**Model Selection in R for College Graduate Earnings**

Yimeng Wang

Rutgers University

4/17/2020

**Abstract**

The goal of this project was to analyze the data from College Scorecard on the earnings and debt of different college graduates to discover different relationships among the variables. Model selection techniques k-fold cross validation and stepwise selection were used to select a best fit linear model for the data to predict earnings and debt based off of other variables.

**Materials and Methods**

This project studied the data collected by the U.S. Department of Education titled “Most Recent Data by Field of Study” that was most recently updated on March 30th, 2020. The data contains the following variables:

* UNITID: University/Institution ID
* OPEID6: Six-digit OPEID, representing the main campus of a multi-branch institution
* INSTNM: Institution Name
* CONTROL: Type of Institution (i.e., public, private, etc.)
* CIPCODE: First four digits of the CIP code
* CIPDESC: Description of the major associated with CIP code
* CREDLEV: Numerical value associated with the degree level
* CREDDESC: Degree level
* DEBTMEDIAN: Median cumulative federal loan debt of federal borrowers by field of study
* DEBTPAYMENT10YR: Estimated monthly payment
* DEBTMEAN: Mean cumulative federal loan debt of federal borrowers by field of study
* TITLEIVCOUNT: Students due to postsecondary enrollment, military service, or death
* EARNINGSCOUNT: Students due to postsecondary enrollment, military service, or death, who were working in the measurement year and whose earnings are included
* MD\_EARN\_WNE: One-year post-completion earning
* IPEDSCOUNT1: Number of recognized postsecondary credentials conferred by field of study for the first award year
* IPEDSCOUNT2: Number of recognized postsecondary credentials conferred by field of study for the second award year

The data was first cleaned by removing all rows with missing variables and renamed some variables and created some new ones based on the data. For this project the focus is on the following variables.

* SchoolType (CONTROL) – Categorial variable that indicates the type of school.
  1. Public
  2. Private, non-profit
  3. Private, for profit
  4. Foreign
* Field – Categorial variable created using CIPCODE and CIPDESC, generalized field of study based on major given in the original data
  1. Humanities
  2. SocialSciences
  3. NaturalSciences
  4. AppliedSciences
  5. Others
* DegreeLevel (CREDLEV) – Numeric integer variable associated with degree level
  1. 1 – Undergraduate Certificate or Diploma
  2. 2 – Associate’s Degree
  3. 3 – Bachelor’s Degree
  4. 4 – Post-baccalaureate Certificate
  5. 5 – Master’s Degree
  6. 6 – Doctoral Degree
  7. 7 – First Professional Degree
  8. 8 – Graduate/Professional Certificate
* DebtMonthlyPaym (DEBTPAYMENT10YR) – Numeric, monthly payment of debt
* DebtMean (DEBTMEAN) – Numeric, mean of total debt by field of study
* Earnings (MD\_EARN\_WNE) – Numeric, one-year post-completion earning

Three different linear models were created for Earnings. The full model includes all variables mentioned above. The second model includes only SchoolType, Field, and DegreeLevel. The third model is most minimalistic and includes only Field and DegreeLevel.

lmod1 <- lm(Earnings ~ SchoolType + Field + DegreeLevel + DebtMonthlyPaym + DebtMean, data = college)

lmod2 <- lm(Earnings ~ SchoolType + Field + DegreeLevel, data = college)

lmod3 <- lm(Earnings ~ Field + DegreeLevel, data = college)

10-fold cross validation was used in order to choose between the three models to find out which is a best fit for predicting Earnings. First the data is split into 10 groups. Each group is then divided into 2 sets, the training set which is used to train the model, and the testing or validation set which is used to test the model by estimating the prediction error. Lastly the results were obtained from the testing fit by summarizing the skill of the model using the sample of model evaluation scores.

require(caret)

# Define training control

set.seed(13245)

train.control <- trainControl(method = "cv", number = 10)

# Train the model

model\_1 <- train (Earnings ~ SchoolType + Field + DegreeLevel + DebtMonthlyPaym + DebtMean,

data = college, method = "lm", trControl = train.control, na.action=na.exclude)

# Summarize the results

print(model\_1)

set.seed(14235)

train.control <- trainControl(method = "cv", number = 10)

model\_2 <- train (Earnings ~ SchoolType + Field + DegreeLevel,

data = college, method = "lm", trControl = train.control)

print(model\_2)

set.seed(13235)

train.control <- trainControl(method = "cv", number = 10)

model\_3 <- train (Earnings ~ Field + DegreeLevel,

data = college, method = "lm", trControl = train.control, na.action=na.exclude)

print(model\_3)

Then stepwise regression is used to confirm the best combination of variables to use to build a best performing model for Earnings. Stepwise regression works to iteratively add or remove predictors from the full predictive model in order to select the best subset of variables that will give the best model. Forward selection, backwards selection, and stepwise selection are all used. Forward selection starts with zero predictors, then iteratively adds the predictor that gives the highest p-value, which means it contributes most to the explanatory power of the model and stops when no other predictors are statistically significant in involvement. Backward selection starts with the full model and iteratively removes the predictor with the least contribution until all the variables left are statistically significant to the model. And stepwise selection starts with predictors, then iteratively adds the next best predictor, but after each addition it removes any previously added variables that are no longer useful to the model after this most recent change.

The selection process for the variable DebtMean was also the same. A 10-fold cross validation and all three stepwise regression was performed on the potential models. The three models used were the full model includes all variables mentioned above. The second and third model includes the same explanatory variables. 10-fold validation and the three methods of stepwise regression are used on the following three linear models for DebtMean.

lmod4 <- lm(DebtMean ~ SchoolType + Field + DegreeLevel + DebtMonthlyPaym + Earnings, data = college)

lmod5 <- lm(DebtMean ~ SchoolType + Field + DegreeLevel, data = college)

lmod6 <- lm(DebtMean ~ Field + DegreeLevel, data = college)

**Results**

***Earnings***

Summary of linear models 1-3 for Earnings gives all predictors p-value less than 0.05.

Cross-validation results for Model 1 (full model) is:

RMSE Rsquared MAE

13697.63 0.4862073 9417.737

Cross-validation results for Model 2 is:

RMSE Rsquared MAE

14022.68 0.4779141 9579.779

Cross-validation results for Model 3 is:

RMSE Rsquared MAE

14187.28 0.4657353 9700.768

Forward selection results:

Parameter Estimates

-----------------------------------------------------------------------------------------------------------------------

model Beta Std. Error Std. Beta t Sig lower upper

-----------------------------------------------------------------------------------------------------------------------

(Intercept) 15460.635 2788.362 5.545 0.000 9995.384 20925.886

SchoolTypePrivate, for-profit 11280.348 2777.566 0.255 4.061 0.000 5836.258 16724.438

SchoolTypePrivate, nonprofit 17507.740 2775.818 0.401 6.307 0.000 12067.076 22948.404

SchoolTypePublic 16919.145 2779.014 0.443 6.088 0.000 11472.217 22366.074

FieldHumanities -22098.919 250.672 -0.524 -88.159 0.000 -22590.241 -21607.597

FieldNaturalSciences -23016.627 370.096 -0.266 -62.191 0.000 -23742.023 -22291.232

FieldOthers -10960.841 237.876 -0.287 -46.078 0.000 -11427.083 -10494.599

FieldSocialSciences -24102.867 322.990 -0.344 -74.624 0.000 -24735.934 -23469.800

