Example Report

Data Analysis Skills Week 3

1 Introduction

Student feedback is an essential source for evaluating teaching in higher education. Feedback can lead to the improvement of teaching approaches and the materials presented. However, studies into student feedback have suggested potential bias factors in the feedback given. Here, we shall examine data from a study into student feedback on professors from The University of Texas at Austin, where the students also provided information on the physical appearance of the teacher. Hence, the relationship between teaching scores and the attractiveness of the teacher will be examined.

Section 2 consists of an exploratory data analysis of teaching evaluation scores and explores the potential relationship between teaching scores and the perceived attractiveness of the teacher. Section 3 contains the results from fitting a linear regression model to the data, as well as the assessment of the model assumptions. Concluding remarks are given in Section 4.

2 Exploratory data analysis

Table 1 contains summary statistics on the teaching and beauty scores of the 463 teaching instructors from the University of Texas at Austin. First, if we look at teaching score, which is an average of the average teaching scores based on students' evaluations on a scale between 1 and 5, we see that the middle 50% of teaching scores lie between 3.8 and 4.6, with an average teaching score of 4.17. Secondly, if we look at the beauty score of the instructors, which is an average of the beauty scores from a panel of six students' scores on a scale between 1 and 10, we see that the middle 50% lies between 3.17 and 5.5, with an average beauty score of 4.42. There is also more variability in beauty scores, as seen from the standard deviation (SD), however, this may well be due to it being on a larger scale than the teaching scores (1 to 10 vs. 1 to 5).

Table 1: Summary statistics on teaching and beauty scores.

Variable	Missing	Complete	Mean	SD	Min.	1st Q.	Median	3rd Q.	Max.
score	0	1	4.17	0.54	2.30	3.80	4.30	4.6	5.00
bty_avg	0	1	4.42	1.53	1.67	3.17	4.33	5.5	8.17

Figure 1 displays a scatterplot of the teaching score against beauty score. Here, there appears to be a positive relationship between teaching and beauty scores. Hence, teaching score tends to increase with beauty score. However, as seen from Figure 1 and the observed correlation coefficient (0.187), it is only a weakly positive relationship. A linear regression model will now be fitted to assess the relationship between teaching and beauty scores.

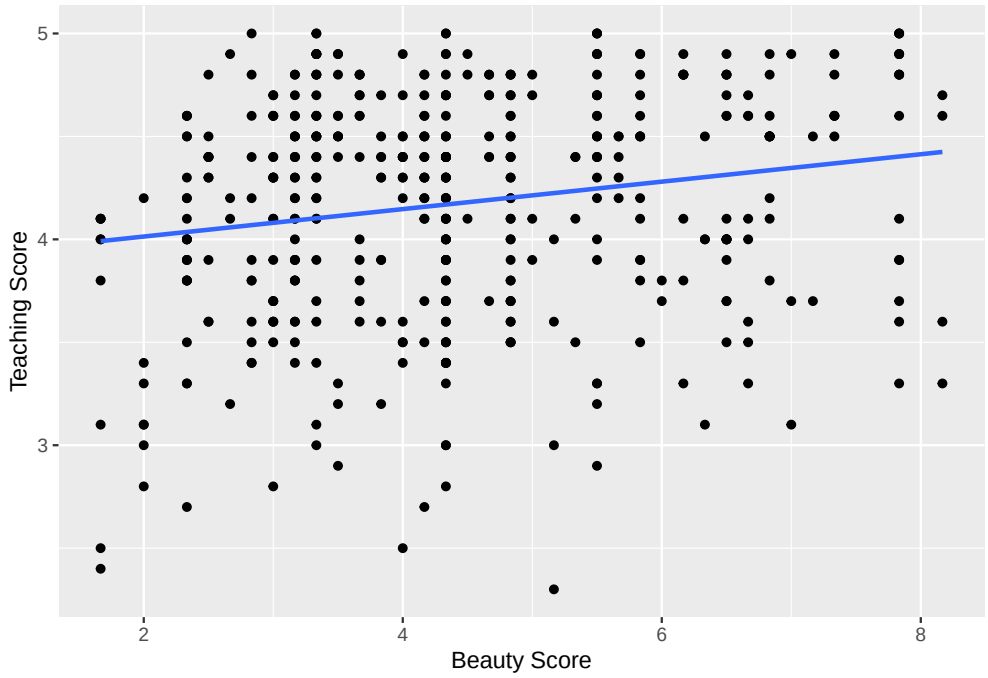

Figure 1: Relationship between teaching and beauty scores. The best-fitting line has been superimposed.

3 Formal data analysis

The linear regression model that will be fitted to the data is as follows:

$$y_i = \alpha + \beta x_i + \epsilon_i, \qquad \epsilon_i \sim N(0, \sigma^2),$$

where y_i and x_i denote the **teaching score** and **beauty score** of the i^{th} teaching instructor, respectively. The intercept of the linear regression line is denoted by α, while the slope is given by β. The i^{th} random error component is denoted by ϵ_i, which are normally distributed with mean zero and variance σ^2.

Table 2 displays the estimated intercept and slope parameters of the best-fitting line from the regression model. Hence, the best-fitting line is given as:

$$\widehat{\text{teaching score}} = 3.88 + 0.067 \cdot \text{beauty score}.$$

That is, for every one unit increase in beauty score, there is an associated increase in teaching score of, on average, 0.067. This agrees with the exploratory analysis carried out in Section 2, that is, the relationship between teaching scores and beauty scores is positive, if only weakly positive. Before concluding, the assumptions corresponding to the linear regression model need to assessed.

