Analysis of Educational Inequality in Korea
- Based on the effect of regional differences -

[Team 9] 20170459 이선오, 20204463 이선구, 20170151 김재현, 20170432 윤웅노, 20203131 김은섭

1 Abstract

It is a well-known fact that there are educational inequalities between metropolitan and non-metropolitan cities in Korea. However, the KICE(Korea Institute for Curriculum and Evaluation) doesn't provide detailed data on the annual changes of educational gaps, to alleviate hierarchical ranking structure by regions. For this reason, it's not easy to quantify educational inequality.

This paper contains the R-based analysis of educational inequality. Our main focus is to answer the questions 1) whether educational inequality exists in Korea, 2) how it exists, and 3) which factor matters to the problem.

2 Introduction

According to article 31 of the Korean Constitution, everyone has the right to get a proper education regardless of their background. The right to education is also a fundamental right guaranteed by the law. However, contrary to our constitution law, the quality and quantity of education seem to vary from area to area. Most of all, these inequalities are getting more severe in a response to the increase of private education.[1] Nowadays, there are clear difference between cities that have 'higher enthusiasm' for education and those that don't.

Especially, Seoul shows a large gap compared to other regions. Educational differences between autonomous districts in Seoul are also wider than in other cities. As the capital of Korea, Seoul is a very big, metropolitan city and has many top great colleges. Thus, people often use the word 'in-Seoul' when they talk about the college entrance exams. It means that Seoul has a unique characteristic that distinguishes the city from other cities.

To alleviate the problem, it is necessary to figure out these imbalances in detail. Therefore, we decided to conduct big-data analysis to show the effects of regional differences in education, by using many statistical methods and visualization tools in R. We expect that works of this project can contribute to let-up on educational problems.

3 Methods

3-1. Dataset

First, we collect data of the number of students by region from National Statistical Office of Korea(http://kostat.go.kr/). Next, we gathered educational data based on the 3 aspects: educational opportunity, educational progress, and academic achievement.[2]

(1) Academic achievement

- Scores of college entrance exam(Su-neung) by region, from 2015 to 2019.

[1] Suehye Kim., et al. "Student Achievement Gap Trends on OECD Equity Indicators: Evidence from Seoul Education Longitudinal Study", SELS: 2010~2018, 2020, pp. 128

[2] HyeKyung Park, "Legislative Evaluation of the “Proposal for the Elimination of the Education Gap", Journal of Legislative Evaluatio 11, 2017, pp. 96

We collected data from Korean Educational Statistics Service (KESS, https://kess.kedi.re.kr)

- The number of students going to SNU(Seoul National University) by region, 2019. We collected data from government audit data of SNU.

(2) Educational progress

- Academic interruption rate: It means data of the disruption to students' academic works. We collected this data from KESS.

(3) Educational opportunity

- Private education: Academy and academy pay rate per students. We collected these data from KESS and Korean Statistical Information Service (KOSIS, https://kosis.kr/index/index.do)
- Public education: Teacher/classroom rate per students. We collected these data from KESS.

3-2. Research Questions

(1) Does educational inequality exist?

This question is answered by performing ANOVA on the college entrance exams of 17 regions. Since we know the mean, the standard deviation, and the number of students in each region, we can still perform ANOVA by using "ind.oneway.second" function in "rpsychi" library. This function conducts a one-way design with independent samples, giving us the ANOVA table. The ANOVA result gives the F-value (thus the p-value) of the data, which we can determine whether the data is significant.

(2) How does educational inequality exist?

This question is answered by performing clustering in 17 regions. Grouped by 3 features (academy rate, academic interruption rate, and entrance exam Korean score), we can see how educational inequality is distributed along these regions. Those 3 features are chosen to represent educational opportunity, educational progress, and academic achievement, respectively. We first calculate the number of clusters using the "complete" method with index Hartigan and KL, and then use both K-means clustering and PAM clustering.

(3) What factors influence educational inequality?

To answer the question, we analyzed the effects of these four factors on the entrance exam score: classroom rate, teacher rate, academy rate, and academy pay rate. For math, Korean high school education offers two different curriculum tracks: liberal arts and natural sciences, so we decided to use Korean scores data only. Among the four factors, the number of classrooms per student and the number of teachers per student represent the effect of public education. Also, the number of academies and spending on academies represent the effect of private education.

The multivariate linear regression was done to analyze each factor's influence on academic achievement.

$$(\text{exam score}) = \theta_0 + \theta_1 x_1 + \theta_2 x_2$$

In this formular, θ_n means regression coefficients of the linear regression model. By analyzing the regression coefficients, we can find a relationship between exam scores and each factor.