DegreeLevel 6613.849 72.202 0.479 91.603 0.000 6472.332 6755.366

DebtMonthlyPaym -12.465 1.727 -0.196 -7.216 0.000 -15.851 -9.080

DebtMean 0.252 0.021 0.325 11.760 0.000 0.210 0.294

-----------------------------------------------------------------------------------------------------------------------

Backward selection results:

Parameter Estimates

-----------------------------------------------------------------------------------------------------------------------

model Beta Std. Error Std. Beta t Sig lower upper

-----------------------------------------------------------------------------------------------------------------------

(Intercept) 15460.635 2788.362 5.545 0.000 9995.384 20925.886

SchoolTypePrivate, for-profit 11280.348 2777.566 0.255 4.061 0.000 5836.258 16724.438

SchoolTypePrivate, nonprofit 17507.740 2775.818 0.401 6.307 0.000 12067.076 22948.404

SchoolTypePublic 16919.145 2779.014 0.443 6.088 0.000 11472.217 22366.074

FieldHumanities -22098.919 250.672 -0.524 -88.159 0.000 -22590.241 -21607.597

FieldNaturalSciences -23016.627 370.096 -0.266 -62.191 0.000 -23742.023 -22291.232

FieldOthers -10960.841 237.876 -0.287 -46.078 0.000 -11427.083 -10494.599

FieldSocialSciences -24102.867 322.990 -0.344 -74.624 0.000 -24735.934 -23469.800

DegreeLevel 6613.849 72.202 0.479 91.603 0.000 6472.332 6755.366

DebtMonthlyPaym -12.465 1.727 -0.196 -7.216 0.000 -15.851 -9.080

DebtMean 0.252 0.021 0.325 11.760 0.000 0.210 0.294

-----------------------------------------------------------------------------------------------------------------------

[1] "No variables have been removed from the model."

Stepwise selection results:

Parameter Estimates

-----------------------------------------------------------------------------------------------------------------------

model Beta Std. Error Std. Beta t Sig lower upper

-----------------------------------------------------------------------------------------------------------------------

(Intercept) 15460.635 2788.362 5.545 0.000 9995.384 20925.886

SchoolTypePrivate, for-profit 11280.348 2777.566 0.255 4.061 0.000 5836.258 16724.438

SchoolTypePrivate, nonprofit 17507.740 2775.818 0.401 6.307 0.000 12067.076 22948.404

SchoolTypePublic 16919.145 2779.014 0.443 6.088 0.000 11472.217 22366.074

FieldHumanities -22098.919 250.672 -0.524 -88.159 0.000 -22590.241 -21607.597

FieldNaturalSciences -23016.627 370.096 -0.266 -62.191 0.000 -23742.023 -22291.232

FieldOthers -10960.841 237.876 -0.287 -46.078 0.000 -11427.083 -10494.599

FieldSocialSciences -24102.867 322.990 -0.344 -74.624 0.000 -24735.934 -23469.800

DegreeLevel 6613.849 72.202 0.479 91.603 0.000 6472.332 6755.366

DebtMonthlyPaym -12.465 1.727 -0.196 -7.216 0.000 -15.851 -9.080

DebtMean 0.252 0.021 0.325 11.760 0.000 0.210 0.294

-----------------------------------------------------------------------------------------------------------------------

***Debt***

Summary of linear models 4-6 for DebtMean gives all predictors p-value less than 0.05.

Cross-validation results for Model 4 (full model) is:

RMSE Rsquared MAE

3152.213 0.9834025 1876.743

Cross-validation results for Model 5 is:

RMSE Rsquared MAE

17491.72 0.4857304 8846.212

Cross-validation results for Model 6 is:

RMSE Rsquared MAE

18106.32 0.4490754 9715.292

Forward selection results:

Parameter Estimates

---------------------------------------------------------------------------------------------------------------------

model Beta Std. Error Std. Beta t Sig lower upper

---------------------------------------------------------------------------------------------------------------------

(Intercept) -6578.852 642.618 -10.238 0.000 -7838.397 -5319.307

SchoolTypePrivate, for-profit 8085.754 639.590 0.141 12.642 0.000 6832.144 9339.363

SchoolTypePrivate, nonprofit 7771.476 639.463 0.138 12.153 0.000 6518.114 9024.838

SchoolTypePublic 7954.386 640.126 0.161 12.426 0.000 6699.725 9209.047

DegreeLevel 649.971 17.998 0.036 36.113 0.000 614.694 685.249

DebtMonthlyPaym 79.308 0.073 0.965 1092.998 0.000 79.166 79.450

Earnings 0.013 0.001 0.010 11.760 0.000 0.011 0.016

FieldHumanities 321.323 63.056 0.006 5.096 0.000 197.732 444.914

FieldNaturalSciences -43.984 89.314 0.000 -0.492 0.622 -219.042 131.074

FieldOthers 252.305 56.263 0.005 4.484 0.000 142.028 362.582

FieldSocialSciences 350.972 79.394 0.004 4.421 0.000 195.358 506.586

---------------------------------------------------------------------------------------------------------------------

Backward selection results:

Parameter Estimates

---------------------------------------------------------------------------------------------------------------------

model Beta Std. Error Std. Beta t Sig lower upper

---------------------------------------------------------------------------------------------------------------------

(Intercept) -6578.852 642.618 -10.238 0.000 -7838.397 -5319.307

SchoolTypePrivate, for-profit 8085.754 639.590 0.141 12.642 0.000 6832.144 9339.363

SchoolTypePrivate, nonprofit 7771.476 639.463 0.138 12.153 0.000 6518.114 9024.838

SchoolTypePublic 7954.386 640.126 0.161 12.426 0.000 6699.725 9209.047

FieldHumanities 321.323 63.056 0.006 5.096 0.000 197.732 444.914

FieldNaturalSciences -43.984 89.314 0.000 -0.492 0.622 -219.042 131.074

FieldOthers 252.305 56.263 0.005 4.484 0.000 142.028 362.582

FieldSocialSciences 350.972 79.394 0.004 4.421 0.000 195.358 506.586

DegreeLevel 649.971 17.998 0.036 36.113 0.000 614.694 685.249

DebtMonthlyPaym 79.308 0.073 0.965 1092.998 0.000 79.166 79.450

Earnings 0.013 0.001 0.010 11.760 0.000 0.011 0.016

---------------------------------------------------------------------------------------------------------------------

[1] "No variables have been removed from the model."

Stepwise selection results:

Parameter Estimates

---------------------------------------------------------------------------------------------------------------------

model Beta Std. Error Std. Beta t Sig lower upper

---------------------------------------------------------------------------------------------------------------------

(Intercept) -6578.852 642.618 -10.238 0.000 -7838.397 -5319.307

SchoolTypePrivate, for-profit 8085.754 639.590 0.141 12.642 0.000 6832.144 9339.363

SchoolTypePrivate, nonprofit 7771.476 639.463 0.138 12.153 0.000 6518.114 9024.838

SchoolTypePublic 7954.386 640.126 0.161 12.426 0.000 6699.725 9209.047

DegreeLevel 649.971 17.998 0.036 36.113 0.000 614.694 685.249

DebtMonthlyPaym 79.308 0.073 0.965 1092.998 0.000 79.166 79.450

Earnings 0.013 0.001 0.010 11.760 0.000 0.011 0.016

FieldHumanities 321.323 63.056 0.006 5.096 0.000 197.732 444.914

FieldNaturalSciences -43.984 89.314 0.000 -0.492 0.622 -219.042 131.074

FieldOthers 252.305 56.263 0.005 4.484 0.000 142.028 362.582

FieldSocialSciences 350.972 79.394 0.004 4.421 0.000 195.358 506.586

---------------------------------------------------------------------------------------------------------------------

**Discussion**

For Earnings, summary results on the full model suggests that all predictors are significant and should be included in the model. Summary also returned the highest R-squared value for this full model compared to the other two with less predictors.

