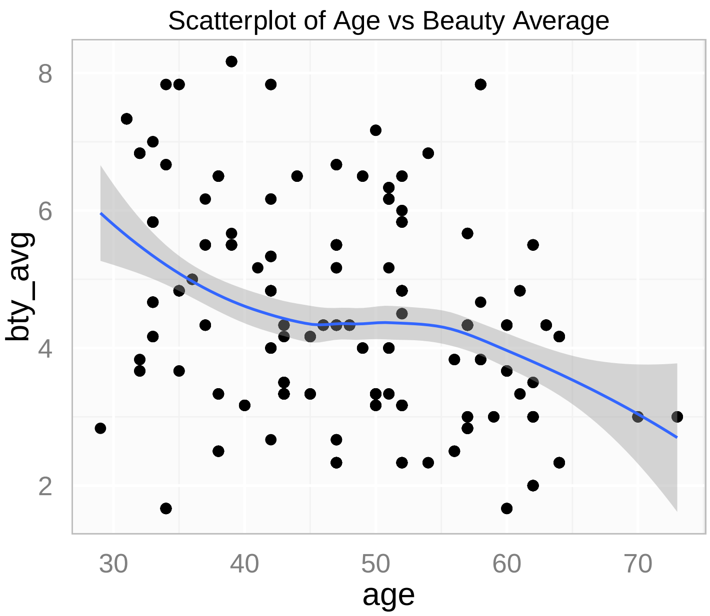
```
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
```



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

```
g1 <- ggplot(data=evals, aes(x=age, y=bty_avg)) +
  geom_point() +
  geom_smooth() +
  labs(title="Scatterplot of Age vs Beauty Average") +
  myTheme
g1
```

```
## geom_smooth: method="auto" and size of largest group is <1000, so using loess. Use 'method = x' to cl
```

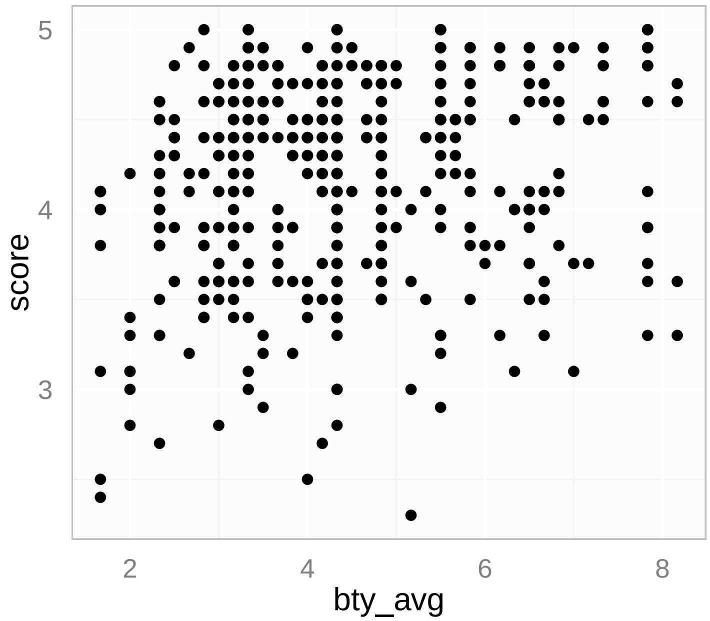


Simple linear regression

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:

```
g1 <- ggplot(data=evals) +
  geom_point(aes(x=bty_avg, y=score), position="dodge") + myTheme
g1
```

```
## ymax not defined: adjusting position using y instead
```




```
nrow(evals)
```

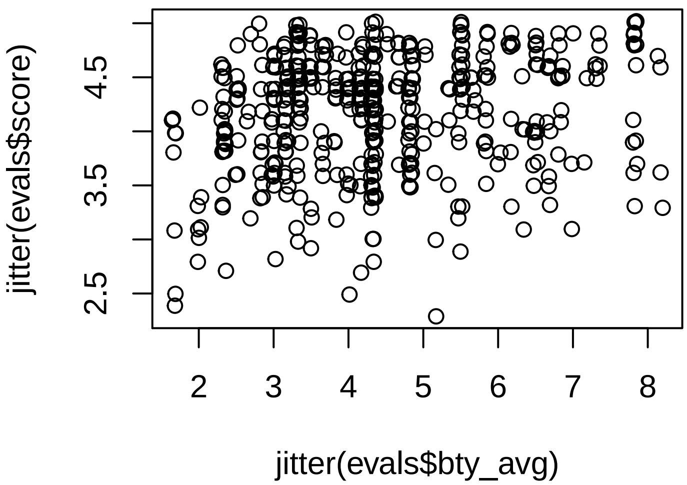
```
## [1] 463
```

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?

