

# Regression Analysis of Higher Education Outcomes

Gilbert Watson

Monday, November 25th, 2013

## Contents

<b>1</b>	<b>The Data</b>	<b>1</b>
1.1	Source: . . . . .	1
1.2	Cleaning: . . . . .	1
1.3	Initial Model Formulation: . . . . .	2
1.4	Initial Model Results: . . . . .	2
<b>2</b>	<b>System Information</b>	<b>6</b>

## 1 The Data

### 1.1 Source:

Data from 2010 NCES survey

Institutional Population:

- a) Title IV participating
- b) US only
- c) Carnegie: Research, Doctoral, Baccalaureate categories
- d) Has Full-Time first time undergraduates
- e) At least one program not offered through distance education
- f) Only Public and Not for Profit Institutions

### 1.2 Cleaning:

Cleaning Steps required:

- a) Removal of observation indicators.
- b) Removal of all variables associated with the response (5 year graduate rates) except for the one of interest.
- c) Removal of for-profit institutions.
- d) Merged financial variables that were reported in two separate fields by public and private institutions. GASB and FASB standards are quite similar and it appeared appropriate to use them concurrently.
- e) Removed variables that were the same for all observations (like degree granting status).

- f) Made fee fields for in-state, in-district partitions equal to total cost in cases where institutions did not report (largely because they are private institutions and the partition has no applicability).
- g) Made graduate demographic percentages 0 where missing (because the institution has no graduate population). Variable behaves like an interaction term with variable "anyGrads."
- h) TO DO: Merge admission and yeild rates for intitutions that do not report gender breakdowns because they do not admit a gender, or because they have open admissions policies. Figure out how to make that distinction.
- i) Create derived percentage and different variables for all variables that apply to alleviate obvious instances of high correlation simply due to levels. Care was taken not to introduce any linear dependencies when making these variables. This is why it may appear as some are omitted. Thier inclusion would simply fully define another variable.
- j) Removed a small subset of variables that do not have any applicability to the whole set of observations (data about a specific cohort tracked by the NCES survey).
- k) Removed variables that cause linear dependencies due to the full partition of another variable.
- l) Replaced infinite values in derived variables with 0 because we often mean that there is no change when we divide by zero.

### 1.3 Initial Model Formulation:

Initial Model Formulation:

- a) Since we have so many variables, some sort of automated technique was required to build an initial model.
- b) First, we replace all missing values in the dataset with zero to eliminate reducing the number of observations, but only to get an initial, large model.
- c) We do forward, backwards, and sequential search for the best regression subset of up to 50 variables using the full model from this modified data. The reason we do this is because exhaustive search would take a prohibitively long time using this many predictors.
- d) We then use all predictors selected by any of these three search methods to build a full model on the original cleaned data (without missing data replaced).
- e) We then conduct the same sort of search as above for a reasonable set of predictors, this time searching for the best regression subset with a max of 20 variables.
- f) Using the all the predictors selected by the three methods, we are left with a model that can be exhaustively searched for the best subset (29 predictors). We are left with an initial model.

Should we allow for a higher number of variables in the second forward, backwards, and sequential search?

### 1.4 Initial Model Results:

Call:

```
lm(formula = DRVGR2011_RV.Graduation.rate...Bachelor.degree.within.5.years..total ~
., data = lastdata)
```

Residuals:

Min	1Q	Median	3Q	Max
-54.272	-4.044	0.383	4.177	79.297

Coefficients:

(Intercept)

DRVIC2011\_RV.Tuition.and.fees..2009.10

``HD2011.Institution.size.categoryUnder 1,000``

``HD2011.Geographic.regionMid East DE DC MD NJ NY PA``

``HD2011.Degree.of.urbanization..Urban.centric.locale.Rural: Distant``

``HD2011.Carnegie.Classification.2010..Undergraduate.ProfileFull-time four-year, more selective, higher

``HD2011.Carnegie.Classification.2010..Undergraduate.ProfileFull-time four-year, more selective, lower

``HD2011.Carnegie.Classification.2010..Undergraduate.ProfileFull-time four-year, selective, higher tra

``HD2011.Carnegie.Classification.2010..Undergraduate.ProfileFull-time four-year, selective, lower tran

``HD2011.Carnegie.Classification.2010..Undergraduate.ProfileMedium full-time four-year, inclusive``

``HD2011.Carnegie.Classification.2010..Size.and.SettingLarge four-year, primarily nonresidential``

``HD2011.Carnegie.Classification.2010..Size.and.SettingMedium four-year, primarily nonresidential``

``HD2011.Carnegie.Classification.2010..Size.and.SettingSmall four-year, primarily nonresidential``

``HD2011.Carnegie.Classification.2010..Size.and.SettingSmall four-year, primarily residential``