Cross-validation results for Model 1 (full model) gives R-squared value of 48.62% with root mean squared error of 13697.63 and mean absolute error of 9417.74. While this model result is not ideal since it has less than 0.5 explanatory power, it is still a better model than Model 2 and Model 3 since models 2-3 had lower R-squared values and higher RMSE and MAE.

Forward selection added the predictors in order of SchoolType-Private-for-profit, SchoolType-Private-nonprofit, SchoolType-Public, FieldHumanities, FieldNaturalSciences, FieldOthers, FieldSocialSciences, DegreeLeve, DebtMonthlyPaym, and DebtMean. Backward selection did not see to remove any predictors. Stepwise selection also gave the same results. All variables have an effect on Earnings. This means the best model for Earnings out of the three is the full model.

Earnings = SchoolTypePrivate\*1.128e04 + SchoolTypePrivateNonprofit\*1.751e4

+ SchoolTypePublic\*1.692e4 + FieldHumanities\*-2.210e4

+ FieldNaturalSciences\*-2.302e4 + FieldOthers\* -1.096e4

+ FieldSocialSciences\*-2.410e4 + DegreeLevel\*6.614e3

+ DebtMonthlyPaym\*-1.247e1 + DebtMean\*2.516e-1

For DebtMean, summary results on the full model suggests that all predictors are significant and should be included in the model. Summary also returned the highest R-squared value for this full model compared to the other two with less predictors.

Cross-validation results for Model 4 (full model) gives R-squared value of 98.34% with root mean squared error of 3125.21 and mean absolute error of 1876.73. Models 4 and 5 both gave significantly less R-squared values, both below 50%. While 4 and 5’s root mean square error values were a lot smaller than Model 4, Model 4 had much lower mean absolute error. Therefore, it seems the best model for DebtMean is also the full model.

Forward selection, backward selection, and stepwise selection all agreed to preserve the full model. The three methods added or eliminated the variables in slightly different order, but all included all the predictors. All variables have an effect in predicting the mean debt. So the best model for DebtMean out of the three is the full model.

DebtMean = SchoolType-Private-for-profit\*8.086e3

+ SchoolType-Private-nonprofit\*7.771e3 + SchoolTypePublic\*7.954e3

+ FieldHumanities\* 3.213e2 + FieldNaturalSciences\*-4.398e1

+ FieldOthers\*2.523e2 + FieldSocialSciences\*3.510e2 + DegreeLevel\*6.500e2

+ DebtMonthlyPaym\*7.931e1 + Earnings\*339e-2

**R-Program Code and Output**

> library(dplyr)

>

> college\_orig <- read.csv("Most-Recent-Field-Data-Elements.csv")

> college1 <- college\_orig[,c(4, 6, 8:9, 13:14, 17)]

> names(college1) <- c("SchoolType", "FieldCode", "Field", "DegreeLevel", "DebtMonthlyPaym", "DebtMean", "Earnings")

>

> college2<-college1[!(college1$Earnings=="PrivacySuppressed"),]

> head(college2)

SchoolType FieldCode

16 Public 11.01

25 Public 13.13

28 Public 14.10

29 Public 14.19

38 Public 24.01

39 Public 26.01

Field

16 Computer and Information Sciences, General.

25 Teacher Education and Professional Development, Specific Subject Areas.

28 Electrical, Electronics and Communications Engineering.

29 Mechanical Engineering.

38 Liberal Arts and Sciences, General Studies and Humanities.

39 Biology, General.

DegreeLevel DebtMonthlyPaym DebtMean Earnings

16 3 376 37194 54300

25 3 327 33085 25400

28 3 361 36482 54500

29 3 383 38900 45800

38 3 311 30970 24000

39 3 353 32783 23800

>

> college2$SchoolType <- as.character(college2$SchoolType)

> college2$Field <- as.character(college2$Field)

> college2$DebtMonthlyPaym <- as.numeric(as.character(college2$DebtMonthlyPaym))

Warning message:

NAs introduced by coercion

> college2$DebtMean <- as.numeric(as.character(college2$DebtMean))

Warning message:

NAs introduced by coercion

> college2$Earnings <- as.numeric(as.character(college2$Earnings))

> college3<-college2 %>%

+ mutate(Field = case\_when(college2$FieldCode<5 ~ "NaturalSciences",

+ college2$FieldCode>=5 & college2$FieldCode<11 ~ "Humanities",

+ college2$FieldCode>=11 & college2$FieldCode<12 ~ "AppliedSciences",

+ college2$FieldCode>=12 & college2$FieldCode<14 ~ "Humanities",

+ college2$FieldCode>=14 & college2$FieldCode<16 ~ "AppliedSciences",

+ college2$FieldCode>=16 & college2$FieldCode<17 ~ "Humanities",

+ college2$FieldCode>=19 & college2$FieldCode<22 ~ "SocialSciences",

+ college2$FieldCode==22 ~ "Others",

+ college2$FieldCode>22 & college2$FieldCode<26 ~ "Humanities",

+ college2$FieldCode>=26 & college2$FieldCode<27 ~ "NaturalSciences",

+ college2$FieldCode>=27 & college2$FieldCode<30 ~ "AppliedSciences",

+ college2$FieldCode==30 ~ "Others",

+ college2$FieldCode>30 & college2$FieldCode<30.1 ~ "AppliedSciences",

+ college2$FieldCode==30.11 & college2$FieldCode<30.18 ~ "SocialSciences",

+ college2$FieldCode>=30.18 & college2$FieldCode<30.2 ~ "NaturalSciences",

+ college2$FieldCode>=30.2 & college2$FieldCode<30.27 ~ "Humanities",

+ college2$FieldCode>=30.27 & college2$FieldCode<30.99 ~ "NaturalSciences",

+ college2$FieldCode==30.99 ~ "Others",

+ college2$FieldCode>=31 & college2$FieldCode<40 ~ "Humanities",

+ college2$FieldCode>=40 & college2$FieldCode<42 ~ "NaturalSciences",

+ college2$FieldCode>=42 & college2$FieldCode<43 ~ "SocialSciences",

+ college2$FieldCode>=43 & college2$FieldCode<45 ~ "Humanities",

+ college2$FieldCode>=45 & college2$FieldCode<46 ~ "SocialSciences",

+ college2$FieldCode>=46 & college2$FieldCode<50 ~ "Others",

+ college2$FieldCode>=50 ~ "Others",

+ college2$FieldCode>=51 & college2$FieldCode<54 ~ "AppliedSciences",

+ TRUE ~ "Humanities"))

> head(college3)

