

Final_project.Rmd

Introduction

The main goal of our reserch is finding variables which affect the student's ability generally through comparison of test scores from differend countries. Especially, we will focus on how inerenet accessibility bring effects on student's ability. Recently, some reserchers found that internet accessibility might bring reverse affect on student's ability because they spend more time on net surfing and gathering information through internet without thinking.

Methodology

To analyze our research theme, we used the Data from Programme for International Student Assessment(PISA) that conducts assesment about the We picked up the result of **Mathmatics, Reading and Problem Solving** that is conductet in 2012. Also, we picked the result of 2003 Problem solving as well, which is a direct assesment of life competencies that apply across different areas of the school curriculum. This data is benefitial to analyze student's ability that is not measured by academic ability.

Picked up coutries are depended on PISA data avairability,

Then, we have analyzed the correlation between these variables.

Data gathering and merging process

The first dataset is from **PISA**, and the second from **the World Bank**. Both datasets are open and can be found in their respective webpages.

Data Source

1. PISA: We downloaded and picked the following three datas up to use as dependent variables:

Dependent Variable	Variable Name	Description
DV	math	Mathmatics mean score(2012)
DV	reading	Readind mean score(2012)
DV	ps	Prolem Solving mean score(2012)

Indivendent Variable	Variable Name	Description
IV	GDPperc	GDP per Capita (current US\$)
IV	expend	Government expenditure on education, total (% of GDP)
IV	pop	Population, total
IV	popd	Population density (people per sq. km of land area)
IV	rteacher	Primary school pupil-teacher ratio is the average number of pupils per teacher in primary school
IV	eyear	Number of years that children are legally obliged to attend school
IV	internet	internet users (per 100 people). Internet users are individuals who have used the Internet (from any location) in the last 12 months. Internet can be used via a computer, mobile phone, personal digital assistant, games machine, digital TV etc.
IV	mobile	Mobile cellular subscriptions (per 100 people)

log(GDPperc)	30.639***	29.944***	16.218*
	(7.175)	(7.404)	(8.606)
log(pop)	5.746	10.372**	8.667*
	(3.764)	(4.617)	(4.288)
popd	0.004*	0.003	0.003
	(0.002)	(0.002)	(0.002)
rteacher	-3.424**	-1.065	
	(1.530)	(1.667)	
eyear	-16.169***	-11.386**	
	(4.192)	(4.211)	
expend	4.506	-2.768	
	(5.686)	(5.851)	
internet	1.561**		
	(0.625)		
mobile	0.125		
	(0.223)		
(intercept)	77.963	214.289*	201.180*
	(88.520)	(101.400)	(104.076)

Observations 42 23 23

R2 0.453 0.760 0.834

Adjusted R2 0.410 0.670 0.740

Residual Std. Error 33.948 (df = 38) 24.438 (df = 16) 21.703 (df = 14)

F Statistic 10.481*** (df = 3; 38) 8.445*** (df = 6; 16) 8.817*** (df = 8; 14) =====

Note: $p < 0.1$; $p < 0.05$; $p < 0.01$

Analysis

According to the result, in the second regression, these variables are statistically significant. - Math:GDP per capita - Reading: GDP per capita, expenditure - Problem solving: GDP per capita, population, pupils-teacher rate and duration educational year in elementary

In the third regression, GDP per capita is not statistically significant any more. Instead of GDP per capita, the number of internet users and population become statistically significant.

Especially, the number of internet users is significant at 0.1% level in math and reading. We can see it bring plus effect on the both score. However, regarding the problem solving, internet is statistically significant only at 5 % level and the effect is weaker than other 2 scores.

In addition, regarding GDP per capita of math and reading, it brings the minus effect on the both scores after the variable internet and mobile are added, though it still brings the plus effect on problem solving. Even though it is not statistically significant for math and reading, but it might be possible to infer if the internet accessibility would be the same level, economically strong country's math and reading scores might be lower than economically weak countries.

Also, we noticed that about the academic ability (math and reading,) internet accessibility would bring the same impact on the scores. However, it would bring less impact on the problem solving ability.

