Optimizing High School Graduation Rates

Alyssa Newman, Collin Powell, Zane Showalter-Castorena

The View From 30k Feet

- Deliverable: A customizable model which produces a strategy to optimize graduation rates for a particular school
- We're presenting the model as a tool. Flexibility and user customization is as big a part of this deliverable as the optimal result.
- The solution generated by this model is NOT designed to be the de-facto solution to the problem of low graduation rates. Rather, this model should be paired alongside a wholistic approach by the administration. People are not machines.

Chosen Variables For Our Model

- Percent of the students in advanced classes (pAC)
- Demographic disparity between students and teachers (dD)
- Percent of male teachers (pMt)
- Student to teacher ratio (sTR)
- Average teacher salary (aTS)
- Turnover rate (tR)
- Special education rate (sER)
- Median house cost (mHC)

Regression And Optimization

- We're using a regression line-of-best-fit to motivate our objective function.
- The regression will produce a linear function of form

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ \dots + \beta_m x_m$$

- We'll rearrange the coefficients and variables

$$C = \begin{matrix} \beta_1 \\ \beta_2 \\ \dots \\ \beta_m \end{matrix} \qquad X = \begin{matrix} x_1 \\ x_2 \\ \dots \\ x_m \end{matrix}$$

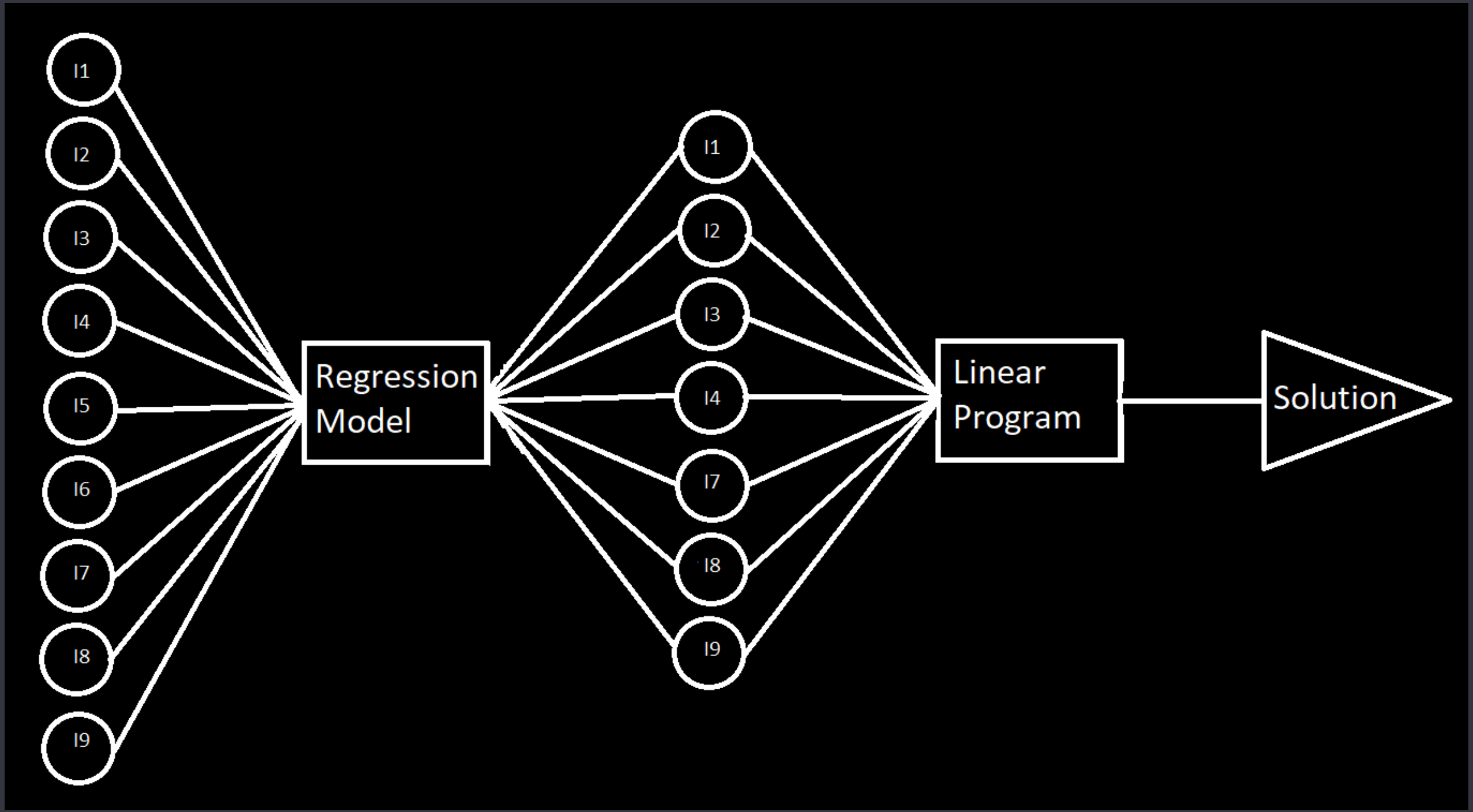
- To formulate the objective function to the linear program