Approximate number of points in the scatterplot? That would be an awfully poor estimate if I made one. I dont' see anything awry. Maybe I'm missing something.

4. 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? **Ok, so there were duplicate scores that were overlapping in the original plot.**

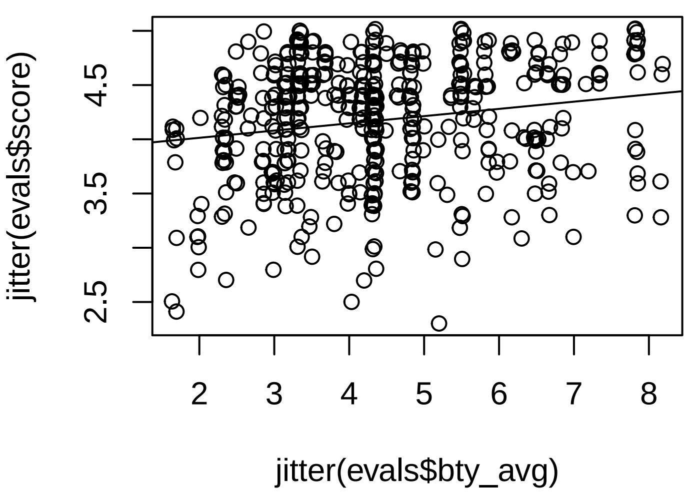
```
plot(jitter(evals$score) ~ jitter(evals$bty_avg))
```



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

```
m_bty <- lm(evals$score ~ evals$bty_avg)
plot(jitter(evals$score) ~ jitter(evals$bty_avg))

abline(m_bty)
```



```r
summary(m_bty)
```

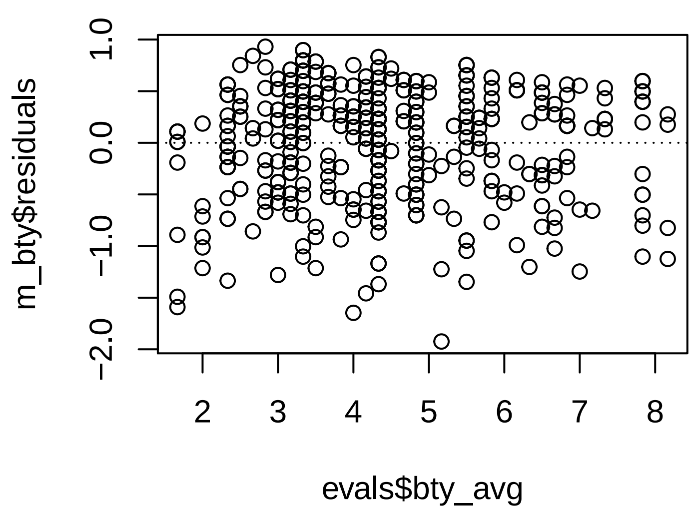
```
##
## Call:
## lm(formula = evals$score ~ evals$bty_avg)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.9246 -0.3690  0.1420  0.3977  0.9309
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)     3.88034    0.07614   50.96  < 2e-16 ***
## evals$bty_avg   0.06664    0.01629    4.09 5.08e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5348 on 461 degrees of freedom
## Multiple R-squared:  0.03502,    Adjusted R-squared:  0.03293
## F-statistic: 16.73 on 1 and 461 DF,  p-value: 5.083e-05
```

$$\hat{y} = 3.880338 + 0.066637x$$

Interpreting the slope, for every 1 unit increase in beauty index, the course evaluation would increase by ~0.67. While the p-value is less than 0.05, practically speaking, it is not a significant predictor as evidenced by the $R^2 \approx 0.03$.