Actually, there is some discussion that internet accessibility might bring reverse effect on student's thinking ability. Therefore, next we focus on the problem solving score.

Table 3: Table:Regression Estimates of Math Score

	<i>Dependent variable:</i>		
	Math Score		
	(1)	(2)	(3)
log(GDPperc)	29.666*** (5.794)	22.703** (9.524)	−7.883 (10.680)
log(pop)	1.642 (3.249)	7.587 (6.643)	11.018** (5.264)
popd	0.004* (0.002)	0.005 (0.003)	0.003 (0.004)
rteacher		−4.035 (2.616)	−0.956 (2.173)
eyear		−8.544 (5.699)	−7.042 (4.473)
expend		11.785 (7.097)	−2.215 (6.412)
internet			2.522*** (0.602)
mobile			0.452 (0.319)
(intercept)	145.762 (87.914)	216.904 (156.559)	240.336 (143.650)
Observations	63	34	34
R ²	0.388	0.451	0.690
Adjusted R ²	0.357	0.328	0.591
Residual Std. Error	41.891 (df = 59)	45.818 (df = 27)	35.750 (df = 25)
F Statistic	12.455*** (df = 3; 59)	3.689*** (df = 6; 27)	6.963*** (df = 8; 25)

Note:

*p<0.1; **p<0.05; ***p<0.01

Table 4: Table:Regression Estimates of Reading Score

	<i>Dependent variable:</i>		
	Reading Score		
	(1)	(2)	(3)
log(GDPperc)	27.480*** (5.174)	20.996** (7.897)	-2.677 (9.208)
log(pop)	3.107 (2.901)	7.920 (5.508)	10.655** (4.538)
popd	0.002 (0.002)	0.003 (0.002)	0.002 (0.003)
rteacher		-2.788 (2.169)	-0.350 (1.873)
eyear		-7.206 (4.725)	-5.995 (3.857)
expend		13.659** (5.884)	2.736 (5.528)
internet			1.960*** (0.519)
mobile			0.383 (0.275)
(intercept)	146.816* (78.509)	187.641 (129.802)	198.843 (123.845)
Observations	63	34	34
R ²	0.365	0.463	0.673
Adjusted R ²	0.332	0.343	0.568
Residual Std. Error	37.409 (df = 59)	37.987 (df = 27)	30.822 (df = 25)
F Statistic	11.284*** (df = 3; 59)	3.877*** (df = 6; 27)	6.419*** (df = 8; 25)

Note:

*p<0.1; **p<0.05; ***p<0.01

Table 6: Table:Detarminants of Student's Ability

	<i>Dependent variable:</i>					
	math			reading		
	(1)	(2)	(3)	(4)	(5)	(6)
log(GDPperc)	29.666*** (5.794)	22.703** (9.524)	-7.883 (10.680)	27.480*** (5.174)	20.996** (7.897)	-2.677 (9.208)
log(pop)	1.642 (3.249)	7.587 (6.643)	11.018** (5.264)	3.107 (2.901)	7.920 (5.508)	10.655** (4.538)
popd	0.004* (0.002)	0.005 (0.003)	0.003 (0.004)	0.002 (0.002)	0.003 (0.002)	0.002 (0.003)
rteacher		-4.035 (2.616)	-0.956 (2.173)		-2.788 (2.169)	-0.350 (1.873)
eyear		-8.544 (5.699)	-7.042 (4.473)		-7.206 (4.725)	-5.995 (3.857)
expend		11.785 (7.097)	-2.215 (6.412)		13.659** (5.884)	2.736 (5.528)
internet			2.522*** (0.602)			1.960*** (0.519)
mobile			0.452 (0.319)			0.383 (0.275)
(intercept)	145.762 (87.914)	216.904 (156.559)	240.336 (143.650)	146.816* (78.509)	187.641 (129.802)	198.843 (123.845)
Observations	63	34	34	63	34	34
R ²	0.388	0.451	0.690	0.365	0.463	0.673
Adjusted R ²	0.357	0.328	0.591	0.332	0.343	0.568
Residual Std. Error	41.891 (df = 59)	45.818 (df = 27)	35.750 (df = 25)	37.409 (df = 59)	37.987 (df = 27)	30.822 (df = 25)
F Statistic	12.455*** (df = 3; 59)	3.689*** (df = 6; 27)	6.963*** (df = 8; 25)	11.284*** (df = 3; 59)	3.877*** (df = 6; 27)	6.419*** (df = 8; 25)