(4) Additional analysis

First, we explained the entrance exam results with graphs additionally. Because some of the data we gathered is from 2015 to 2019, when we plot data on a 2D map graph, it is hard to get intuitive information from it. Thus, we thought that 3D visualization might be a more meaningful approach to make graphs easy for analysis.

From the visualization, we found out

that some data on educational inequality have changed over time. so secondly, we analyzed data to know how the gap of educational inequality is changed qualitatively.

Lastly, we tried to analyze exam results for the group of 'Gu', which means a smaller administrative area than the city. As we mentioned in the introduction, data from Seoul shows different trends compared with other cities. We thought that is because Seoul is a too massive city. As the capital of Korea, groups of 'Gu' of Seoul have a similar population with other smaller cities.

4 Analysis Results

4-1. ANOVA

In Figure 1, "df" stands for the degree of freedom, and "F" stands for F-value. We can see that F-value is 287.68. Using the F distribution with parameter 16 and 387203, the 0.05 critical value is 2.35. Since the F-value is much greater than the critical value, we can conclude that the college entrance exam scores are different among 17 regions in Korea.

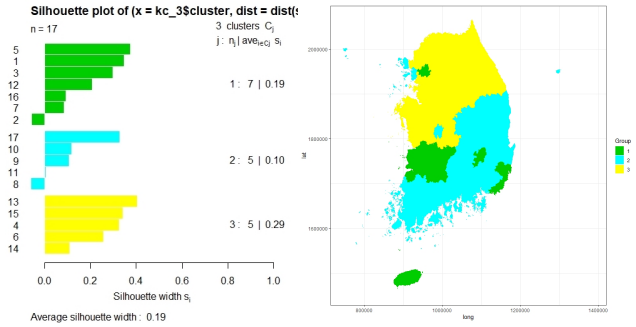
```
$anova.table
                    SS     df     MS      F
Between (A)    1753479     16 109592 287.68
Within       147505041 387203    381
Total        149258519 387219
```

Figure 1. Result of ANOVA

4-2. Clustering

Cities in Korea are clustered with 3 features: academy rate, academic interruption rate, and entrance exam Korean score. We tried to use various clustering algorithms including K-means and PAM. Similar results were found in the analyses, and the reported result in Figure 2 is derived from the K-means clustering. The result shows below three things.

First, the number of clusters tend to be 3. Second, Seoul, Busan, Daegu tend to be in one group, Gyeonggi-do and Chungcheong-do tend to be in another group, and Daejeon and Gyeongsang-do tend to be in the other group. Third, the first group consists of Seoul and most of the metropolitan cities, which have the highest academy rate and exam score with the lowest academic interruption rate. Therefore, we can conclude that Seoul and most of the metropolitan cities have a better educational environment than others.



(a) Silhouette plot and map visualization

	group1	group2	group3
korean	99.07143000	96.06000000	95.8800000
academyNumRate	0.01653657	0.01306660	0.0135132
interruptedRatio	0.01188808	0.01679778	0.0132602

(b) Mean value of exam score/academy rate/academic interruption rate for each clusters

Figure 2. Result of the K-means clustering

4-3. Linear Regression

Figure 3 shows the result of the linear regression analysis on the Korean exam score. In Figure 3(a), coefficients of classroom rate and teacher rate are both smaller than 0, but all of those p-values are also much higher than 0.05. So, we can say that there is no significant relationship between Korean exam score and factors of public educations.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-9.179e-16	2.273e-01	0.000	1.00
classroomNumRate	-2.102e-01	5.159e-01	-0.407	0.69
teacherNumRate	-2.843e-01	5.159e-01	-0.551	0.59

(a) Coefficients of factors representing public education

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	-1.058e-15	2.063e-01	0.000	1.0000	
academyNumRate	1.939e-01	2.165e-01	0.896	0.3856	
academyPayRate	5.385e-01	2.165e-01	2.487	0.0261	*

(b) Coefficients of factors representing private education

On the contrary, Figure 3(b) shows that the academy pay rate gets a p-value smaller than 0.05, and have a meaningful correlation to exam score. Therefore, we concluded that spending on private education is important to get better exam scores.