6. Use residual plots to evaluate whether the conditions of least squares regression are reasonable. Provide plots and comments for each one (see the Simple Regression Lab for a reminder of how to make these).

```r
plot(m_bty$residuals ~ evals$bty_avg)
abline(h = 0, lty = 3)  # adds a horizontal dashed line at y = 0
```



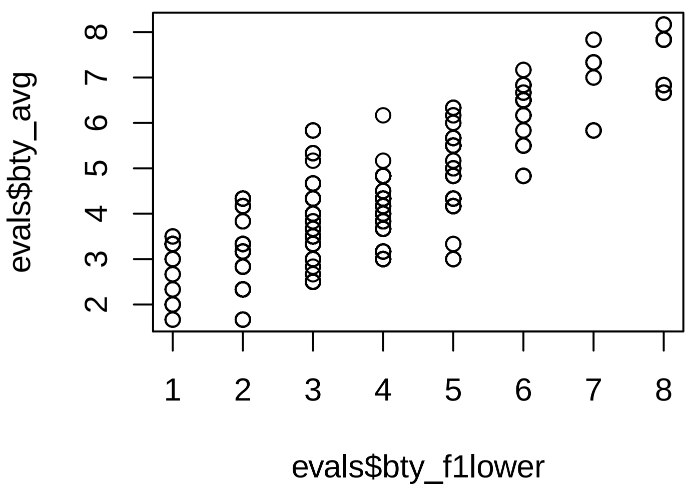
Nearly normal residuals: Based on the plot above, the residuals appear to be nearly normal.

Linearity: The data are clearly not having a narrow linear relationship. The points are wide in their trend.

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)
```

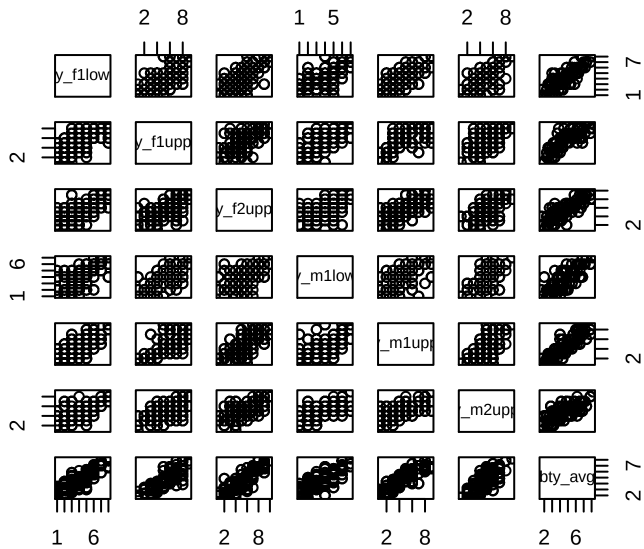