Note:

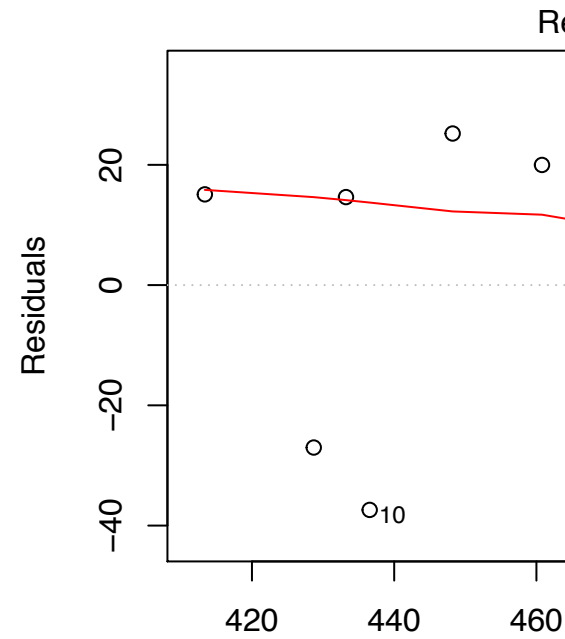
Table 7: Table:Regression Estimates of Three Modeles

	<i>Dependent variable:</i>		
	math	reading	ps
	(1)	(2)	(3)
log(GDPperc)	-7.883 (10.680)	-2.677 (9.208)	16.218* (8.606)
log(pop)	11.018** (5.264)	10.655** (4.538)	8.667* (4.288)
popd	0.003 (0.004)	0.002 (0.003)	0.003 (0.002)
rteacher	-0.956 (2.173)	-0.350 (1.873)	-1.065 (1.667)
eyear	-7.042 (4.473)	-5.995 (3.857)	-11.386** (4.211)
expend	-2.215 (6.412)	2.736 (5.528)	-2.768 (5.851)
internet	2.522*** (0.602)	1.960*** (0.519)	1.561** (0.625)
mobile	0.452 (0.319)	0.383 (0.275)	0.125 (0.223)
(intercept)	240.336 (143.650)	198.843 (123.845)	201.180* (104.076)
Observations	34	34	23
R ²	0.690	0.673	0.834
Adjusted R ²	0.591	0.568	0.740
Residual Std. Error	35.750 (df = 25)	30.822 (df = 25)	21.703 (df = 14)
F Statistic	6.963*** (df = 8; 25)	6.419*** (df = 8; 25)	8.817*** (df = 8; 14)

Note:

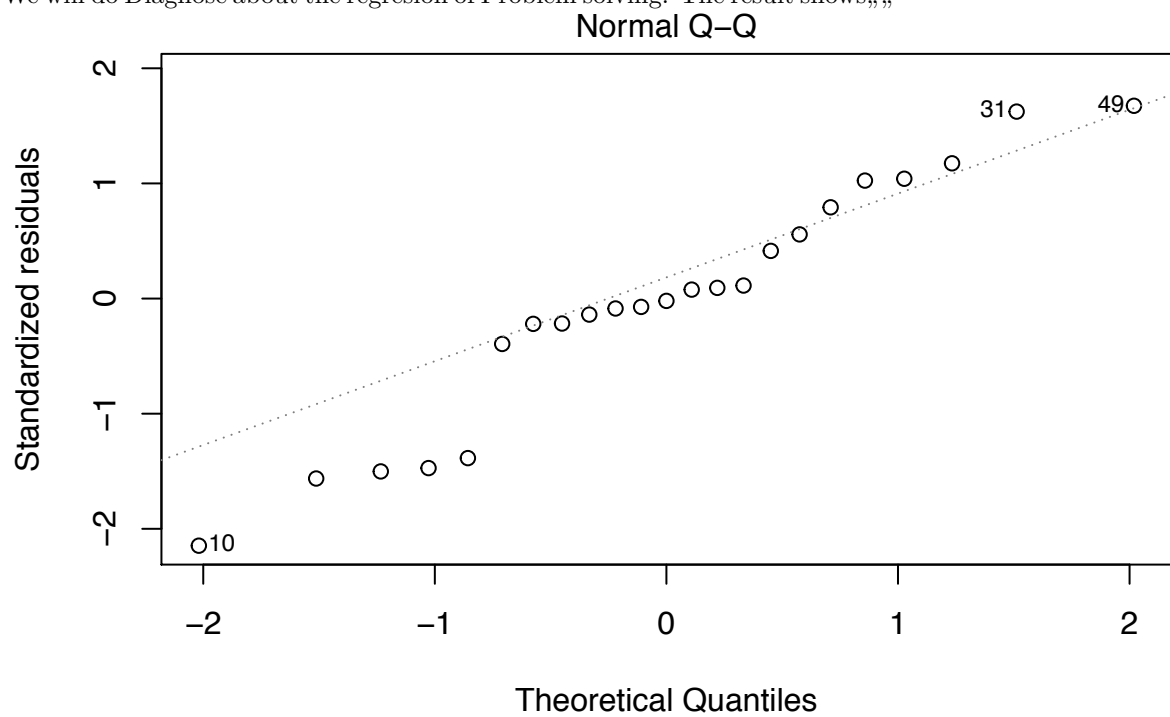
*p<0.1; **p<0.05; ***p<0.01

Correlation between Problem Solving and Internet

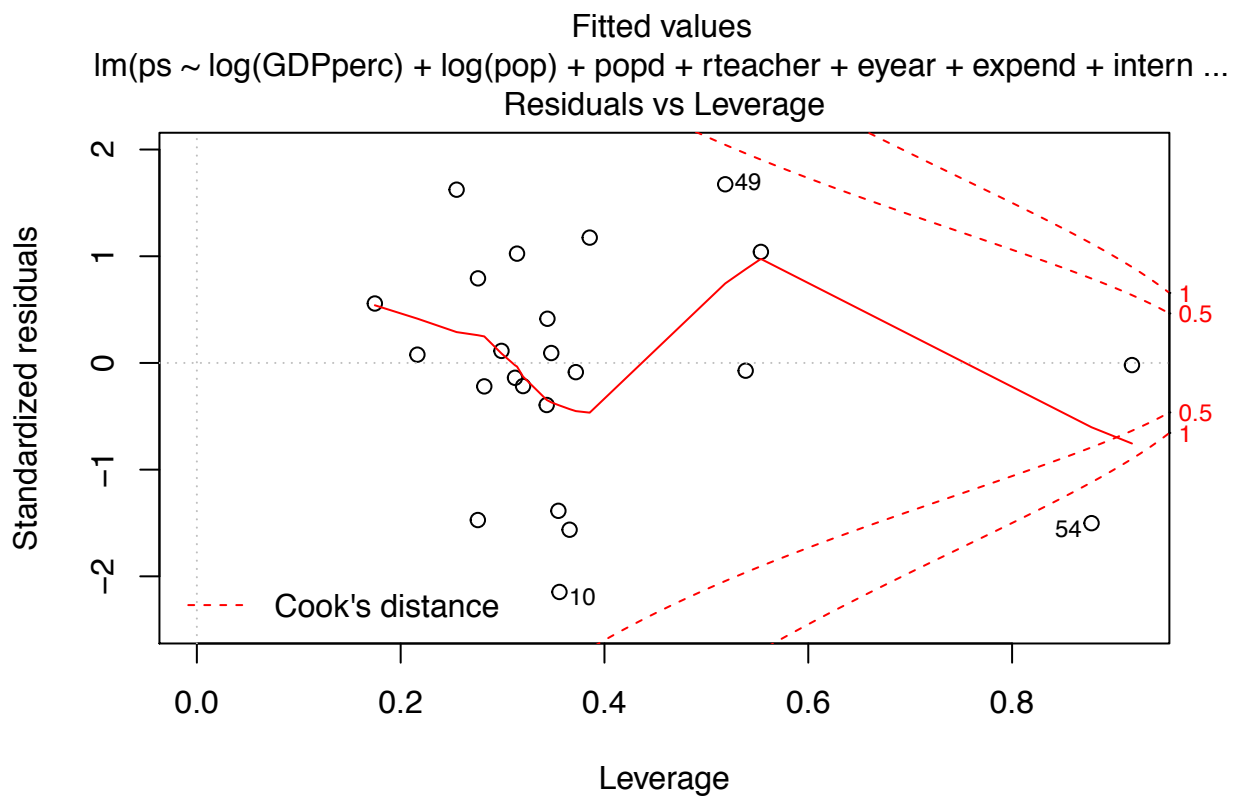
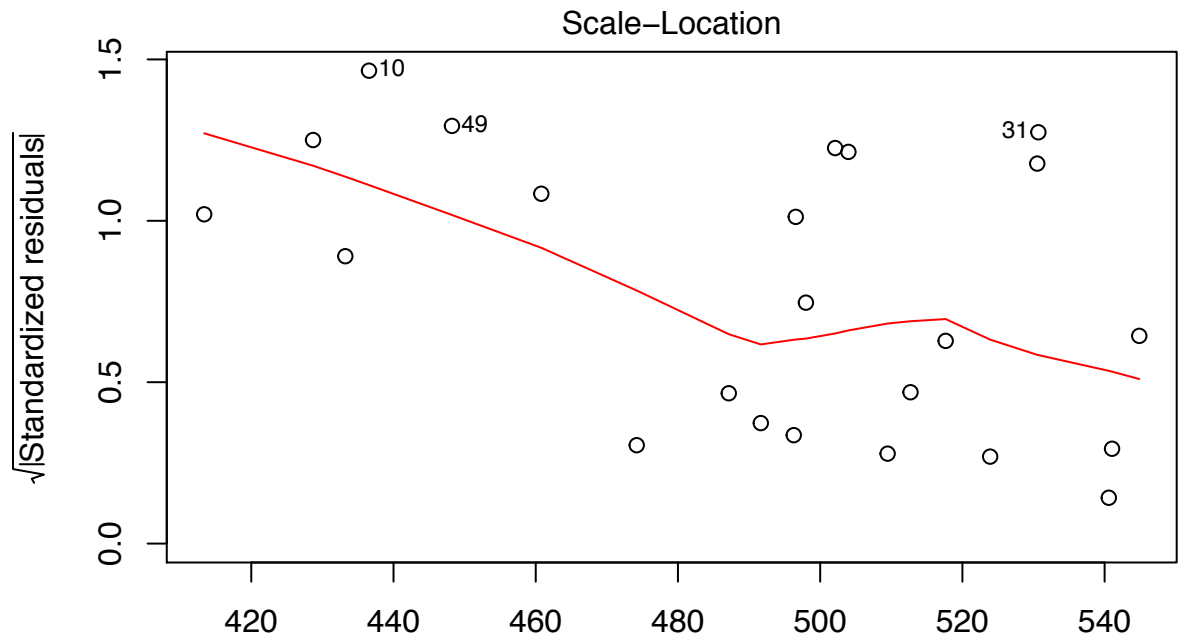


We will do Diagnose about the regresion of Problem solving. The result shows,,,

$\text{lm}(\text{ps} \sim \log(\text{GDPperc}) + \log(\text{pop}))$



$\text{lm}(\text{ps} \sim \log(\text{GDPperc}) + \log(\text{pop}) + \text{popd} + \text{rtteacher} + \text{eyear} + \text{expend} + \text{intern} \dots)$



Im(ps ~ log(GDPperc) + log(pop) + popd + rteacher + eyear + expend + intern ...