``HD2011.Carnegie.Classification.2000Other separate health profession schools``

DRVEF2011\_RV.Full.time.equivalent.fall.enrollment

DRVEF2011\_RV.Percent.of.undergraduate.enrollment.that.are.White

DRVEF2011\_RV.Percent.of.undergraduate.enrollment.that.are.women

EF2011D\_RV.Full.time.retention.rate..2011

DRVEF2011\_RV.Adult.age..25.64..enrollment..full.time.students

DRVEF2011\_RV.Adult.age..25.64..enrollment..full.time.undergraduate

DRVEF2011\_RV.Percent.of.undergraduate.enrollment.18.24

DRVEF2011\_RV.Percent.of.undergraduate.enrollment.over.65

DRVC2011\_RV.Bachelor.s.degree

SFA1011\_RV.Percent.of.full.time.first.time.undergraduates.receiving.Pell.grants

DRVF2011\_RV.Endowment.assets..year.end..per.FTE.enrollment..MERGE

percent.assistant.salary

(Intercept)

DRVIC2011\_RV.Tuition.and.fees..2009.10

``HD2011.Institution.size.categoryUnder 1,000``

``HD2011.Geographic.regionMid East DE DC MD NJ NY PA``

``HD2011.Degree.of.urbanization..Urban.centric.locale.Rural: Distant``

``HD2011.Carnegie.Classification.2010..Undergraduate.ProfileFull-time four-year, more selective, higher

``HD2011.Carnegie.Classification.2010..Undergraduate.ProfileFull-time four-year, more selective, lower

``HD2011.Carnegie.Classification.2010..Undergraduate.ProfileFull-time four-year, selective, higher tra

``HD2011.Carnegie.Classification.2010..Undergraduate.ProfileFull-time four-year, selective, lower tran

``HD2011.Carnegie.Classification.2010..Undergraduate.ProfileMedium full-time four-year, inclusive``

``HD2011.Carnegie.Classification.2010..Size.and.SettingLarge four-year, primarily nonresidential``

``HD2011.Carnegie.Classification.2010..Size.and.SettingMedium four-year, primarily nonresidential``

``HD2011.Carnegie.Classification.2010..Size.and.SettingSmall four-year, primarily nonresidential``

``HD2011.Carnegie.Classification.2010..Size.and.SettingSmall four-year, primarily residential``

``HD2011.Carnegie.Classification.2000Other separate health profession schools``

DRVEF2011\_RV.Full.time.equivalent.fall.enrollment

DRVEF2011\_RV.Percent.of.undergraduate.enrollment.that.are.White

DRVEF2011\_RV.Percent.of.undergraduate.enrollment.that.are.women

EF2011D\_RV.Full.time.retention.rate..2011  
 DRVEF2011\_RV.Adult.age..25.64..enrollment..full.time.students  
 DRVEF2011\_RV.Adult.age..25.64..enrollment..full.time.undergraduate  
 DRVEF2011\_RV.Percent.of.undergraduate.enrollment.18.24  
 DRVEF2011\_RV.Percent.of.undergraduate.enrollment.over.65  
 DRVC2011\_RV.Bachelor.s.degree  
 SFA1011\_RV.Percent.of.full.time.first.time.undergraduates.receiving.Pell.grants  
 DRVF2011\_RV.Endowment.assets..year.end..per.FTE.enrollment..MERGE  
 percent.assistant.salary

(Intercept)

DRVIC2011\_RV.Tuition.and.fees..2009.10  
 ``HD2011.Institution.size.categoryUnder 1,000``  
 ``HD2011.Geographic.regionMid East DE DC MD NJ NY PA``  
 ``HD2011.Degree.of.urbanization..Urban.centric.locale.Rural: Distant``  
 ``HD2011.Carnegie.Classification.2010..Undergraduate.ProfileFull-time four-year, more selective, higher tra``  
 ``HD2011.Carnegie.Classification.2010..Undergraduate.ProfileFull-time four-year, more selective, lower tra``  
 ``HD2011.Carnegie.Classification.2010..Undergraduate.ProfileFull-time four-year, selective, higher tra``  
 ``HD2011.Carnegie.Classification.2010..Undergraduate.ProfileFull-time four-year, selective, lower tran``  
 ``HD2011.Carnegie.Classification.2010..Undergraduate.ProfileMedium full-time four-year, inclusive``  
 ``HD2011.Carnegie.Classification.2010..Size.and.SettingLarge four-year, primarily nonresidential``  
 ``HD2011.Carnegie.Classification.2010..Size.and.SettingMedium four-year, primarily nonresidential``  
 ``HD2011.Carnegie.Classification.2010..Size.and.SettingSmall four-year, primarily nonresidential``  
 ``HD2011.Carnegie.Classification.2010..Size.and.SettingSmall four-year, primarily residential``  
 ``HD2011.Carnegie.Classification.2000Other separate health profession schools``  
 DRVEF2011\_RV.Full.time.equivalent.fall.enrollment  
 DRVEF2011\_RV.Percent.of.undergraduate.enrollment.that.are.White  
 DRVEF2011\_RV.Percent.of.undergraduate.enrollment.that.are.women  
 EF2011D\_RV.Full.time.retention.rate..2011  
 DRVEF2011\_RV.Adult.age..25.64..enrollment..full.time.students  
 DRVEF2011\_RV.Adult.age..25.64..enrollment..full.time.undergraduate  
 DRVEF2011\_RV.Percent.of.undergraduate.enrollment.18.24  
 DRVEF2011\_RV.Percent.of.undergraduate.enrollment.over.65  
 DRVC2011\_RV.Bachelor.s.degree  
 SFA1011\_RV.Percent.of.full.time.first.time.undergraduates.receiving.Pell.grants  
 DRVF2011\_RV.Endowment.assets..year.end..per.FTE.enrollment..MERGE  
 percent.assistant.salary