```
cor(evals$bty_avg, evals$bty_f1lower)
```

```
## [1] 0.8439112
```

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.

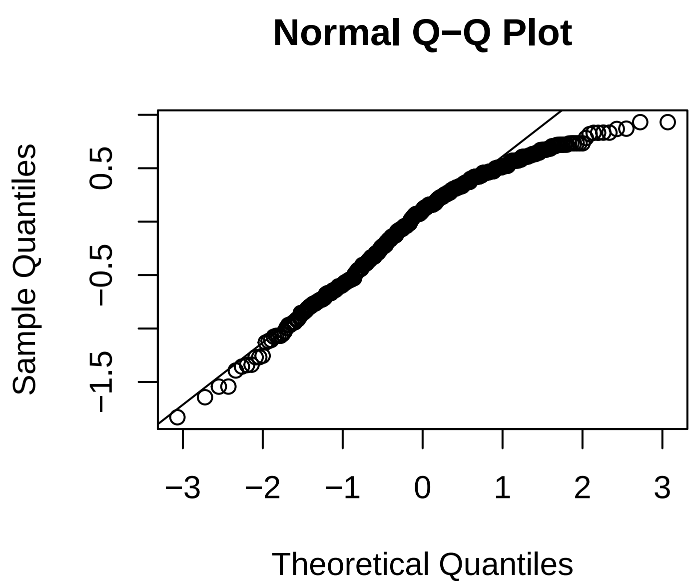
In order to see if beauty is still a significant predictor of professor score after we've accounted for the gender of the professor, we can add the gender term into the model.

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

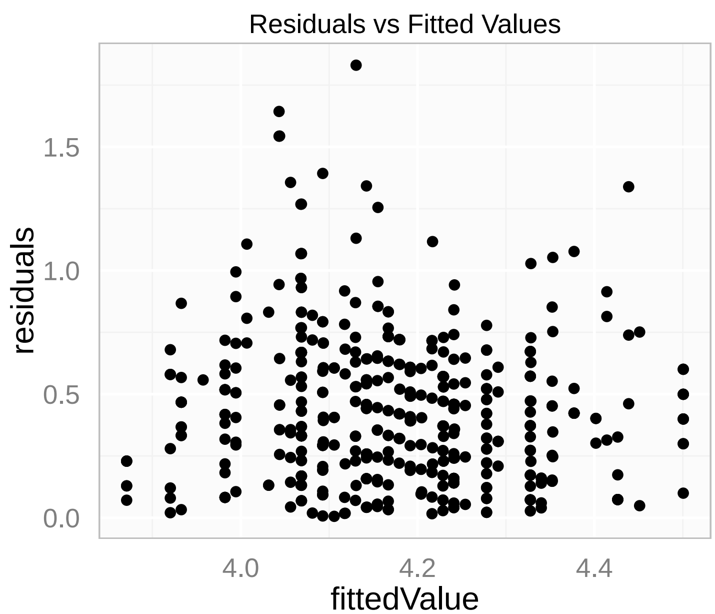
```
##
## Call:
## lm(formula = score ~ bty_avg + gender, data = evals, y = TRUE)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.8305 -0.3625  0.1055  0.4213  0.9314
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.74734    0.08466  44.266  < 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
```

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

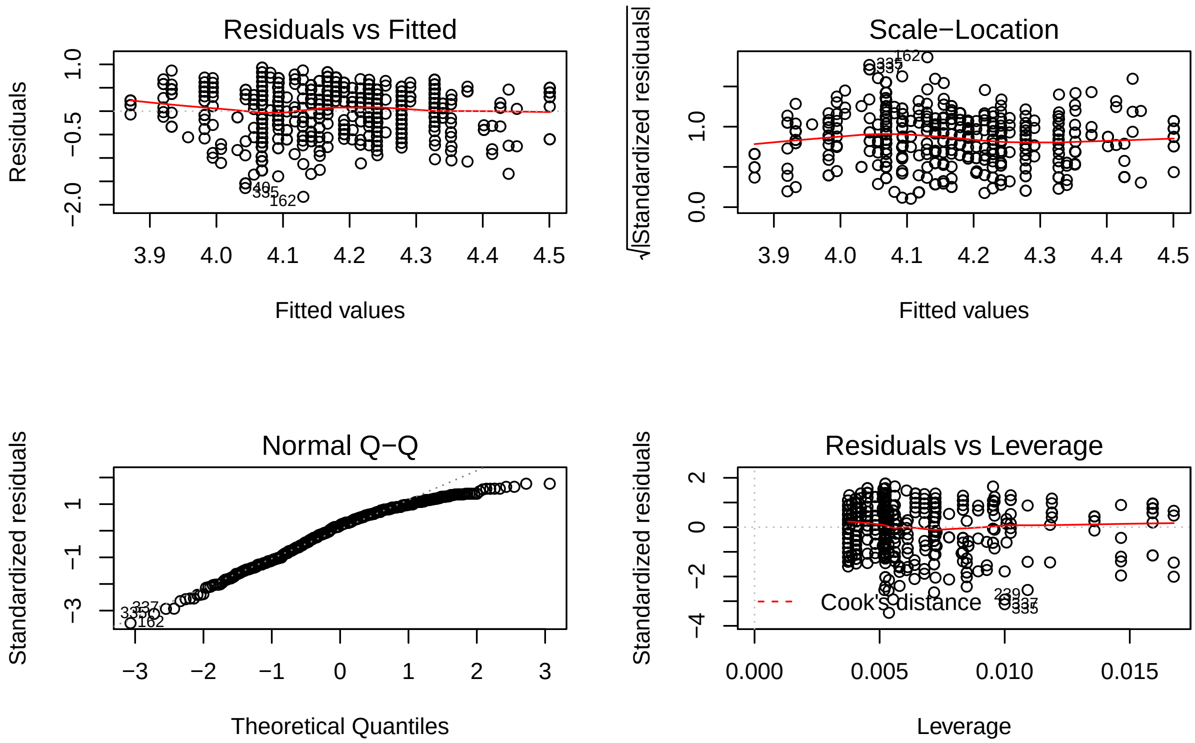
Nearly normal residuals seem to hold as shown in the QQ plot below.



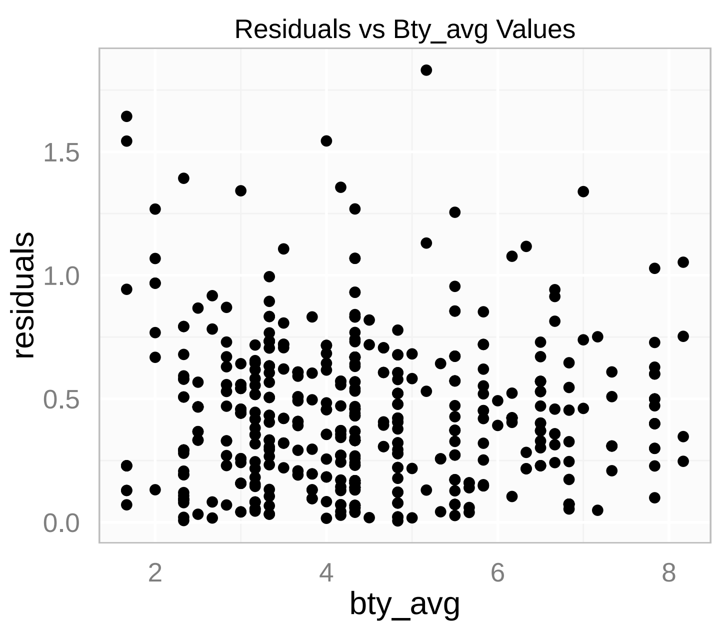
The constant variability of residuals is checked via a comparison to the fitted values. As shown in the following visualization, the data appears to satisfy the conditions for regression.



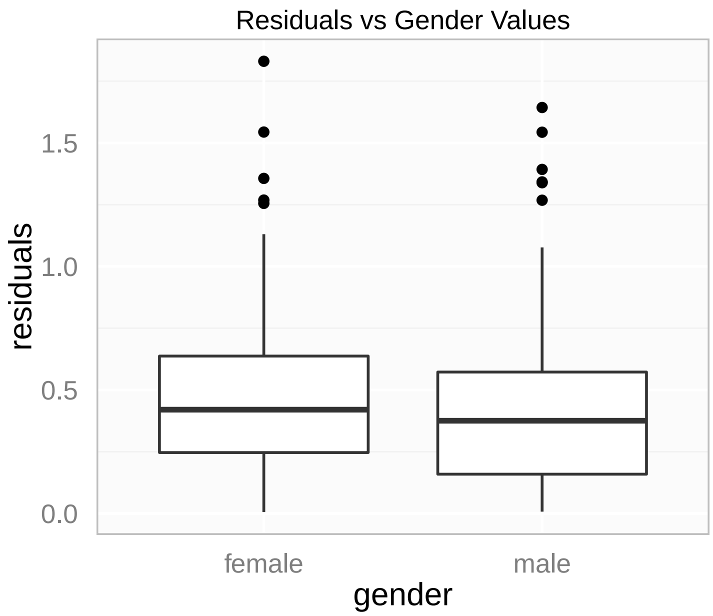
Based on code from http://www.statmethods.net/stats/regression.html, we can view a full set of diagnostic plots to verify the conditions:



When comparing the residuals to the `bty_avg` predictor values, we see there is not specific structure and the residuals are mostly having constant variability. This supports the concept that the residuals are independent.



Comparing to the `gender` variable, the distribution is slightly different but similar. The difference doesn't appear significant enough to invalidate the model.



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`?

   Revisiting the summary of the model we see the parameter estimte for `bty_avg` is now 0.07416 whereas before gender was introduced, the estimate was 0.06664, a 0.00752 increase. Although the parameter estimate changed for `bty_avg`, the t-value is effectively identical and stronger than gender. Seperately, the R^2 is 0.05912 with both `bty_avg` and `gender`, and with only `bty_avg` the R^2 is 0.03502. Therefore I conclude that adding `gender` to the model is useful, but on the whole, `bty_avg` is not a significant predictor of `score`.

