The Effect of Research Experience on Graduate Admissions

2383746W

1 Introduction

Admission to graduate programmes has become increasingly competitive, and many universities have addressed this issue by introducing additional entry requirements for applicants. In particular, institutions in the United States emphasize the importance of obtaining high scores in the "Graduate Record Examinations (GRE)", which is a standardised test that assesses graduates' academic skill set. In this report, GRE scores from 400 UCLA graduates were considered along with a record of prior research experience in order to investigate whether the latter has an effect on their chance of being admitted to graduate programmes.

In particular, this work presents numerical and graphical summaries to reveal the structure and underlying relationship between the response variable `Chance.of.Admit` and the explanatory variable `GRE.Score`, when taking into account the categorical input `research`. Two linear regression models are fit to the data to estimate the difference, on average, between the chance of admission of those students who have prior research experience, and those who do not. The model selection procedure is carried out to determine the preferred model.

2 Exploratory Data Analysis

In this section, we identify simple relationships between our variables of interest, before proceeding with more detailed formal analysis. The summary statistics of the chance of admission are presented below. For each group i.e. researchers ("Yes") and non-researchers ("No"), the corresponding figures are shown separately.

Table 1: Summary statistics on the chance of admission by research experience of 400 graduate applicants.

research	n	Mean	St.Dev	Min	Q1	Median	Q3	Max
No	181	0.6	0.1	0.34	0.57	0.65	0.72	0.89
Yes	219	0.8	0.1	0.36	0.73	0.80	0.90	0.97

Table 1 reveals that the dataset of interest contains 9.5% more graduates with research experience, than those without prior research exposure (219 compared to 181). Moreover, apart from equal standard deviations (0.1 for both groups), it is evident that the summaries of the research group for the chance of being admitted are consistently greater than the corresponding summaries of non-researchers.

The mean likelihood of gaining admission to a graduate programme is 0.8 compared to 0.6 for researchers and non-researchers respectively. There is a comparatively smaller difference in sample medians between the two groups, with 0.8 vs 0.65 for the same comparison. Interestingly, the range of data is higher for researchers (Max-Min=0.61) when compared to non-researchers (Max-Min=0.55). However, despite this result, the first quantile of the research group (0.73) is one per cent higher than the third quantile of the non-researcher group (0.72). Given the data, this provides the first evidence that research experience appears to positively affect the chance of admission to a graduate programme. The magnitude of this influence is approximately 15 to 20 per cent, depending on which measure of location is being taken into consideration.

Figure 1: Chance of admission by GRE score and research experience.

Figure 1 shows two different intercepts and slopes, consequently `research` interacts in different ways with `GRE.Score` for both researchers and non-researchers. Comparatively to the group of non-researchers the following can be noted:

- There are more observations associated with high GRE scores in the group of researchers.
- The associated effect of rising GRE scores seems to be more pronounced for researchers, i.e the chance of admission for those increases faster with GRE scores above 300.
- For low GRE scores (below 300) research experience appears to be irrelevant in improving the chance of admission.
- The correlation between GRE Score and the chance of admission appears to be higher for researchers since observations are less spread out around the fitted line.

We obtain the correlation by research experience where the first coefficient is calculated for the GRE score of researchers, while the second stands for non-researchers. Thus, it is of interest to compare the impact of GRE scores obtained by the two groups on the chance of admission.

Table 2: Correlation coefficients by research experience.

research	corr
No	0.5791
Yes	0.7993

As can be seen from the table above, the correlation between the GRE score and the chance of admission is approximately 20 percent higher for the research group, hence suggesting a stronger linear association, compared to the non-research category. Based on this result, it is more likely that using an interaction model leads to more accurate inferences because, compared to the parallel slopes model, it is able to capture the different trends within the two levels of the `research` variable.

3 Formal Data Analysis

The aim of this section is to set up a statistical model which describes our given data in the most accurate way. For this reason, we will compare the performance of the "interaction model" with that of the "parallel lines model" and conduct model selection based on the obtained results.