Max $C^T X$

S.T various constraints

Regression And Optimization

Future Enhancements: Bridging the gap, no manual work.



Linear Regression Model For Objective Function

- Recall the objective function comes from the regression equation:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ \dots + \beta_m x_m$$

where y is the graduation rate, $x_i's$ are the variables used in the model, and the coefficients $\beta_i's$ represent the contribution of each variable to the graduation rate based on a linear regression model.

- We use a Least Absolution Deviation (LAD) Regression in order to generate this model.
  - This form of regression is robust to outliers in data.

Construction Of The Objective Function

- The full regression equation from the data has the form:

$$y_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \cdots + \beta_m x_{i,m}$$

- Constructing this equation we define the following:
  y_i is the graduation rates for each district considered in the analysis
  x_{ij} is the observed variables, j, for the i^{th} district with $x_{i0} = 1$,and
  β_j is the variable representing the coefficient for the j^{th} variable

- i ranges from 1 to n, where n is the number of districts in our sample
  j ranges from 0 to m, where m is the number of variables considered in the model

Construction Of The Objective Function

- The LP then to generate this regression equation has the form:

$$\min \sum_{i=1}^{n} t_i$$

$$s.t.$$

$$-t_i \leq y_i - \sum_{j=0}^{m} \beta_j x_{ij} \leq t_i \,, i = 1, \dots, n$$

Results Of Objective Function

- $y = 98.33 + 46.26pAC - 19.3542dD - 8.80pMT - 0.047sTR - 7.4 \cdot 10^{-6}aTS - 5.2 \cdot 10^{-6}mHC + 1.54tR - 0.30sER$
- Interpretation of some of these values:
  - 98.33 is the expected graduation rate if all other values are zero, not useful in final objective function.
  - $46.26pAC$ represents that an increase of 1% (0.01) in Advanced Classes increases the expected graduation rate by 0.4626%, similar to dD, pMT, and tR.
  - $-0.047sTR$ represents that an increase of 1 student per teacher drops graduation rate by 0.047%, similar to sER.
  - $-7.46 \cdot 10^{-6}aTS$ is of important note, it represents teach salary and that a change in average salary of $100,000 only amounts to a change in graduation rate of 0.746%, similar to mHC.

Assumptions About The Test School

- Assumed all of the variable’s starting value was the same as the average value for the state
  - Percent in Advanced class: 7.33%; Demographic disparity: 42.662%; Percent of male teachers: 24.111%; Student to teacher ratio: 25.935; Average teacher salary: $45,251.67; Median house cost: $257,565.37; Turnover rate: 20.821%; Special education ratio: 0.313
- Assumed we have 2,000 students, 100 classrooms, and that the classrooms can fit at most 30 students.
- We considered two different scenarios, where we had $150,000 of extra budget, and when we did not have any extra budget.