```
##
## Call:
## lm(formula = score ~ bty_avg + gender, data = evals, y = TRUE)
##
## 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
```

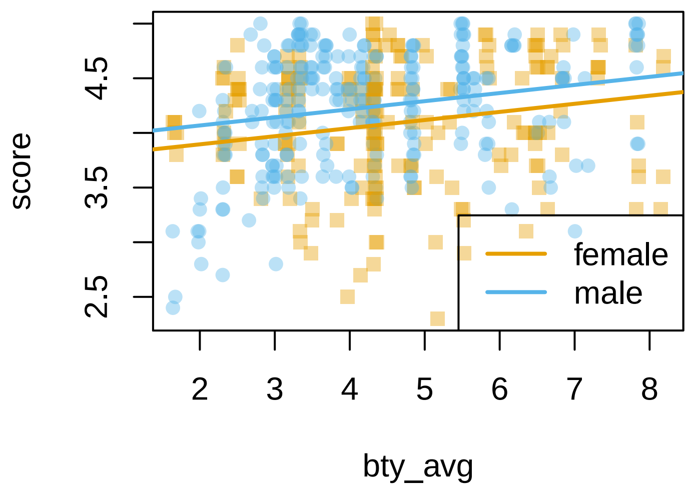
Note that the estimate for `gender` is now called `gendermale`. You'll see this name change whenever you introduce a categorical variable. The reason is that R recodes `gender` from having the values of `female` and `male` to being an indicator variable called `gendermale` that takes a value of 0 for 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.

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

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