SchoolType FieldCode Field DegreeLevel

1 Public 11.01 AppliedSciences 3

2 Public 13.13 Humanities 3

3 Public 14.10 AppliedSciences 3

4 Public 14.19 AppliedSciences 3

5 Public 24.01 Humanities 3

6 Public 26.01 NaturalSciences 3

DebtMonthlyPaym DebtMean Earnings

1 376 37194 54300

2 327 33085 25400

3 361 36482 54500

4 383 38900 45800

5 311 30970 24000

6 353 32783 23800

> college <- college3[,c(1,3:7)]

> head(college)

SchoolType Field DegreeLevel

1 Public AppliedSciences 3

2 Public Humanities 3

3 Public AppliedSciences 3

4 Public AppliedSciences 3

5 Public Humanities 3

6 Public NaturalSciences 3

DebtMonthlyPaym DebtMean Earnings

1 376 37194 54300

2 327 33085 25400

3 361 36482 54500

4 383 38900 45800

5 311 30970 24000

6 353 32783 23800

>

> public <- subset(college, SchoolType == "Public")

> private1 <- subset(college, SchoolType == "Private, for-profit")

> private2 <- subset(college, SchoolType == "Private, nonprofit")

> foreign <- subset(college, SchoolType == "Foreign")

>

>

> #install.packages("tidyverse")

> #install.packages("ggpubr")

> library(tidyverse)

> library(ggplot2)

> library(ggpubr)

> # basic plot for Earnings vs SchoolType

> p1 <- ggplot(public, aes(x=DegreeLevel, y=Earnings)) + geom\_point() + ggtitle("Public")

> p2 <- ggplot(private1, aes(x=DegreeLevel, y=Earnings)) + geom\_point() + ggtitle("Private, for-profit")

> p3 <- ggplot(private2, aes(x=DegreeLevel, y=Earnings)) + geom\_point() + ggtitle("Private, nonprofit")

> p4 <- ggplot(foreign, aes(x=DegreeLevel, y=Earnings)) + geom\_point() + ggtitle("Foreign")

> ggarrange(p1, p2, p3, p4, widths=1)

A screenshot of a cell phone

Description automatically generated

> lmod1 <- lm(Earnings ~ SchoolType + Field + DegreeLevel + DebtMonthlyPaym + DebtMean, data = college)

> summary(lmod1)

Call:

lm(formula = Earnings ~ SchoolType + Field + DegreeLevel + DebtMonthlyPaym +

DebtMean, data = college)

Residuals:

Min 1Q Median 3Q Max

-85131 -7186 -1228 5522 155691

Coefficients:

Estimate Std. Error

(Intercept) 1.546e+04 2.788e+03

SchoolTypePrivate, for-profit 1.128e+04 2.778e+03

SchoolTypePrivate, nonprofit 1.751e+04 2.776e+03

SchoolTypePublic 1.692e+04 2.779e+03

FieldHumanities -2.210e+04 2.507e+02

FieldNaturalSciences -2.302e+04 3.701e+02

FieldOthers -1.096e+04 2.379e+02

FieldSocialSciences -2.410e+04 3.230e+02

DegreeLevel 6.614e+03 7.220e+01

DebtMonthlyPaym -1.247e+01 1.727e+00

DebtMean 2.516e-01 2.140e-02

t value Pr(>|t|)

(Intercept) 5.545 2.96e-08 \*\*\*

SchoolTypePrivate, for-profit 4.061 4.89e-05 \*\*\*

SchoolTypePrivate, nonprofit 6.307 2.87e-10 \*\*\*

SchoolTypePublic 6.088 1.15e-09 \*\*\*

FieldHumanities -88.159 < 2e-16 \*\*\*

FieldNaturalSciences -62.191 < 2e-16 \*\*\*

FieldOthers -46.078 < 2e-16 \*\*\*

FieldSocialSciences -74.624 < 2e-16 \*\*\*

DegreeLevel 91.603 < 2e-16 \*\*\*

DebtMonthlyPaym -7.216 5.44e-13 \*\*\*

DebtMean 11.760 < 2e-16 \*\*\*

---

Signif. codes:

0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 13700 on 40918 degrees of freedom

(4442 observations deleted due to missingness)

Multiple R-squared: 0.4863, Adjusted R-squared: 0.4861

F-statistic: 3873 on 10 and 40918 DF, p-value: < 2.2e-16

> lmod2 <- lm(Earnings ~ SchoolType + Field + DegreeLevel, data = college)

> summary(lmod2)

Call:

lm(formula = Earnings ~ SchoolType + Field + DegreeLevel, data = college)

Residuals:

Min 1Q Median 3Q Max

-74243 -7533 -1322 5489 195719

Coefficients:

Estimate Std. Error

(Intercept) 21311.77 2774.50

SchoolTypePrivate, for-profit 5142.63 2759.59

SchoolTypePrivate, nonprofit 11190.31 2755.93

SchoolTypePublic 10107.64 2755.75

FieldHumanities -22193.56 242.13

FieldNaturalSciences -22823.30 363.50

FieldOthers -10698.10 229.44

FieldSocialSciences -24017.76 317.75

DegreeLevel 7877.07 49.61

t value Pr(>|t|)

(Intercept) 7.681 1.61e-14 \*\*\*

SchoolTypePrivate, for-profit 1.864 0.062392 .

SchoolTypePrivate, nonprofit 4.060 4.91e-05 \*\*\*

SchoolTypePublic 3.668 0.000245 \*\*\*

FieldHumanities -91.658 < 2e-16 \*\*\*

FieldNaturalSciences -62.787 < 2e-16 \*\*\*

FieldOthers -46.628 < 2e-16 \*\*\*

FieldSocialSciences -75.586 < 2e-16 \*\*\*

DegreeLevel 158.790 < 2e-16 \*\*\*

---

Signif. codes:

0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 14020 on 45362 degrees of freedom

Multiple R-squared: 0.4779, Adjusted R-squared: 0.4779

F-statistic: 5191 on 8 and 45362 DF, p-value: < 2.2e-16

> lmod3 <- lm(Earnings ~ Field + DegreeLevel, data = college)

> summary(lmod3)

Call:

lm(formula = Earnings ~ Field + DegreeLevel, data = college)

Residuals:

Min 1Q Median 3Q Max

-73985 -7585 -1474 5526 199021

Coefficients:

Estimate Std. Error t value

(Intercept) 28957.84 252.60 114.64

FieldHumanities -22181.41 244.86 -90.59

FieldNaturalSciences -21817.84 366.41 -59.55

FieldOthers -10962.74 231.55 -47.34

FieldSocialSciences -23136.29 320.23 -72.25

DegreeLevel 8384.30 47.29 177.28

Pr(>|t|)

(Intercept) <2e-16 \*\*\*

FieldHumanities <2e-16 \*\*\*

FieldNaturalSciences <2e-16 \*\*\*

FieldOthers <2e-16 \*\*\*

FieldSocialSciences <2e-16 \*\*\*

DegreeLevel <2e-16 \*\*\*

---

Signif. codes:

0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 14190 on 45365 degrees of freedom

Multiple R-squared: 0.4653, Adjusted R-squared: 0.4653

F-statistic: 7897 on 5 and 45365 DF, p-value: < 2.2e-16

> require(caret)

> # Define training control

> set.seed(13245)

> train.control <- trainControl(method = "cv", number = 10)

> # Train the model

> model\_1 <- train(Earnings ~ SchoolType + Field + DegreeLevel + DebtMonthlyPaym + DebtMean,

+ data = college, method = "lm", trControl = train.control, na.action=na.exclude)

> # Summarize the results

> print(model\_1)

Linear Regression

45371 samples

5 predictor

No pre-processing

Resampling: Cross-Validated (10 fold)

Summary of sample sizes: 36837, 36835, 36835, 36836, 36836, 36837, ...