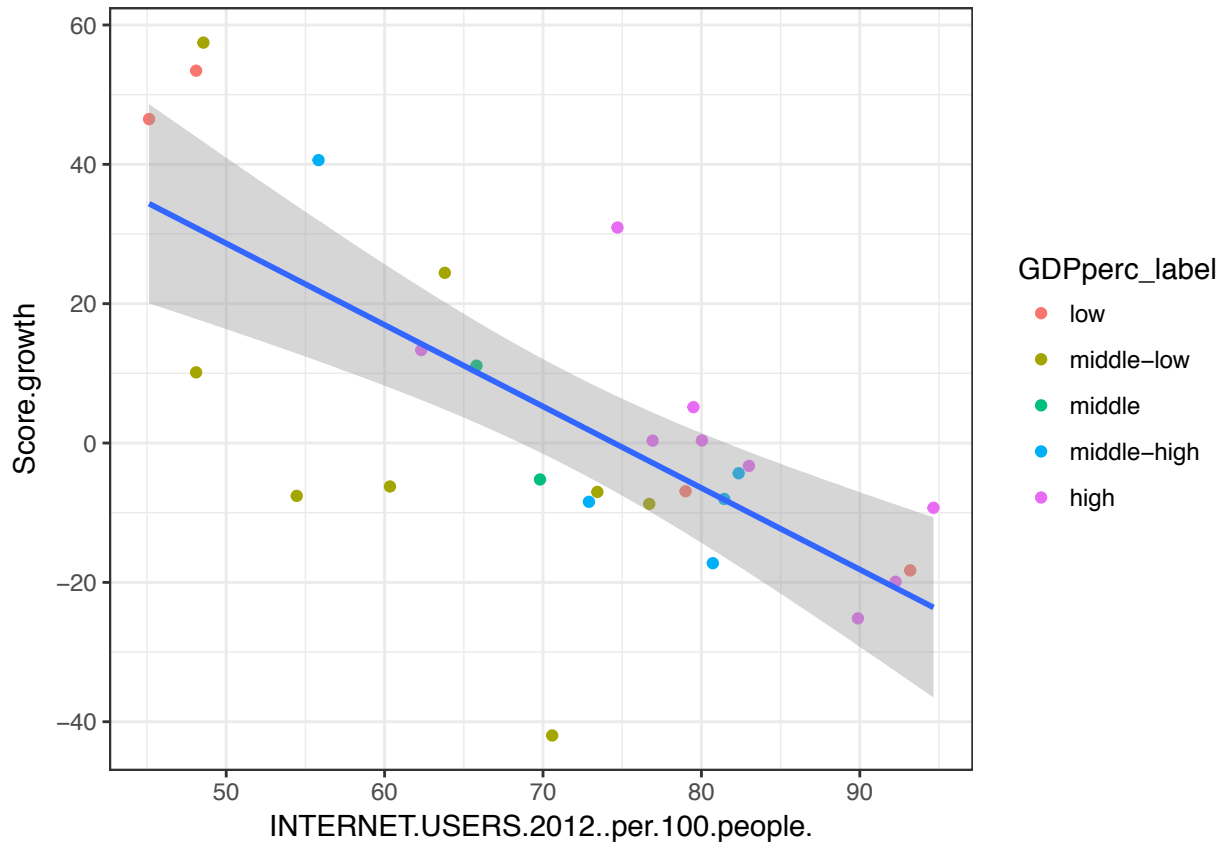
	2.5 %	97.5 %	(Intercept)	-22.041157797	424.400379451	log(GDPperc)	-2.241003905	34.676962325	log(pop)	-0.530610003	17.863716671	popd	-0.002706016	0.007889451	rteacher	-4.641069200	2.510159015	eyear	-20.416747602	-2.354732539	expend	-15.316598142	9.781025773	internet	0.221215890	2.901338214	mobile	-0.352832156	0.603257434
Min.	1755	10690	20580	31350	45720	149200																							

Analysis about score growth

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      3989  13140   24450   29960   44730   101600
```

```
## Warning: Removed 13 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 13 rows containing missing values (geom_point).
```



```
## Warning: Removed 19 rows containing missing values (geom_point).
```

