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 afffect 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)

2. The World Bank: Taking aveirability into account, we picked the following variables up as explanetory variables from the World Bank Data.

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)

Cleaned and Merged

We have cleaned and merged the relevant data of both datasets. Then, we have generate tables and figures to relate the data with the aim of addressing, as well as possible, our researching question to begin to determine whether our hypothesis is correct or not.

Multiple Regression

Summary statistics of the variables

Statistic N Mean St. Dev. Min Max

math 63 469.743 52.225 368.103 573.468
reading 63 471.675 45.780 384.151 544.600
ps 42 485.544 44.181 399.166 562.421
GDPperc 63 32,608.030 29,376.850 1,754.548 149,160.800 pop 63 36,014,700.000 60,090,250.000 36,791 314,102,623 popd 63 635.244 2,625.005 2.959 18,654.280 rteacher 51 15.061 4.335 7.444 28.016
eyear 62 10.226 1.750 6 14
expend 39 4.897 1.270 2.922 7.656
internet 63 66.580 18.362 14.520 96.210
mobile 63 128.770 33.182 79.568 289.782

Table:Regression Estimates of Problem Solving Score

	Dependent variable:			
	Problem Solving Score			
	(1)	(2)	(3)	(4)

log(GDPperc)	26.447**	39.002***	-6.536	2.227
	(10.146)	(12.141)	(17.757)	(19.536)
log(pop)	7.770	6.484	6.960	8.256
	(5.252)	(5.445)	(5.041)	(4.981)
popd	0.003	0.001	0.007*	0.003
	(0.002)	(0.002)	(0.004)	(0.004)
rteacher	-2.914	-2.135	-2.051	-0.441
	(1.860)	(1.856)	(1.808)	(1.828)
eyear	-2.907	-11.611**	-0.202	-8.845*
	(4.202)	(4.802)	(4.172)	(4.473)
expend	-8.131	-6.580		
	(7.474)	(6.747)		
internet	1.938**	2.230**		
	(0.875)	(0.944)		
mobile	-0.137	0.144		
	(0.276)	(0.251)		
(intercept)	161.860	181.296	345.194**	279.162*
	(119.071)	(126.734)	(151.692)	(146.168)
Observations	32	23	32	23
R2	0.464	0.705	0.556	0.793
Adjusted R2	0.361	0.594	0.426	0.674
Residual Std. Error	32.753 (df = 26)	27.096 (df = 16)	31.048 (df = 24)	24.289 (df = 14)
F Statistic	4.507*** (df = 5; 26)	6.371*** (df = 6; 16)	4.287*** (df = 7; 24)	6.687*** (df = 8; 14)

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

Table 3: Table:Regression Estimates of Math Score

	Dependent variable:			
	Math Score			
	(1)	(2)	(3)	(4)
log(GDPperc)	20.933**	19.279*	−19.929*	−36.782***
	(7.847)	(10.881)	(10.282)	(12.139)
log(pop)	5.468	6.384	7.355**	12.210**
	(4.187)	(6.875)	(3.365)	(4.564)
popd	0.004*	0.005	0.008**	0.007*
	(0.002)	(0.003)	(0.003)	(0.003)
rteacher	−3.751*	−3.690	−3.329**	−2.585
	(1.889)	(2.912)	(1.540)	(1.959)
eyear	−3.606	−7.555	−1.067	−6.329
	(3.898)	(5.896)	(3.183)	(3.836)
expend		8.809		−3.060
		(7.507)		(5.219)
internet			2.570***	3.699***
			(0.502)	(0.614)
mobile			−0.009	0.274
			(0.221)	(0.282)
(intercept)	266.053**	267.479	437.763***	471.732***
	(109.259)	(166.564)	(105.489)	(139.107)
Observations	50	34	50	34
R^2	0.399	0.404	0.635	0.769
Adjusted R^2	0.331	0.272	0.574	0.694
Residual Std. Error	41.402 (df = 44)	47.712 (df = 27)	33.031 (df = 42)	30.906 (df = 25)
F Statistic	5.844*** (df = 5; 44)	3.052** (df = 6; 27)	10.433*** (df = 7; 42)	10.374*** (df = 8; 25)
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 4: Table:Regression Estimates of Reading Score