Resampling results:

RMSE Rsquared MAE

13697.63 0.4862073 9417.737

Tuning parameter 'intercept' was held constant at

a value of TRUE

> ##

>

> # Define training control

> set.seed(14235)

> train.control <- trainControl(method = "cv", number = 10)

> # Train the model

> model\_2 <- train(Earnings ~ SchoolType + Field + DegreeLevel,

+ data = college, method = "lm", trControl = train.control)

> # Summarize the results

> print(model\_2)

Linear Regression

45371 samples

3 predictor

No pre-processing

Resampling: Cross-Validated (10 fold)

Summary of sample sizes: 40834, 40835, 40833, 40834, 40834, 40834, ...

Resampling results:

RMSE Rsquared MAE

14022.68 0.4779141 9579.779

Tuning parameter 'intercept' was held constant at

a value of TRUE

> ##

>

> # Define training control

> set.seed(13235)

> train.control <- trainControl(method = "cv", number = 10)

> # Train the model

> model\_3 <- train(Earnings ~ Field + DegreeLevel,

+ data = college, method = "lm", trControl = train.control, na.action=na.exclude)

> # Summarize the results

> print(model\_3)

Linear Regression

45371 samples

2 predictor

No pre-processing

Resampling: Cross-Validated (10 fold)

Summary of sample sizes: 40834, 40834, 40833, 40835, 40833, 40833, ...

Resampling results:

RMSE Rsquared MAE

14187.28 0.4657353 9700.768

Tuning parameter 'intercept' was held constant at

a value of TRUE

> ##

> ols\_step\_forward\_p(lmod1,details=TRUE)

Forward Selection Method

---------------------------

Candidate Terms:

1. SchoolType

2. Field

3. DegreeLevel

4. DebtMonthlyPaym

5. DebtMean

We are selecting variables based on p value...

Forward Selection: Step 1

+ DebtMean

Model Summary

---------------------------------------------------------------------

R 0.697 RMSE 13697.688

R-Squared 0.486 Coef. Var 34.380

Adj. R-Squared 0.486 MSE 187626661.068

Pred R-Squared 0.486 MAE 9414.962

---------------------------------------------------------------------

RMSE: Root Mean Square Error

MSE: Mean Square Error

MAE: Mean Absolute Error

ANOVA

----------------------------------------------------------------------------------

Sum of

Squares DF Mean Square F Sig.

----------------------------------------------------------------------------------

Regression 7.266989e+12 10 726698888593.011 3873.111 0.0000

Residual 7.677308e+12 40918 187626661.068

Total 1.49443e+13 40928

----------------------------------------------------------------------------------

Parameter Estimates

-----------------------------------------------------------------------------------------------------------------------

model Beta Std. Error Std. Beta t Sig lower upper

-----------------------------------------------------------------------------------------------------------------------

(Intercept) 15460.635 2788.362 5.545 0.000 9995.384 20925.886

SchoolTypePrivate, for-profit 11280.348 2777.566 0.255 4.061 0.000 5836.258 16724.438

SchoolTypePrivate, nonprofit 17507.740 2775.818 0.401 6.307 0.000 12067.076 22948.404

SchoolTypePublic 16919.145 2779.014 0.443 6.088 0.000 11472.217 22366.074

FieldHumanities -22098.919 250.672 -0.524 -88.159 0.000 -22590.241 -21607.597

FieldNaturalSciences -23016.627 370.096 -0.266 -62.191 0.000 -23742.023 -22291.232

FieldOthers -10960.841 237.876 -0.287 -46.078 0.000 -11427.083 -10494.599

FieldSocialSciences -24102.867 322.990 -0.344 -74.624 0.000 -24735.934 -23469.800

DegreeLevel 6613.849 72.202 0.479 91.603 0.000 6472.332 6755.366

DebtMonthlyPaym -12.465 1.727 -0.196 -7.216 0.000 -15.851 -9.080

DebtMean 0.252 0.021 0.325 11.760 0.000 0.210 0.294

-----------------------------------------------------------------------------------------------------------------------

> ols\_step\_backward\_p(lmod,details=TRUE)

Backward Elimination Method

---------------------------

Candidate Terms:

1 . SchoolType

2 . Field

3 . DegreeLevel

4 . DebtMonthlyPaym

5 . DebtMean

We are eliminating variables based on p value...

No more variables satisfy the condition of p value = 0.3

Variables Removed:

Final Model Output

------------------

Model Summary

---------------------------------------------------------------------

R 0.697 RMSE 13697.688

R-Squared 0.486 Coef. Var 34.380

Adj. R-Squared 0.486 MSE 187626661.068

Pred R-Squared 0.486 MAE 9414.962

---------------------------------------------------------------------

RMSE: Root Mean Square Error

MSE: Mean Square Error

MAE: Mean Absolute Error

ANOVA

----------------------------------------------------------------------------------

Sum of

Squares DF Mean Square F Sig.

----------------------------------------------------------------------------------

Regression 7.266989e+12 10 726698888593.011 3873.111 0.0000

Residual 7.677308e+12 40918 187626661.068

Total 1.49443e+13 40928

----------------------------------------------------------------------------------

Parameter Estimates

-----------------------------------------------------------------------------------------------------------------------

model Beta Std. Error Std. Beta t Sig lower upper

-----------------------------------------------------------------------------------------------------------------------

(Intercept) 15460.635 2788.362 5.545 0.000 9995.384 20925.886

SchoolTypePrivate, for-profit 11280.348 2777.566 0.255 4.061 0.000 5836.258 16724.438

SchoolTypePrivate, nonprofit 17507.740 2775.818 0.401 6.307 0.000 12067.076 22948.404

SchoolTypePublic 16919.145 2779.014 0.443 6.088 0.000 11472.217 22366.074

FieldHumanities -22098.919 250.672 -0.524 -88.159 0.000 -22590.241 -21607.597

FieldNaturalSciences -23016.627 370.096 -0.266 -62.191 0.000 -23742.023 -22291.232

FieldOthers -10960.841 237.876 -0.287 -46.078 0.000 -11427.083 -10494.599

FieldSocialSciences -24102.867 322.990 -0.344 -74.624 0.000 -24735.934 -23469.800

DegreeLevel 6613.849 72.202 0.479 91.603 0.000 6472.332 6755.366

DebtMonthlyPaym -12.465 1.727 -0.196 -7.216 0.000 -15.851 -9.080

DebtMean 0.252 0.021 0.325 11.760 0.000 0.210 0.294

-----------------------------------------------------------------------------------------------------------------------

[1] "No variables have been removed from the model."

> ols\_step\_both\_p(lmod,details=TRUE)

Stepwise Selection Method

---------------------------

Candidate Terms:

1. SchoolType

2. Field

3. DegreeLevel

4. DebtMonthlyPaym

5. DebtMean

We are selecting variables based on p value...