Constructed Constraints For Test School

- Restrictions:
  - $-7.334\% \leq pAC \leq 10\%$
  - $-(\mathrm{tR} + 20.82\%) \leq dD \leq \mathrm{tR} + 20.82\%$
  - $-42.66\% \leq dD \leq 57.34\%$
  - $-(\mathrm{tR} + 20.82\%) \leq pMT \leq \mathrm{tR} + 20.82\%$
  - $-24.11\% \leq pMT \leq 75.89\%$
  - $-6 \leq sTR \leq 4$
  - $-20.83\% \leq tR \leq 30\%$
  - $0 \leq sER \leq 0.2$

Budget Constraint For Test School

- $\$1{,}842{,}000\ pAC - \$135{,}753\ sTR + \$750{,}000\ tR + \$162{,}580\ sER \leq 150{,}000$(or 0)
  - $\$1{,}842{,}000\ pAC$: Each additional student in Advanced classes results in an increased cost of $921 and since 1% of the student body of our mock school is 20 students this gives us a rate or change of 1,842,000. [5]
  - $\$750{,}000\ tR$: An increase in turnover requires extra funds to, and increase by 1% is one additional teacher which requires $7,500. [3]
  - $\$162{,}580\ sER$:Each additional student in special education classes results in an increased cost of $1,478 so an increase in the ratio of students in special education by 1 would result in $162,580 increased budget. [5]
  - $-\$135{,}753\ sTR$: This is our only negative coefficient since if the student to teacher ratio is increased, there can be less teachers employed and thus saves the school money. For our mock school if the ration is increased by 1 then there needs to be 3 less teachers, and since their salary is $45,251.67 the school saves $135,753

Actual Model

```
#Remember, there is relevant information not included in this model.
#From the regression model we've determined coefficients (parameters)
#for the linear model we're optimizing.

#Parameters
#These coefficients are generated by the regression model
param percentAdvancedClasses;
param demographicDispartiy;
param percentMaleTeachers;
param studentTeacherRatio;
param turnoverRate;
param specialEducationRatio;

#Variables
#All these variables use the same language as their parameter counterparts.
#A change in variable represents change to the value, not
var pAC; #percentAdvancedClasses
var dD; #demographicDispartiy
var pMT; #percentMaleTeachers
var sTR; #studentTeacherRatio
var tR; #turnoverRate
var sER; #specialEducationRatio
```

```
subject to c11: 1842000*pAC - 135753*sTR + 750000*tR + 162580*sER <= 150000;
#The above references the budgetary constraint
#Assumed an increase of $150k to budget, these are extraFunds
#Coefficients determined using a variety of sources on budgetary impacts

#Objective function
maximize extraFunds:
percentAdvancedClasses*pAC + demographicDispartiy*dD + percentMaleTeachers*pMT
+ studentTeacherRatio*sTR + turnoverRate*tR + specialEducationRatio*sER;
#The coefficients generated by the regression file correspond to the parameters here.
```