```
multiLines(m_bty_gen)
```



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?

$$\widehat{score} = 3.7473382 + 0.0741554 \times bty\_avg + 0.0741554 \times (1)$$

   Since we are adding 0.0741554 for men, this means men will tend to have a higher course evaluation score. This is evident in the chart shown above also, where the blue line (male) is above the brown line (female).

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.)

10. 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`.

R appears to handle categorical variabels having more than two levels by adding an additional parameter for each additional level above 2. Therefore `teaching` would yield $\hat{B}_2 = 0, \hat{B}_3 = 0$. `tenure track` would be $\hat{B}_2 = 1, \hat{B}_3 = 0$, and 'tenured would be $\hat{B}_2 = 0, \hat{B}_3 = 1$

```
m_bty_rank <- lm(score ~ bty_avg + rank, data=evals)
summary(m_bty_rank)
```

```
##
## Call:
## lm(formula = score ~ bty_avg + rank, data = evals)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.8713 -0.3642  0.1489  0.4103  0.9525
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       3.98155    0.09078  43.860  < 2e-16 ***
## bty_avg           0.06783    0.01655   4.098 4.92e-05 ***
## ranktenure track -0.16070    0.07395  -2.173   0.0303 *
## ranktenured      -0.12623    0.06266  -2.014   0.0445 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5328 on 459 degrees of freedom
## Multiple R-squared:  0.04652,    Adjusted R-squared:  0.04029
## F-statistic: 7.465 on 3 and 459 DF,  p-value: 6.88e-05
```

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.

11. 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.

I would hypothesize that `language`, language of the university where they got their degree, would have the highest p-value.

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
             + pic_outfit + pic_color, data = evals)
summary(m_full)
```

```
##
## Call:
## lm(formula = score ~ rank + ethnicity + gender + language + age +
##     cls_perc_eval + cls_students + cls_level + cls_profs + cls_credits +
##     bty_avg + pic_outfit + pic_color, data = evals)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.77397 -0.32432  0.09067  0.35183  0.95036
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)            4.0952141  0.2905277  14.096  < 2e-16 ***
## ranktenure track      -0.1475932  0.0820671  -1.798  0.07278 .
## ranktenured           -0.0973378  0.0663296  -1.467  0.14295
## ethnicitynot minority  0.1234929  0.0786273   1.571  0.11698
## gendermale             0.2109481  0.0518230   4.071 5.54e-05 ***
## languagenon-english   -0.2298112  0.1113754  -2.063  0.03965 *
## age                   -0.0090072  0.0031359  -2.872  0.00427 **
## cls_perc_eval          0.0053272  0.0015393   3.461  0.00059 ***
## cls_students           0.0004546  0.0003774   1.205  0.22896
## cls_levelupper         0.0605140  0.0575617   1.051  0.29369
## cls_profssingle       -0.0146619  0.0519885  -0.282  0.77806
## cls_creditsone credit  0.5020432  0.1159388   4.330 1.84e-05 ***
## bty_avg                0.0400333  0.0175064   2.287  0.02267 *
## pic_outfitnot formal  -0.1126817  0.0738800  -1.525  0.12792
## pic_colorcolor        -0.2172630  0.0715021  -3.039  0.00252 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.498 on 448 degrees of freedom
## Multiple R-squared:  0.1871, Adjusted R-squared:  0.1617
## F-statistic: 7.366 on 14 and 448 DF,  p-value: 6.552e-14
```