3.1 Interaction Model

As seen in the linear association plot (Figure 1), there is an interaction effect in our model. The effect of `GRE.score` here will depend on whether the score was achieved by researchers or non-researchers, that is, the effect of `GRE.score` on the admission chance will differ by research experience. The interaction model can be formally expressed as:

$$y_i = \alpha \ + \ \beta_{\text{gre}} \cdot \text{gre}_i \ + \ \beta_{\text{researchYes}} \cdot \mathbb{I}_{\text{researchYes}}(i) \ + \ \beta_{\text{gre, researchYes}} \cdot \text{gre}_i \cdot \mathbb{I}_{\text{researchYes}}(i) \ + \ \epsilon_i$$

where

- α is the intercept of the regression line for non-researchers;
- β_{gre} is the slope of the regression line for both researchers and non-researchers;
- gre_i is the GRE score of ith observation;
- $\beta_{\text{researchYes}}$ is the additional term added to α to get the intercept of the regression line for males; and
- $\mathbb{I}_{\text{researchYes}}(i)$ is an indicator function such that

$$\mathbb{I}_{\text{researchYes}}(i) = \begin{cases} 1 & \text{if the } i\text{th observation is researcher,} \\ 0 & \text{Otherwise.} \end{cases}$$

- $\beta_{\text{gre, researchYes}} \cdot \text{gre}_i \cdot \mathbb{I}_{\text{researchYes}}(i)$ corresponds to the interaction term.

After having defined the model specifications, we run the regression with the involved interaction term and acquire the associated estimates for the model coefficients. These are displayed in Table 3 below:

Table 3: Estimates of the parameters from the fitted interaction model.

term	estimate
intercept	-1.686
GRE.Score	0.008
researchYes	-0.755
GRE.Score:researchYes	0.003

Hence, after combining terms, the regression lines for researchers and non-researchers are given by:

$$\widehat{Chance.of.Admit}_i = \begin{cases} -1.686 + 0.008 \cdot GRE.Score_i & non\ researchers \\ -2.441 + 0.011 \cdot GRE.Score_i & researchers \end{cases}$$

It can be noted that the interaction model allows for different slopes and intercepts for researchers and non-researchers. The increment for every GRE score unit in the group of researchers is expected to lead to a 0.003 higher chance of admission, compared to the group of non-researchers. Since it was previously shown that for low GRE scores (below 300), the fitted line for researchers lies below those of non-researchers, a lower intercept for researchers is modelled to consider this difference (-2.441 compared to -1.686).

3.2 Parallel Lines Model

For simplicity reasons, it can sometimes be useful to consider a parallel lines model by removing the interaction term from the interaction model. This model implies that the slope of the relationship between `Chance.of.Admit` and `GRE.score` is the same for both researchers and non-researchers, and only the intercepts of the regression lines are different. Therefore, the equation for the parallel lines model is given by:

$$y_i = \alpha \ + \ \beta_{\text{gre}} \cdot \text{gre}_i \ + \ \beta_{\text{researchYes}} \cdot \mathbb{I}_{\text{researchYes}}(i) \ + \ \epsilon_i$$

We fit the parallel lines model to the data, obtaining the following estimates for the model coefficients:

Table 4: Estimates of the parameters from the fitted parallel lines model.

term	estimate
intercept	-2.156
GRE.Score	0.009
researchYes	0.038

Thus, after combining terms, the regression lines for researchers and non-researchers are:

$$\widehat{Chance.of.Admit}_i = \begin{cases} -2.156 + 0.009 \cdot GRE.Score_i & non\ researchers \\ -2.118 + 0.009 \cdot GRE.Score_i & researchers \end{cases}$$

By this model, both researchers and non-researchers have the same slope, i.e. the associated effect of GRE scores on the chance of admission is the same for both groups. Therefore, for every one point increase in `GRE.score`, there is a corresponding increase in the chance of admission of 0.009. Students with prior research experience have a higher intercept term.