Stepwise Selection: Step 1

+ DebtMean

Model Summary

---------------------------------------------------------------------

R 0.697 RMSE 13697.688

R-Squared 0.486 Coef. Var 34.380

Adj. R-Squared 0.486 MSE 187626661.068

Pred R-Squared 0.486 MAE 9414.962

---------------------------------------------------------------------

RMSE: Root Mean Square Error

MSE: Mean Square Error

MAE: Mean Absolute Error

ANOVA

----------------------------------------------------------------------------------

Sum of

Squares DF Mean Square F Sig.

----------------------------------------------------------------------------------

Regression 7.266989e+12 10 726698888593.011 3873.111 0.0000

Residual 7.677308e+12 40918 187626661.068

Total 1.49443e+13 40928

----------------------------------------------------------------------------------

Parameter Estimates

-----------------------------------------------------------------------------------------------------------------------

model Beta Std. Error Std. Beta t Sig lower upper

-----------------------------------------------------------------------------------------------------------------------

(Intercept) 15460.635 2788.362 5.545 0.000 9995.384 20925.886

SchoolTypePrivate, for-profit 11280.348 2777.566 0.255 4.061 0.000 5836.258 16724.438

SchoolTypePrivate, nonprofit 17507.740 2775.818 0.401 6.307 0.000 12067.076 22948.404

SchoolTypePublic 16919.145 2779.014 0.443 6.088 0.000 11472.217 22366.074

FieldHumanities -22098.919 250.672 -0.524 -88.159 0.000 -22590.241 -21607.597

FieldNaturalSciences -23016.627 370.096 -0.266 -62.191 0.000 -23742.023 -22291.232

FieldOthers -10960.841 237.876 -0.287 -46.078 0.000 -11427.083 -10494.599

FieldSocialSciences -24102.867 322.990 -0.344 -74.624 0.000 -24735.934 -23469.800

DegreeLevel 6613.849 72.202 0.479 91.603 0.000 6472.332 6755.366

DebtMonthlyPaym -12.465 1.727 -0.196 -7.216 0.000 -15.851 -9.080

DebtMean 0.252 0.021 0.325 11.760 0.000 0.210 0.294

-----------------------------------------------------------------------------------------------------------------------

>

> lmod4 <- lm(DebtMean ~ SchoolType + Field + DegreeLevel + DebtMonthlyPaym + Earnings, data = college)

> summary(lmod4)

Call:

lm(formula = DebtMean ~ SchoolType + Field + DegreeLevel + DebtMonthlyPaym +

Earnings, data = college)

Residuals:

Min 1Q Median 3Q Max

-35331 -1415 -172 1208 96292

Coefficients:

Estimate Std. Error

(Intercept) -6.579e+03 6.426e+02

SchoolTypePrivate, for-profit 8.086e+03 6.396e+02

SchoolTypePrivate, nonprofit 7.771e+03 6.395e+02

SchoolTypePublic 7.954e+03 6.401e+02

FieldHumanities 3.213e+02 6.306e+01

FieldNaturalSciences -4.398e+01 8.931e+01

FieldOthers 2.523e+02 5.626e+01

FieldSocialSciences 3.510e+02 7.939e+01

DegreeLevel 6.500e+02 1.800e+01

DebtMonthlyPaym 7.931e+01 7.256e-02

Earnings 1.339e-02 1.138e-03

t value Pr(>|t|)

(Intercept) -10.238 < 2e-16 \*\*\*

SchoolTypePrivate, for-profit 12.642 < 2e-16 \*\*\*

SchoolTypePrivate, nonprofit 12.153 < 2e-16 \*\*\*

SchoolTypePublic 12.426 < 2e-16 \*\*\*

FieldHumanities 5.096 3.49e-07 \*\*\*

FieldNaturalSciences -0.492 0.622

FieldOthers 4.484 7.33e-06 \*\*\*

FieldSocialSciences 4.421 9.87e-06 \*\*\*

DegreeLevel 36.113 < 2e-16 \*\*\*

DebtMonthlyPaym 1092.998 < 2e-16 \*\*\*

Earnings 11.760 < 2e-16 \*\*\*

---

Signif. codes:

0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3160 on 40918 degrees of freedom

(4442 observations deleted due to missingness)

Multiple R-squared: 0.9836, Adjusted R-squared: 0.9836

F-statistic: 2.455e+05 on 10 and 40918 DF, p-value: < 2.2e-16

> lmod5 <- lm(DebtMean ~ SchoolType + Field + DegreeLevel, data = college)

> summary(lmod5)

Call:

lm(formula = DebtMean ~ SchoolType + Field + DegreeLevel, data = college)

Residuals:

Min 1Q Median 3Q Max

-113933 -5687 -679 4075 290913

Coefficients:

Estimate Std. Error

(Intercept) 78087.09 3531.02

SchoolTypePrivate, for-profit -81948.74 3511.88

SchoolTypePrivate, nonprofit -87107.81 3506.92

SchoolTypePublic -92735.99 3506.71

FieldHumanities -1660.81 307.75

FieldNaturalSciences -856.21 458.89

FieldOthers 3805.12 291.72

FieldSocialSciences 869.11 401.35

DegreeLevel 12198.58 64.33

t value Pr(>|t|)

(Intercept) 22.115 < 2e-16 \*\*\*

SchoolTypePrivate, for-profit -23.335 < 2e-16 \*\*\*

SchoolTypePrivate, nonprofit -24.839 < 2e-16 \*\*\*

SchoolTypePublic -26.445 < 2e-16 \*\*\*

FieldHumanities -5.397 6.83e-08 \*\*\*

FieldNaturalSciences -1.866 0.0621 .

FieldOthers 13.044 < 2e-16 \*\*\*

FieldSocialSciences 2.165 0.0304 \*

DegreeLevel 189.636 < 2e-16 \*\*\*

---

Signif. codes:

0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 17500 on 43408 degrees of freedom

(1954 observations deleted due to missingness)

Multiple R-squared: 0.4856, Adjusted R-squared: 0.4856

F-statistic: 5123 on 8 and 43408 DF, p-value: < 2.2e-16

> lmod6 <- lm(DebtMean ~ Field + DegreeLevel, data = college)

> summary(lmod6)

Call:

lm(formula = DebtMean ~ Field + DegreeLevel, data = college)

Residuals:

Min 1Q Median 3Q Max

-78560 -7213 -1464 4665 293226

Coefficients:

Estimate Std. Error t value

(Intercept) -9554.00 330.26 -28.929

FieldHumanities -1234.91 318.85 -3.873

FieldNaturalSciences -2778.06 473.90 -5.862

FieldOthers 5065.59 301.44 16.805

FieldSocialSciences 92.92 414.33 0.224

DegreeLevel 11691.89 62.70 186.481

Pr(>|t|)

(Intercept) < 2e-16 \*\*\*

FieldHumanities 0.000108 \*\*\*

FieldNaturalSciences 4.6e-09 \*\*\*

FieldOthers < 2e-16 \*\*\*

FieldSocialSciences 0.822549

DegreeLevel < 2e-16 \*\*\*

---

Signif. codes:

0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 18140 on 43411 degrees of freedom

(1954 observations deleted due to missingness)

Multiple R-squared: 0.447, Adjusted R-squared: 0.4469

F-statistic: 7017 on 5 and 43411 DF, p-value: < 2.2e-16

>

> require(caret)

> # Define training control

> set.seed(13245)

> train.control <- trainControl(method = "cv", number = 10)

> # Train the model

> model\_4 <- train(DebtMean ~ SchoolType + Field + DegreeLevel + DebtMonthlyPaym + Earnings,

+ data = college, method = "lm", trControl = train.control, na.action=na.exclude)

> # Summarize the results

> print(model\_4)

Linear Regression

45371 samples

5 predictor

No pre-processing

Resampling: Cross-Validated (10 fold)

Summary of sample sizes: 36835, 36837, 36836, 36837, 36837, 36835, ...