	Dependent variable:			
	Reading Score			
	(1)	(2)	(3)	(4)
log(GDPperc)	19.883***	17.940*	−14.930	−25.810**
	(6.961)	(9.098)	(9.516)	(10.973)
log(pop)	7.070*	6.813	8.735***	11.460**
	(3.714)	(5.749)	(3.114)	(4.125)
popd	0.002	0.003	0.007**	0.005
	(0.002)	(0.002)	(0.003)	(0.003)
rteacher	−3.271*	−2.455	−3.151**	−1.519
	(1.676)	(2.435)	(1.425)	(1.771)
eyear	−3.269	−6.301	−1.551	−5.295
	(3.458)	(4.930)	(2.946)	(3.467)
expend		10.894*		1.507
		(6.277)		(4.717)
internet			2.154***	2.901***
			(0.464)	(0.555)
mobile			−0.169	0.252
			(0.205)	(0.255)
(intercept)	241.887**	233.204	417.458***	384.247***
	(96.925)	(139.280)	(97.636)	(125.744)
Observations	50	34	50	34
R^2	0.374	0.407	0.586	0.731
Adjusted R^2	0.303	0.276	0.517	0.645
Residual Std. Error	36.728 (df = 44)	39.896 (df = 27)	30.572 (df = 42)	27.937 (df = 25)
F Statistic	5.259*** (df = 5; 44)	3.095** (df = 6; 27)	8.494*** (df = 7; 42)	8.492*** (df = 8; 25)
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 6: Table:Detarminants of Student's Ability

	math					*De*
	(1)	(2)	(3)	(4)	(5)	(6)
log(GDPperc)	20.933**	19.279*	−19.929*	−36.782***	19.883***	17.940*
	(7.847)	(10.881)	(10.282)	(12.139)	(6.961)	(9.098)
log(pop)	5.468	6.384	7.355**	12.210**	7.070*	6.813
	(4.187)	(6.875)	(3.365)	(4.564)	(3.714)	(5.749)
popd	0.004*	0.005	0.008**	0.007*	0.002	0.003
	(0.002)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)
rteacher	−3.751*	−3.690	−3.329**	−2.585	−3.271*	−2.455
	(1.889)	(2.912)	(1.540)	(1.959)	(1.676)	(2.435)
eyear	−3.606	−7.555	−1.067	−6.329	−3.269	−6.301
	(3.898)	(5.896)	(3.183)	(3.836)	(3.458)	(4.930)
expend		8.809		−3.060		10.894*
		(7.507)		(5.219)		(6.277)
internet			2.570***	3.699***		
			(0.502)	(0.614)		
mobile			−0.009	0.274		
			(0.221)	(0.282)		
(intercept)	266.053**	267.479	437.763***	471.732***	241.887**	233.204
	(109.259)	(166.564)	(105.489)	(139.107)	(96.925)	(139.280)
Observations	50	34	50	34	50	34
R^2	0.399	0.404	0.635	0.769	0.374	0.407
Adjusted R^2	0.331	0.272	0.574	0.694	0.303	0.276
Residual Std. Error	41.402 (df = 44)	47.712 (df = 27)	33.031 (df = 42)	30.906 (df = 25)	36.728 (df = 44)	39.896 (df = 27
F Statistic	5.844*** (df = 5; 44)	3.052** (df = 6; 27)	10.433*** (df = 7; 42)	10.374*** (df = 8; 25)	5.259*** (df = 5; 44)	3.095** (df = 6;
Note:						