Table 2: Estimates of the intercept and slope from the fitted linear regression model.

term	estimate
intercept	3.880
bty_avg	0.067

Figure 2 shows scatterplots of the residuals against the explanatory variable beauty score, as well as the fitted values. Here we can see that there is no obvious pattern in the residuals, which appear to be evenly scattered above and below the zero line, and hence have mean zero. Also, the spread of the residuals is constant across all values of the explanatory variable and fitted values, and as such display no obvious change in variability. Hence, the random error component of the regression model satisfies the assumptions of having mean zero and constant variance. Now we need to examine whether they are normally distributed.

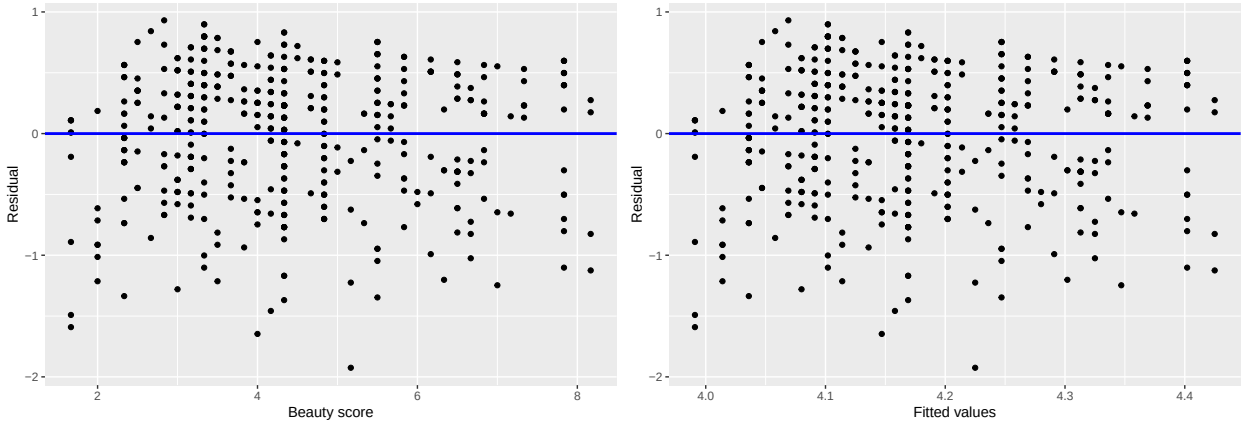

Figure 2: Scatterplots of the residuals against beauty score (left) and the fitted values (right).

Figure 3 displays a histogram of the residuals. Here, the histogram appears to be slightly left-skewed, however, it appears to be relatively bell-shaped and centred around zero. Hence, the assumption of normally distributed errors appears to hold for the fitted regression model.

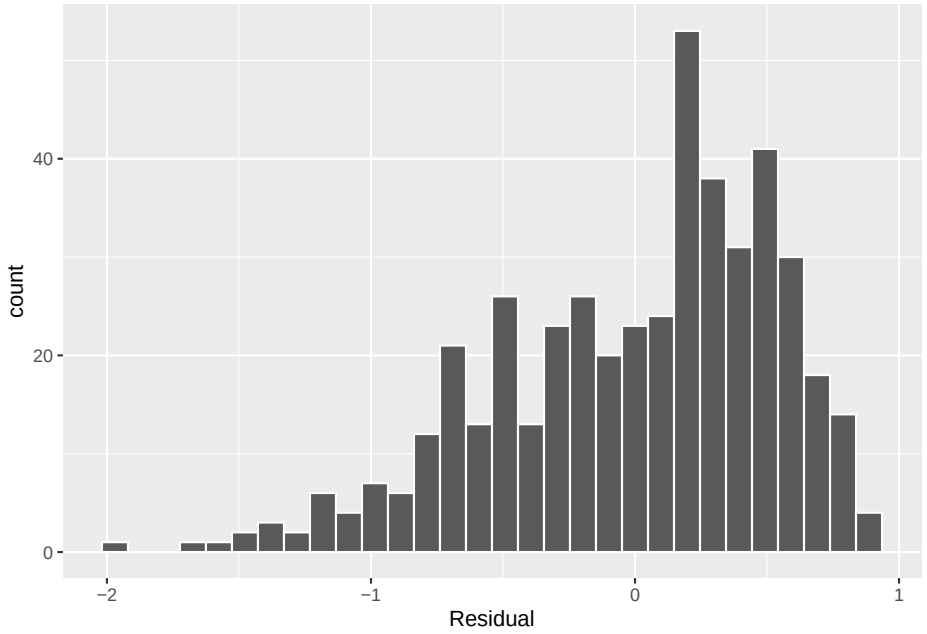

Figure 3: Histogram of the residuals.

4 Conclusions

The student feedback on professors from the University of Texas at Austin suggests that there may be potential for bias factors, such as the attractiveness of the teacher, when it comes to providing feedback. This is observed from the fitted regression model, where teaching score was found to increase with increasing beauty score. However, the relationship observed is only weakly positive (correlation coefficient of 0.187), with an associated increase in teaching score of, on average, 0.067, with every one unit increase in beauty score.