OfS report on "grade inflation" in English universities

David Firth

Updated 7 January 2019

This file is written in *R Markdown.*

The small example from my original blog post (2019-01-02)

See https://statgeek.net/2019/01/02/office-for-students-report-on-grade-inflation/

Here I will attempt to use the OfS fixed-effects logistic regression method, for the artificial data that appears in my blog post.

First, set up the data:

```
provider <- factor(rep(c(rep("A", 2), rep("B", 2)), 2))
year <- factor(c(rep("2010-11", 4), rep("2016-17", 4)))
student_type <- factor(rep(c("h", "i"), 4))
n_firsts <- c(1000, 0, 500, 500, 1800, 0, 0, 500)
n_other <- c(0, 1000, 500, 500, 200, 0, 0, 1500)
total <- n_firsts + n_other
the_data <- data.frame(provider, year, student_type, n_firsts, n_other)
the_data
```

```
##   provider    year student_type n_firsts n_other
## 1        A 2010-11            h     1000       0
## 2        A 2010-11            i        0    1000
## 3        B 2010-11            h      500     500
## 4        B 2010-11            i      500     500
## 5        A 2016-17            h     1800     200
## 6        A 2016-17            i        0       0
## 7        B 2016-17            h        0       0
## 8        B 2016-17            i      500    1500
```

Now fit the logistic regression model as specified in Annex D (p42) of the OfS report.

```
outcome <- cbind(n_firsts, n_other)
the_model <- glm(outcome ~ provider + year + provider:year + student_type,
                 family = binomial, data = the_data)
```

Compute the fitted values (expected numbers of Firsts) from that model, and print those out:

```
expected_firsts <- total * fitted(the_model)
model_predictions <- data.frame(year, provider, student_type, expected_firsts)
model_predictions
```

```
##      year provider student_type expected_firsts
## 1 2010-11        A            h             750
## 2 2010-11        A            i             250
## 3 2010-11        B            h             750
```

1

```
## 4 2010-11          B          i              250
## 5 2016-17          A          h             1800
## 6 2016-17          A          i                0
## 7 2016-17          B          h                0
## 8 2016-17          B          i              500
```

Now get "predicted" counts, as described at point 5 of Annex C in the OfS report. First we make a new dataset in which the year is changed to "2010-11" throughout:

```
newdata <- the_data
newdata$year <- factor(rep("2010-11", 8))
```

And then we compute the predicted values based on graduation in year 2010-11 (as predicted probabilities multiplied by the binomial totals):

```
predicted_2010_11_counts <- predict(the_model, newdata, type = "response") * total
predicted_2010_11_counts
```

```
##    1    2    3    4    5    6    7    8
##  750  250  750  250 1500    0    0  500
```

The whole-sector "unexplained" Firsts in 2016-17, as defined at point 5 of Annex C in the OfS report, can now be counted:

```
sum((n_firsts - predicted_2010_11_counts)[year == "2016-17"])
```

```
## [1] 300
```

— that's 300 "unexplained" Firsts (or 7.5 percentage points, as a fraction of the total 4000 graduates here).

The better model ("SASA model") from blog post Part 2 (2019-01-07)

This is as described in Sec 2.1 of the "Part 2" blog post at https://statgeek.net/2019/01/07/part-2-further-comments-on-ofs-grade-inflation-report/

```
SASA_model <- glm(outcome ~ -1 + provider:student_type + provider:year,
                  family = binomial, data = the_data)
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

(The reported warning relates to the 100% probability of a First in University A in year 2010-11. That's correct, and in this instance the warning can be safely ignored.)

Then as above we compute the predicted values based on graduation in year 2010-11 (as predicted probabilities multiplied by the binomial totals):

```
predicted_2010_11_counts <- predict(SASA_model, newdata, type = "response") * total
```

The whole-sector "unexplained" Firsts in 2016-17 can now be counted as before:

```
sum((n_firsts - predicted_2010_11_counts)[year == "2016-17"])
```

```
## [1] -700
```

— so that's a measured *decrease* of 700 Firsts (which is 23.3% of the expected firsts based on 2010-11 data, or 17.5% of the total 4000 graduates here).