Actual Model

```
#Constraints
#The reasoning behind these constraints is too verbose to variabilize
subject to c1: -.07334 <= pAC <= .1;
#The above references limits on percentAdvancedClasses
#Starting coefficient of .07334
#The minimum value of pAC is 0%
#The maximum value of pAC is 17%, determined by 'reasonable' maximum

subject to c2: dD - tR <= .2082;
subject to c3: dD + tR >= -.2082;
subject to c4: -.4266 <= dD <= .5734;
#The above references limits on demographicDisparity
#Starting coefficient of .4226
#The maximum total change of dD is limited by turnover rate
#The minimum value of dD is 0%

subject to c5: pMT - tR <= .2082;
subject to c6: pMT + tR >= -.2082;
subject to c7: -.2411 <= pMT <= .7589;
#The above references limits on percentMaleTeachers
#Starting coefficient of .2411
#The maximum total change of pMT is limited by turnover rate
#The minimum value of pMT is 0%
```

```
subject to c8: -6 <= sTR <= 4;
#The above references limits on studentTeacherRatio
#Starting ratio 1 teacher per 26 students
#The maximum ratio is 1 teacher per 30 students, determined by Colorado/federal law
#The minimum ratio is 1 teacher per 20 students, determined by test schools' classroom count

subject to c9: -.2082 <= tR <= .3;
#The above references limits on turnoverRate
#Starting coefficient of .2082
#The maximum turnover rate is 50%, determined by a 'reasonable' maximum
#The minimum turnover rate is 0%

subject to c10: 0 <= sER <= .2;
#The above references limits on specialEducationRatio
#The starting ratio is assumed to be the minimum, .313 determined by Colorado/federal law
```

Results

Budget Increase $150,000

- pAC:(Percent Advanced Classes) +10% to 17.33%
- dD:(Demographic Disparity) -43% to 0%
- pMT:(Percent Male Teacher) -24% to 0%
- sTR:(Student Teacher Ratio) +1.9 to 27.8
- tR:(Turnover Rate) +30% to 50.8%
- sER:(Special Education Ratio) +0.0
- Expected Graduation increase +15.38%

Budget Increase $0

- pAC (Percent Advanced Classes) +10% to 17.33%
- dD (Demographic Disparity) -43% to 0%
- pMT (Percent Male Teacher) -24% to 0%
- sTR (Student Teacher Ratio) +3.0 to 28.9
- tR (Turnover Rate) +30% to 50.8%
- sER (Special Education Ratio) +0.0
- Expected Graduation Increase +15.32%

Holes In Results/ Needs More Research

- A few of the results show that more research is needed:
  - That this model suggests the percentage of male teachers is dropped to zero is controversial and shows that more research is needed for the impact of male/female teachers in the classroom.
  - Special Education unchanged is a limitation of the model and the true costs associated with changing this value needs additional research.
  - Given that Student Teacher ratio has a very low priority, this potentially is a flaw in the models processing of the data.

Limitations And Potential Extensions

- Use for several different states
  - Compare different objective functions and different solutions for different states
- Add more variables
  - Teacher experience, after school programs, parent education level, etc
- Account for the potential interrelation of certain variables
- Constraints are sometimes formed off of old research that should be updated
- Extend to include non-linear constraints and objective
- Streamline input process

References And Resources

1. AMPL: A Modeling Language for Mathematical Programming by Robert Fourer, David M. Gay, and Brian W. Kerninghan, 2nd edition, 2003 http://ampl.com/resources/the-ampl-book/chapter-downloads/
2. Colorado Department of Education: Colorado Education Statistics. (n.d.). Retrieved from https://www.cde.state.co.us/cdereval
3. Kavanagh, D. (2016, April 13). Teacher Turnover Costs How Much?!?!?! Retrieved November 27, 2020, from https://www.sais.org/news/283215/Teacher-Turnover-Costs-How-Much.htm#:~:text=The%20cost%20of%20teacher%20turnover,year%20(Carroll%2C%201)
4. Olberding, E. (n.d.). Linear Regression as Linear Programming. Retrieved from http://math.ucdenver.edu/~sborgwardt/wiki/index.php/Linear_Regression_as_Linear_Programming
5. Roza, M. (2009, August 16). Breaking Down School Budgets. Retrieved November 27, 2020, from https://www.educationnext.org/breaking-down-school-budgets-2/
6. Vanderbei, R. J. (2007, October 17). Linear Programming: Chapter 12 Regression. Lecture presented in NJ, Princeton. Retrieved from https://vanderbei.princeton.edu/542/lectures/lec9.pdf

Any Questions?