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

    It appears that cls_prof, number of professors, has the least impact on score with a p-value of 0.77806.

13. Interpret the coefficient associated with the ethnicity variable.

Holding all else equal, when ethnicity is `not minority` the score tends to increase by 0.1234929.

14. 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?

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

```
##
## Call:
## lm(formula = score ~ rank + ethnicity + gender + language + age +
##     cls_perc_eval + cls_students + cls_level + cls_credits +
##     bty_avg + pic_outfit + pic_color, data = evals)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.7836 -0.3257  0.0859  0.3513  0.9551
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)            4.0872523  0.2888562  14.150  < 2e-16 ***
## ranktenure track      -0.1476746  0.0819824  -1.801 0.072327 .
## ranktenured           -0.0973829  0.0662614  -1.470 0.142349
## ethnicitynot minority  0.1274458  0.0772887   1.649 0.099856 .
## gendermale             0.2101231  0.0516873   4.065 5.66e-05 ***
## languagenon-english   -0.2282894  0.1111305  -2.054 0.040530 *
## age                   -0.0089992  0.0031326  -2.873 0.004262 **
## cls_perc_eval          0.0052888  0.0015317   3.453 0.000607 ***
## cls_students           0.0004687  0.0003737   1.254 0.210384
## cls_levelupper         0.0606374  0.0575010   1.055 0.292200
## cls_creditsone credit  0.5061196  0.1149163   4.404 1.33e-05 ***
## bty_avg                0.0398629  0.0174780   2.281 0.023032 *
## pic_outfitnot formal  -0.1083227  0.0721711  -1.501 0.134080
## pic_colorcolor        -0.2190527  0.0711469  -3.079 0.002205 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4974 on 449 degrees of freedom
## Multiple R-squared:  0.187,  Adjusted R-squared:  0.1634
## F-statistic: 7.943 on 13 and 449 DF,  p-value: 2.336e-14
```

Coefficients and p-values changed, through only slightly, which suggests some collinearity between the `cls_profs` and other variables.

15. Using backward-selection and p-value as the selection criterion, determine the best model. 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 you settle on.

```r
m_backSel_pval <- lm(score ~ ethnicity + gender + language + age + cls_perc_eval
             + cls_credits + bty_avg
             + pic_color, data = evals)
summary(m_backSel_pval)
```

```
##
## Call:
## lm(formula = score ~ ethnicity + gender + language + age + cls_perc_eval +
##     cls_credits + bty_avg + pic_color, data = evals)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.85320 -0.32394  0.09984  0.37930  0.93610
```