3.3 Model Comparison and Confidence Intervals of Parameters

In the following figure, the fitted interaction and parallel lines models are shown next to each other. We particularly take note of the different slopes and intercepts between the two models.

Figure 2: Fitted regression lines for the interaction model (left) and the parallel lines model (right).

Tables 5 and 6 below show regression summaries for the interaction and the parallel lines models respectively. All parameters are significant in both models since the 95% confidence intervals do not include zero.

Table 5: Regression table with interaction effect (interaction)

term	estimate	std_error	statistic	p_value	lower_ci	upper_ci
intercept	-1.686	0.219	-7.707	0.000	-2.116	-1.256
GRE.Score	0.008	0.001	10.625	0.000	0.006	0.009
researchYes	-0.755	0.287	-2.632	0.009	-1.319	-0.191
GRE.Score:researchYes	0.003	0.001	2.766	0.006	0.001	0.004

Table 6: Regression table without interaction effect (parallel lines)

term	estimate	std_error	statistic	p_value	lower_ci	upper_ci
intercept	-2.156	0.139	-15.480	0	-2.429	-1.882
GRE.Score	0.009	0.000	20.079	0	0.008	0.010
researchYes	0.038	0.010	3.643	0	0.017	0.058

3.4 Model Selection

From table 7 below, we can see that the adjusted R-squared (`adj.r.squared`) is higher while at the same time the AIC and BIC are lower when using the interaction model over the parallel lines one. As a result, the interaction model achieves a better overall fit to the data and will be our model of choice because it describes the relationship between the response and explanatory variables more accurately. Allowing for different slopes seems necessary and was previously suggested by the differing correlation coefficients between the two levels of `research`.

Table 7: Model comparison values for two models

Model	adj.r.squared	AIC	BIC
Interaction	0.660	-848.112	-828.154
Parallel	0.654	-842.459	-826.494

The interaction model an adjusted R-squared value of 0.66, thus two-thirds of the variation in the response variable `Chance.of.Admit` is explained by the explanatory variables `GRE.Score` and `research`. This represents a moderate fit to the data and is a common result when the number of included covariates is relatively low (as is the case here).

3.5 Modelling Assumptions

Before proceeding with the interaction model, it is important to examine the validity of the modelling assumptions in our chosen model. These include that the residuals should be normally distributed, have a mean of zero and display a constant variance within the group of researchers and non-researchers across all observations. Formally, we test for:

$$E(\widehat{\epsilon_i}) = 0 \quad , \quad Var(\widehat{\epsilon_i}) = \sigma^2 \qquad s.t. \qquad \widehat{\epsilon_i} \sim N(0,\ \sigma^2)$$

In order to reveal systematic patterns, we produce scatterplots of the residuals against the explanatory variable `GRE.Score` as well as against the fitted values for the response variable `Chance.of.Admit`. The associated plots can be seen in Figure 3 below.