Resampling results:

RMSE Rsquared MAE

3152.213 0.9834025 1876.743

Tuning parameter 'intercept' was held constant at

a value of TRUE

> ##

>

> # Define training control

> set.seed(14235)

> train.control <- trainControl(method = "cv", number = 10)

> # Train the model

> model\_5 <- train(DebtMean ~ SchoolType + Field + DegreeLevel,

+ data = college, method = "lm", trControl = train.control, na.action=na.exclude)

> # Summarize the results

> print(model\_5)

Linear Regression

45371 samples

3 predictor

No pre-processing

Resampling: Cross-Validated (10 fold)

Summary of sample sizes: 39075, 39075, 39075, 39075, 39075, 39076, ...

Resampling results:

RMSE Rsquared MAE

17491.72 0.4857304 8846.212

Tuning parameter 'intercept' was held constant at

a value of TRUE

> ##

>

> # Define training control

> set.seed(13235)

> train.control <- trainControl(method = "cv", number = 10)

> # Train the model

> model\_6 <- train(DebtMean ~ Field + DegreeLevel,

+ data = college, method = "lm", trControl = train.control, na.action=na.exclude)

> # Summarize the results

> print(model\_6)

Linear Regression

45371 samples

2 predictor

No pre-processing

Resampling: Cross-Validated (10 fold)

Summary of sample sizes: 39074, 39075, 39075, 39076, 39076, 39076, ...

Resampling results:

RMSE Rsquared MAE

18106.32 0.4490754 9715.292

Tuning parameter 'intercept' was held constant at

a value of TRUE

> ##

>

>

> ols\_step\_forward\_p(lmod4,details=TRUE)

Forward Selection Method

---------------------------

Candidate Terms:

1. SchoolType

2. Field

3. DegreeLevel

4. DebtMonthlyPaym

5. Earnings

We are selecting variables based on p value...

Forward Selection: Step 1

+ Earnings

Model Summary

-------------------------------------------------------------------

R 0.992 RMSE 3161.396

R-Squared 0.984 Coef. Var 11.155

Adj. R-Squared 0.984 MSE 9994427.754

Pred R-Squared 0.984 MAE 1877.718

-------------------------------------------------------------------

RMSE: Root Mean Square Error

MSE: Mean Square Error

MAE: Mean Absolute Error

ANOVA

--------------------------------------------------------------------------------

Sum of

Squares DF Mean Square F Sig.

--------------------------------------------------------------------------------

Regression 2.451349e+13 6 4.085582e+12 408785.971 0.0000

Residual 4.08992e+11 40922 9994427.754

Total 2.492248e+13 40928

--------------------------------------------------------------------------------

Parameter Estimates

---------------------------------------------------------------------------------------------------------------------

model Beta Std. Error Std. Beta t Sig lower upper

---------------------------------------------------------------------------------------------------------------------

(Intercept) -6328.988 640.867 -9.876 0.000 -7585.102 -5072.874

SchoolTypePrivate, for-profit 8100.985 639.887 0.142 12.660 0.000 6846.792 9355.178

SchoolTypePrivate, nonprofit 7794.855 639.736 0.138 12.184 0.000 6540.958 9048.753

SchoolTypePublic 7958.195 640.397 0.161 12.427 0.000 6703.004 9213.387

DegreeLevel 666.132 17.155 0.037 38.830 0.000 632.508 699.756

DebtMonthlyPaym 79.325 0.072 0.965 1099.440 0.000 79.183 79.466

Earnings 0.011 0.001 0.009 11.410 0.000 0.010 0.013

---------------------------------------------------------------------------------------------------------------------

Forward Selection: Step 2

+ Field

Model Summary

-------------------------------------------------------------------

R 0.992 RMSE 3159.681

R-Squared 0.984 Coef. Var 11.149

Adj. R-Squared 0.984 MSE 9983583.865

Pred R-Squared 0.984 MAE 1876.110

-------------------------------------------------------------------

RMSE: Root Mean Square Error

MSE: Mean Square Error

MAE: Mean Absolute Error

ANOVA

------------------------------------------------------------------------------------

Sum of

Squares DF Mean Square F Sig.

------------------------------------------------------------------------------------

Regression 2.451397e+13 10 2.451397e+12 245542.835 0.0000

Residual 408508284594.389 40918 9983583.865

Total 2.492248e+13 40928

------------------------------------------------------------------------------------

Parameter Estimates

---------------------------------------------------------------------------------------------------------------------

model Beta Std. Error Std. Beta t Sig lower upper

---------------------------------------------------------------------------------------------------------------------

(Intercept) -6578.852 642.618 -10.238 0.000 -7838.397 -5319.307

SchoolTypePrivate, for-profit 8085.754 639.590 0.141 12.642 0.000 6832.144 9339.363

SchoolTypePrivate, nonprofit 7771.476 639.463 0.138 12.153 0.000 6518.114 9024.838

SchoolTypePublic 7954.386 640.126 0.161 12.426 0.000 6699.725 9209.047

DegreeLevel 649.971 17.998 0.036 36.113 0.000 614.694 685.249

DebtMonthlyPaym 79.308 0.073 0.965 1092.998 0.000 79.166 79.450

Earnings 0.013 0.001 0.010 11.760 0.000 0.011 0.016

FieldHumanities 321.323 63.056 0.006 5.096 0.000 197.732 444.914

FieldNaturalSciences -43.984 89.314 0.000 -0.492 0.622 -219.042 131.074

FieldOthers 252.305 56.263 0.005 4.484 0.000 142.028 362.582

FieldSocialSciences 350.972 79.394 0.004 4.421 0.000 195.358 506.586

---------------------------------------------------------------------------------------------------------------------

> ols\_step\_backward\_p(lmod4,details=TRUE)

Backward Elimination Method

---------------------------

Candidate Terms:

1 . SchoolType

2 . Field

3 . DegreeLevel

4 . DebtMonthlyPaym

5 . Earnings

We are eliminating variables based on p value...

No more variables satisfy the condition of p value = 0.3

Variables Removed:

Final Model Output

------------------

Model Summary

-------------------------------------------------------------------

R 0.992 RMSE 3159.681

R-Squared 0.984 Coef. Var 11.149

Adj. R-Squared 0.984 MSE 9983583.865

Pred R-Squared 0.984 MAE 1876.110

-------------------------------------------------------------------

RMSE: Root Mean Square Error

MSE: Mean Square Error

MAE: Mean Absolute Error

ANOVA

------------------------------------------------------------------------------------

Sum of

Squares DF Mean Square F Sig.