(Intercept)

DRVIC2011\_RV.Tuition.and.fees..2009.10  
 ``HD2011.Institution.size.categoryUnder 1,000``  
 ``HD2011.Geographic.regionMid East DE DC MD NJ NY PA``  
 ``HD2011.Degree.of.urbanization..Urban.centric.locale.Rural: Distant``  
 ``HD2011.Carnegie.Classification.2010..Undergraduate.ProfileFull-time four-year, more selective, higher tra``  
 ``HD2011.Carnegie.Classification.2010..Undergraduate.ProfileFull-time four-year, more selective, lower tra``  
 ``HD2011.Carnegie.Classification.2010..Undergraduate.ProfileFull-time four-year, selective, higher tra``  
 ``HD2011.Carnegie.Classification.2010..Undergraduate.ProfileFull-time four-year, selective, lower tran``  
 ``HD2011.Carnegie.Classification.2010..Undergraduate.ProfileMedium full-time four-year, inclusive``  
 ``HD2011.Carnegie.Classification.2010..Size.and.SettingLarge four-year, primarily nonresidential``  
 ``HD2011.Carnegie.Classification.2010..Size.and.SettingMedium four-year, primarily nonresidential``  
 ``HD2011.Carnegie.Classification.2010..Size.and.SettingSmall four-year, primarily nonresidential``

```

``HD2011.Carnegie.Classification.2010..Size.and.SettingSmall four-year, primarily residential``
``HD2011.Carnegie.Classification.2000Other separate health profession schools``
DRVEF2011_RV.Full.time.equivalent.fall.enrollment
DRVEF2011_RV.Percent.of.undergraduate.enrollment.that.are.White
DRVEF2011_RV.Percent.of.undergraduate.enrollment.that.are.women
EF2011D_RV.Full.time.retention.rate..2011
DRVEF2011_RV.Adult.age..25.64..enrollment..full.time.students
DRVEF2011_RV.Adult.age..25.64..enrollment..full.time.undergraduate
DRVEF2011_RV.Percent.of.undergraduate.enrollment.18.24
DRVEF2011_RV.Percent.of.undergraduate.enrollment.over.65
DRVC2011_RV.Bachelor.s.degree
SFA1011_RV.Percent.of.full.time.first.time.undergraduates.receiving.Pell.grants
DRVF2011_RV.Endowment.assets..year.end..per.FTE.enrollment..MERGE
percent.assistant.salary