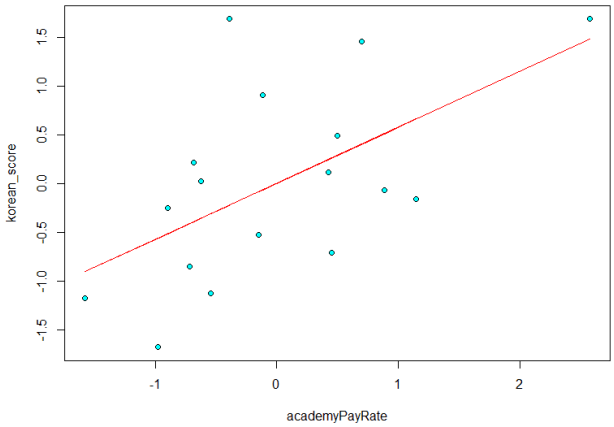


Figure 4. Linear model of academy pay rate and Korean exam score

Because of the large number of students, the number of teachers and classrooms per student in metropolitan cities is smaller than in non-metropolitan cities. That could be one of the reasons why regression coefficient of teacher rate and classroom rate have a negative value.

4-4. Additional Analysis

We used "rayshader" package which converts graphs made by ggplot to 3D. Figure 5 is based on Korean exam data from 2015 to 2019. The 3D graph is much more intuitive than a 2D graph.

Figure 5. Korean exam data from 2015 to 2019

We checked all entrance exam results for 5 years. The gap of Korean score didn't change, math A score gap increases significantly. But, the math B score gap was reduced a lot too. The government policy to divide math courses as A and B was changed in 2016. Because of that, there was a huge difference between before 2017 and after 2017. However, the math A score gap was increased after 2017 despite the math B score gap remains similar. Also, there is a difference of deviation only smaller than 0.05% of math A and B which means there is an increasing educational inequality trend on math A. Figure 6 shows detailed values about them.

mean_math_B_score_L	mean_math_B_score_H	mean_math_A_score_L	mean_math_A_score_H	mean_kor_score_L	mean_kor_score_H	year
91.58	103.66	97.58	102.62	96.96	102.13	2015
90.70	103.90	97.42	101.70	96.74	101.12	2016
97.44	101.86	92.18	103.10	95.72	100.84	2017
97.34	101.44	91.36	102.52	95.28	100.34	2018
96.92	100.96	91.06	101.90	94.84	99.94	2019

Figure 6. Lowest and highest means of exam scores, from 2015 to 2019

Based on the location of each Gu, Seoul could be divided into 5 areas. We get the sum of entrance exam scores for 2015. Because of government policy, there is no up-to-date data. The graph shows three area 도심, 동북, 서북 have a similar mean value of exam scores, but 서남 is lower than others, and 동

남 have much higher scores. So, we could find out there is a huge educational inequality gap in Seoul too.

These inequalities are shown visually in Figure 7.

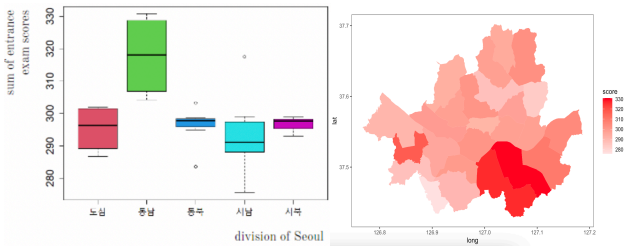


Figure 7. Analysis by regions in Seoul

5 Conclusion

As a result of statistical analysis, in Korea, there were significant educational inequalities by regions. Especially, the gap of academic achievements between metropolitan and non-metropolitan cities was large. Seoul and most of the metropolitan cities show a better educational environment than others. Also, our project suggests that this educational environment is to a great extent the products of private education.

According to the linear regression model, factors representing public educations aren't important to get better exam scores. Because there is a definite result that only spending on the academy has a significant relationship with academic achievements, it seems to be true that private education is leading the exam scores.

In Korea, the word 'high enthusiasm for education' is often used for the cities that have many famous private education institutes. In Seoul, there is even a street that is now known as 'the first avenue' for private education. In consideration of those disproportions, policymakers should make changes in an effort to help lower the incentive of private lessons.

6 References

1. Suehye Kim., et al. "Student Achievement Gap Trends on OECD Equity Indicators: Evidence from Seoul Education Longitudinal Study", SELS: 2010~2018, 2020. 6.
2. SungBae Oh, "Analysis of the factors creating education districts in scholastic achievements: Effect of the regional differences and pupil mobility", The Journal of Korean Education, 2004.
3. HyeKyung Park, "Legislative Evaluation of the “Proposal for the Elimination of the Education Gap", Journal of Legislative Evaluatio 11, 2017. 6.