------------------------------------------------------------------------------------

Regression 2.451397e+13 10 2.451397e+12 245542.835 0.0000

Residual 408508284594.389 40918 9983583.865

Total 2.492248e+13 40928

------------------------------------------------------------------------------------

Parameter Estimates

---------------------------------------------------------------------------------------------------------------------

model Beta Std. Error Std. Beta t Sig lower upper

---------------------------------------------------------------------------------------------------------------------

(Intercept) -6578.852 642.618 -10.238 0.000 -7838.397 -5319.307

SchoolTypePrivate, for-profit 8085.754 639.590 0.141 12.642 0.000 6832.144 9339.363

SchoolTypePrivate, nonprofit 7771.476 639.463 0.138 12.153 0.000 6518.114 9024.838

SchoolTypePublic 7954.386 640.126 0.161 12.426 0.000 6699.725 9209.047

FieldHumanities 321.323 63.056 0.006 5.096 0.000 197.732 444.914

FieldNaturalSciences -43.984 89.314 0.000 -0.492 0.622 -219.042 131.074

FieldOthers 252.305 56.263 0.005 4.484 0.000 142.028 362.582

FieldSocialSciences 350.972 79.394 0.004 4.421 0.000 195.358 506.586

DegreeLevel 649.971 17.998 0.036 36.113 0.000 614.694 685.249

DebtMonthlyPaym 79.308 0.073 0.965 1092.998 0.000 79.166 79.450

Earnings 0.013 0.001 0.010 11.760 0.000 0.011 0.016

---------------------------------------------------------------------------------------------------------------------

[1] "No variables have been removed from the model."

> ols\_step\_both\_p(lmod4,details=TRUE)

Stepwise Selection Method

---------------------------

Candidate Terms:

1. SchoolType

2. Field

3. DegreeLevel

4. DebtMonthlyPaym

5. Earnings

We are selecting variables based on p value...

Stepwise Selection: Step 1

+ Earnings

Model Summary

-------------------------------------------------------------------

R 0.992 RMSE 3161.396

R-Squared 0.984 Coef. Var 11.155

Adj. R-Squared 0.984 MSE 9994427.754

Pred R-Squared 0.984 MAE 1877.718

-------------------------------------------------------------------

RMSE: Root Mean Square Error

MSE: Mean Square Error

MAE: Mean Absolute Error

ANOVA

--------------------------------------------------------------------------------

Sum of

Squares DF Mean Square F Sig.

--------------------------------------------------------------------------------

Regression 2.451349e+13 6 4.085582e+12 408785.971 0.0000

Residual 4.08992e+11 40922 9994427.754

Total 2.492248e+13 40928

--------------------------------------------------------------------------------

Parameter Estimates

---------------------------------------------------------------------------------------------------------------------

model Beta Std. Error Std. Beta t Sig lower upper

---------------------------------------------------------------------------------------------------------------------

(Intercept) -6328.988 640.867 -9.876 0.000 -7585.102 -5072.874

SchoolTypePrivate, for-profit 8100.985 639.887 0.142 12.660 0.000 6846.792 9355.178

SchoolTypePrivate, nonprofit 7794.855 639.736 0.138 12.184 0.000 6540.958 9048.753

SchoolTypePublic 7958.195 640.397 0.161 12.427 0.000 6703.004 9213.387

DegreeLevel 666.132 17.155 0.037 38.830 0.000 632.508 699.756

DebtMonthlyPaym 79.325 0.072 0.965 1099.440 0.000 79.183 79.466

Earnings 0.011 0.001 0.009 11.410 0.000 0.010 0.013

---------------------------------------------------------------------------------------------------------------------

Stepwise Selection: Step 2

+ Field

Model Summary

-------------------------------------------------------------------

R 0.992 RMSE 3159.681

R-Squared 0.984 Coef. Var 11.149

Adj. R-Squared 0.984 MSE 9983583.865

Pred R-Squared 0.984 MAE 1876.110

-------------------------------------------------------------------

RMSE: Root Mean Square Error

MSE: Mean Square Error

MAE: Mean Absolute Error

ANOVA

------------------------------------------------------------------------------------

Sum of

Squares DF Mean Square F Sig.

------------------------------------------------------------------------------------

Regression 2.451397e+13 10 2.451397e+12 245542.835 0.0000

Residual 408508284594.389 40918 9983583.865

Total 2.492248e+13 40928

------------------------------------------------------------------------------------

Parameter Estimates

---------------------------------------------------------------------------------------------------------------------

model Beta Std. Error Std. Beta t Sig lower upper

---------------------------------------------------------------------------------------------------------------------

(Intercept) -6578.852 642.618 -10.238 0.000 -7838.397 -5319.307

SchoolTypePrivate, for-profit 8085.754 639.590 0.141 12.642 0.000 6832.144 9339.363

SchoolTypePrivate, nonprofit 7771.476 639.463 0.138 12.153 0.000 6518.114 9024.838

SchoolTypePublic 7954.386 640.126 0.161 12.426 0.000 6699.725 9209.047

DegreeLevel 649.971 17.998 0.036 36.113 0.000 614.694 685.249

DebtMonthlyPaym 79.308 0.073 0.965 1092.998 0.000 79.166 79.450

Earnings 0.013 0.001 0.010 11.760 0.000 0.011 0.016

FieldHumanities 321.323 63.056 0.006 5.096 0.000 197.732 444.914

FieldNaturalSciences -43.984 89.314 0.000 -0.492 0.622 -219.042 131.074

FieldOthers 252.305 56.263 0.005 4.484 0.000 142.028 362.582

FieldSocialSciences 350.972 79.394 0.004 4.421 0.000 195.358 506.586

---------------------------------------------------------------------------------------------------------------------

Model Summary

-------------------------------------------------------------------

R 0.992 RMSE 3159.681

R-Squared 0.984 Coef. Var 11.149

Adj. R-Squared 0.984 MSE 9983583.865

Pred R-Squared 0.984 MAE 1876.110

-------------------------------------------------------------------

RMSE: Root Mean Square Error

MSE: Mean Square Error

MAE: Mean Absolute Error

ANOVA

------------------------------------------------------------------------------------

Sum of

Squares DF Mean Square F Sig.

------------------------------------------------------------------------------------

Regression 2.451397e+13 10 2.451397e+12 245542.835 0.0000

Residual 408508284594.389 40918 9983583.865

Total 2.492248e+13 40928

------------------------------------------------------------------------------------

Parameter Estimates

---------------------------------------------------------------------------------------------------------------------

model Beta Std. Error Std. Beta t Sig lower upper

---------------------------------------------------------------------------------------------------------------------

(Intercept) -6578.852 642.618 -10.238 0.000 -7838.397 -5319.307

SchoolTypePrivate, for-profit 8085.754 639.590 0.141 12.642 0.000 6832.144 9339.363

SchoolTypePrivate, nonprofit 7771.476 639.463 0.138 12.153 0.000 6518.114 9024.838

SchoolTypePublic 7954.386 640.126 0.161 12.426 0.000 6699.725 9209.047

DegreeLevel 649.971 17.998 0.036 36.113 0.000 614.694 685.249

DebtMonthlyPaym 79.308 0.073 0.965 1092.998 0.000 79.166 79.450

Earnings 0.013 0.001 0.010 11.760 0.000 0.011 0.016

FieldHumanities 321.323 63.056 0.006 5.096 0.000 197.732 444.914

FieldNaturalSciences -43.984 89.314 0.000 -0.492 0.622 -219.042 131.074

FieldOthers 252.305 56.263 0.005 4.484 0.000 142.028 362.582

FieldSocialSciences 350.972 79.394 0.004 4.421 0.000 195.358 506.586

---------------------------------------------------------------------------------------------------------------------

**Works Cited**

Data source: <https://collegescorecard.ed.gov/data/>