```

(Intercept)

```

DRVIC2011_RV.Tuition.and.fees..2009.10
``HD2011.Institution.size.categoryUnder 1,000``
``HD2011.Geographic.regionMid East DE DC MD NJ NY PA``
``HD2011.Degree.of.urbanization..Urban.centric.locale.Rural: Distant``
``HD2011.Carnegie.Classification.2010..Undergraduate.ProfileFull-time four-year, more selective, higher``
``HD2011.Carnegie.Classification.2010..Undergraduate.ProfileFull-time four-year, more selective, lower``
``HD2011.Carnegie.Classification.2010..Undergraduate.ProfileFull-time four-year, selective, higher tra``
``HD2011.Carnegie.Classification.2010..Undergraduate.ProfileFull-time four-year, selective, lower tran``
``HD2011.Carnegie.Classification.2010..Undergraduate.ProfileMedium full-time four-year, inclusive``
``HD2011.Carnegie.Classification.2010..Size.and.SettingLarge four-year, primarily nonresidential``
``HD2011.Carnegie.Classification.2010..Size.and.SettingMedium four-year, primarily nonresidential``
``HD2011.Carnegie.Classification.2010..Size.and.SettingSmall four-year, primarily nonresidential``
``HD2011.Carnegie.Classification.2010..Size.and.SettingSmall four-year, primarily residential``
``HD2011.Carnegie.Classification.2000Other separate health profession schools``
DRVEF2011_RV.Full.time.equivalent.fall.enrollment
DRVEF2011_RV.Percent.of.undergraduate.enrollment.that.are.White
DRVEF2011_RV.Percent.of.undergraduate.enrollment.that.are.women
EF2011D_RV.Full.time.retention.rate..2011
DRVEF2011_RV.Adult.age..25.64..enrollment..full.time.students
DRVEF2011_RV.Adult.age..25.64..enrollment..full.time.undergraduate
DRVEF2011_RV.Percent.of.undergraduate.enrollment.18.24
DRVEF2011_RV.Percent.of.undergraduate.enrollment.over.65
DRVC2011_RV.Bachelor.s.degree
SFA1011_RV.Percent.of.full.time.first.time.undergraduates.receiving.Pell.grants
DRVF2011_RV.Endowment.assets..year.end..per.FTE.enrollment..MERGE
percent.assistant.salary
---

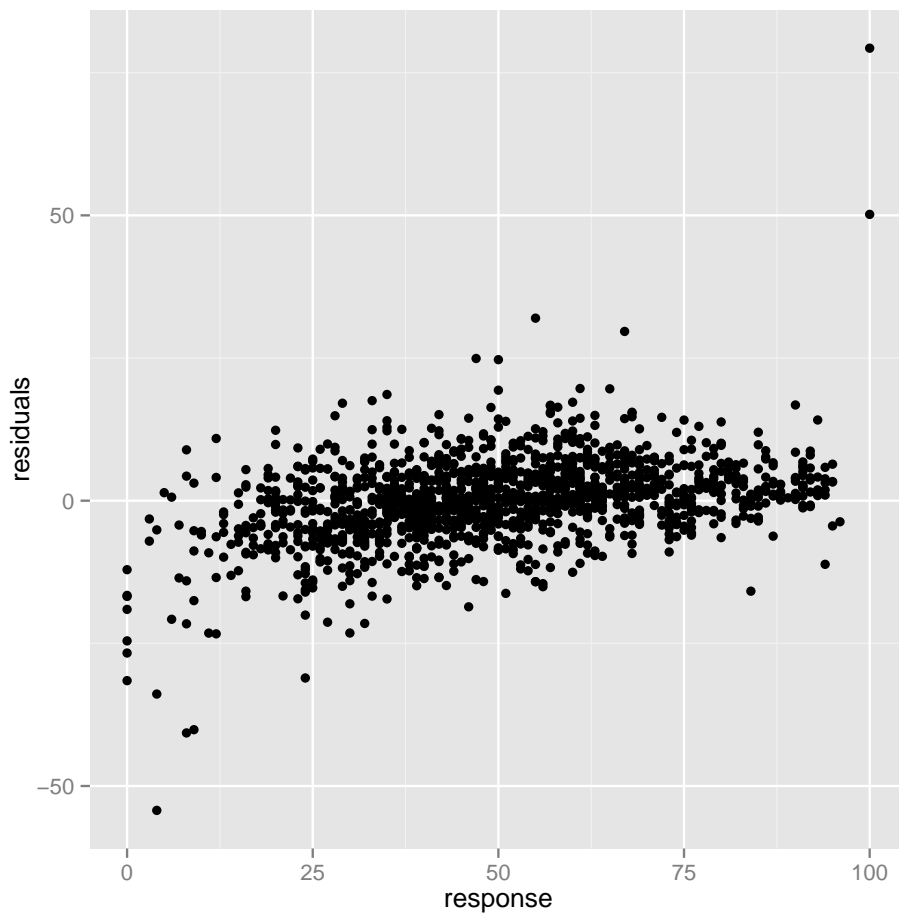
```

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7.814 on 1318 degrees of freedom

Multiple R-squared: 0.8549, Adjusted R-squared: 0.852

F-statistic: 298.6 on 26 and 1318 DF, p-value: < 2.2e-16



## 2 System Information

```
> sessionInfo();
```

```
R version 3.0.1 (2013-05-16)
```

```
Platform: x86_64-apple-darwin10.8.0 (64-bit)
```

```
locale:
```

```
[1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
```

```
attached base packages:
```

```
[1] stats      graphics  grDevices  utils      datasets  methods    base
```

```
other attached packages:
```

```
[1] ggplot2_0.9.3.1 xtable_1.7-1    leaps_2.9      MASS_7.3-26
```

```
loaded via a namespace (and not attached):
```

```
[1] colorspace_1.2-2  dichromat_2.0-0    digest_0.6.3      grid_3.0.1  
[5] gtable_0.1.2      labeling_0.2       munsell_0.4.2     plyr_1.8
```

```
[9] proto_0.3-10      RColorBrewer_1.0-5 reshape2_1.2.2    scales_0.2.3
[13] stringr_0.6.2     tools_3.0.1
```