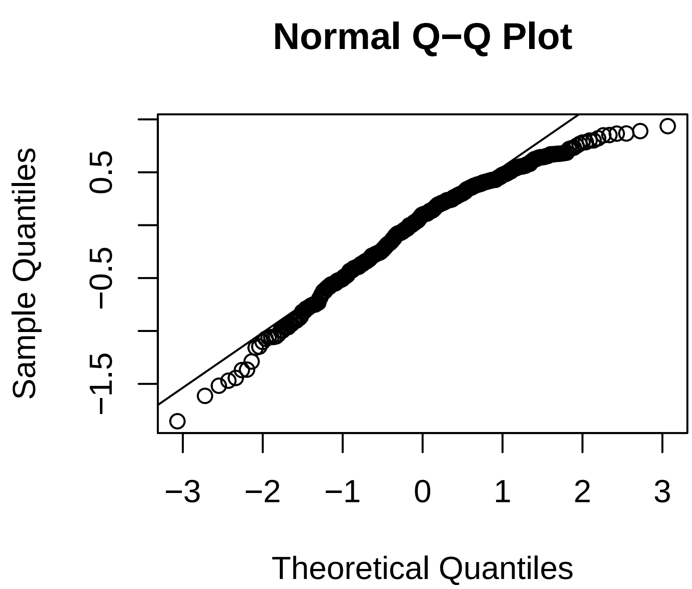
```
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)             3.771922   0.232053  16.255  < 2e-16 ***
## ethnicitynot minority   0.167872   0.075275   2.230  0.02623 *
## gendermale              0.207112   0.050135   4.131 4.30e-05 ***
## languagenon-english    -0.206178   0.103639  -1.989  0.04726 *
## age                    -0.006046   0.002612  -2.315  0.02108 *
## cls_perc_eval           0.004656   0.001435   3.244  0.00127 **
## cls_creditsone credit   0.505306   0.104119   4.853 1.67e-06 ***
## bty_avg                 0.051069   0.016934   3.016  0.00271 **
## pic_colorcolor         -0.190579   0.067351  -2.830  0.00487 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4992 on 454 degrees of freedom
## Multiple R-squared:  0.1722, Adjusted R-squared:  0.1576
## F-statistic:  11.8 on 8 and 454 DF,  p-value: 2.58e-15
```

```
coeffBackSel <- coefficients(m_backSel_pval)
```

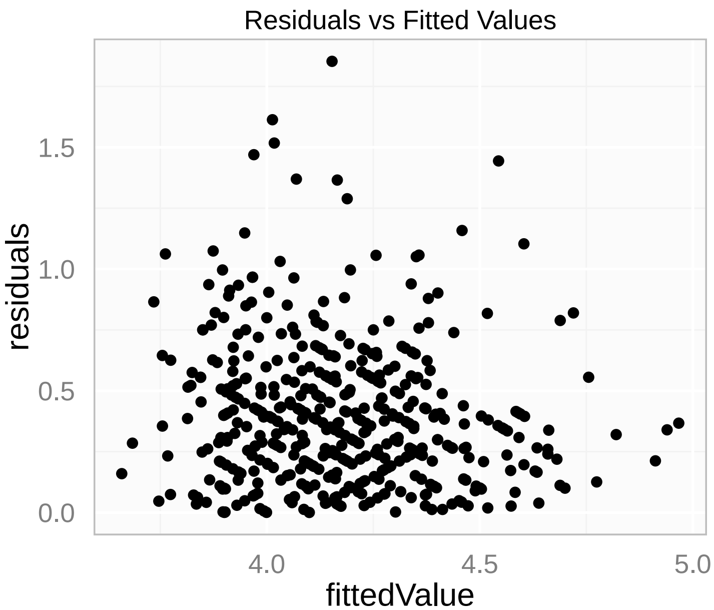
$$
\begin{aligned}
\widehat{score} =& 3.7719215 + 0.1678723 \text{ ethnicity}_{notminority} + 0.2071121 \text{ gender}_{male} + \\
& - 0.2061781 \text{ language}_{non-english} + -0.0060459 \text{ age} + 0.0046559 \text{ cls\_perc\_eval} + \\
& 0.5053062 \text{ cls\_credits}_{onecredit} + 0.0510693 \text{ bty\_avg} + -0.1905788 \text{ pic\_color}_{color}
\end{aligned}
$$

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

Normalness appears fine:



The constant variability of residuals is checked via a comparison to the fitted values. As shown in the following visualization, the data appears to satisfy the conditions for regression.



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?

    The new information (all courses they have taught) would introduce bias toward professors who have been at the universtiy longer whom would be represented more in sample.

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.

```
#summary(evals)
# Characteristics of a high scoring professor
vHigh <- c(1,      # Intercept
           1,      # Ethnicity: Non minority
           1,      # Gender: Male
           0,      # Language: Non-english
           29,     # Age
           100,    # Percent of Class Evaluations
           1,      # Credits: One Credit
           8.167,  # Average Beauty Score
           0)      # Pic Color: Color
# Compute predicted score
score <- coefficients(m_backSel_pval) %*% vHigh
# Exceeds the maximum
score
```

```
##          [,1]
## [1,] 5.359551
```

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

I would not be comfortable generalizing these conclusions to professors at other univesities, but could be used more generally at the University of Texas Austin. The population being sampled is specifically the UoTA professors. Inorder to generalize to any university, we would need to sample from all unversities in some way and would likely include a geographic factor in the model as well.