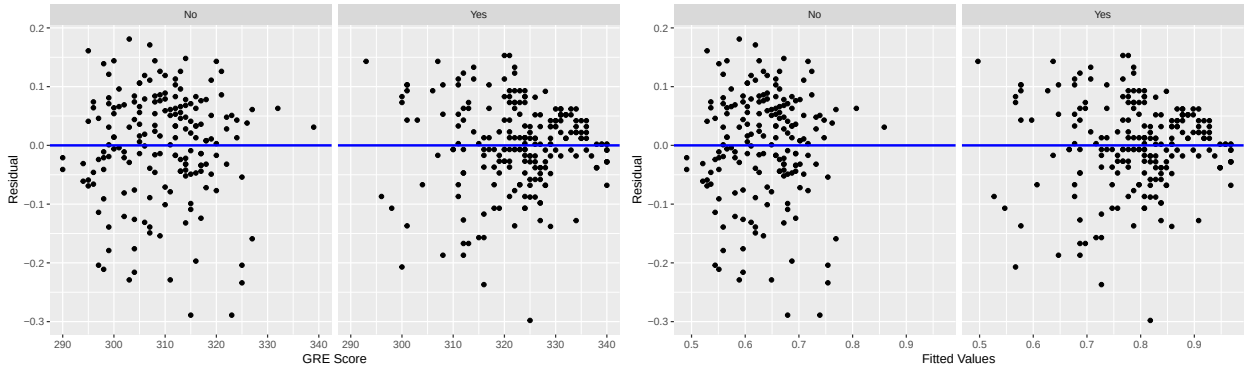

Figure 3: Scatterplots of the residuals against GRE score (left) and the fitted values (right) by research.

The scatterplots show a relatively even spread of the residuals above and below the zero line for each level in the variable research, and hence have mean zero. Moreover, since the spread across all explanatory and fitted values appears to be constant (on average), there is no sufficient evidence for the existence of a systematic pattern in the residuals. Based on this examination, we conclude that the random error component of our interaction model satisfies the assumptions of having a mean of zero and a constant variance.

Subsequently, we check for normality of the residuals.

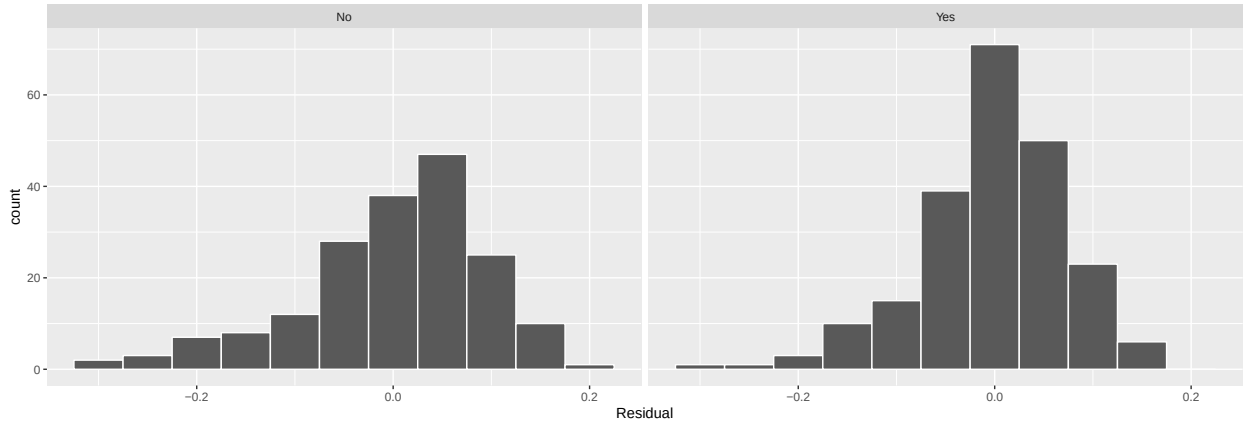

Figure 4: Histogram of the residual distribution.

Figure 4 displays the histograms of the residuals for both levels of research. Although the data seems to be slightly left-skewed for both groups of researchers, this does not represent a serious issue as the histograms appear to be relatively symmetrical and centred around zero.

4 Conclusion

To sum up, there was a significant effect of research on the chance of admission for both, the interaction and the parallel lines models. Model selection procedure revealed that the interaction model, which allows for different slopes and intercepts for the two levels of research, describes the relationship between the response and explanatory variables more accurately than the parallel lines model. Adjusted R-squared for the interaction model was 0.66, which despite giving a moderate fit to the data, is normal when considering a limited number of included covariates.

We have estimated that, on average, with every point increase in GRE score graduates with prior research experience have a 0.003 higher chance of being admitted to a graduate programme, as compared to students with no research experience. It was also observed that for GRE scores below 300, the fitted line for researchers lies below that of non-researchers, indicating the positive effect of research experience in particular for higher GRE scores.