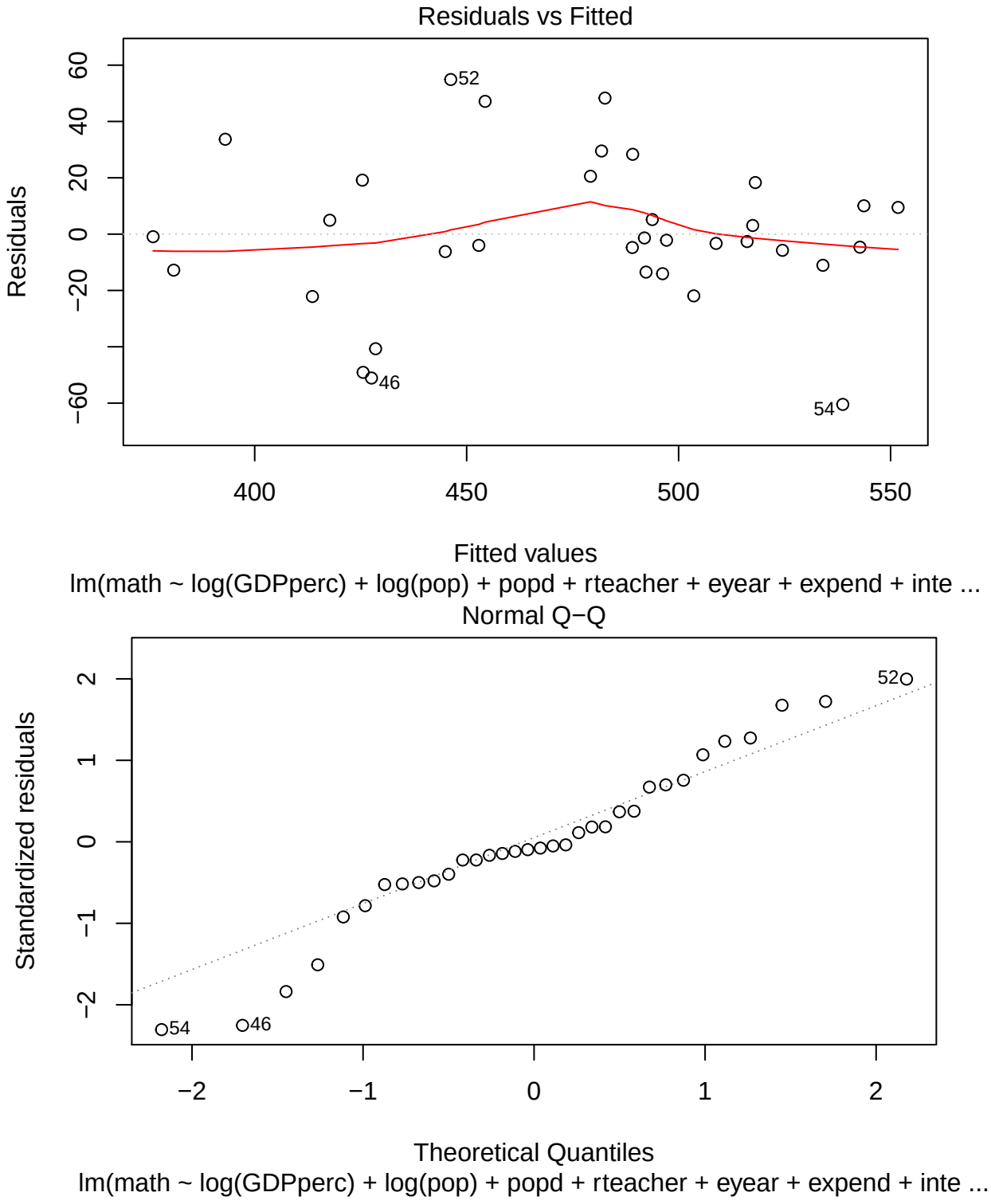
Table 7: Table:Detarminants of Student's Ability

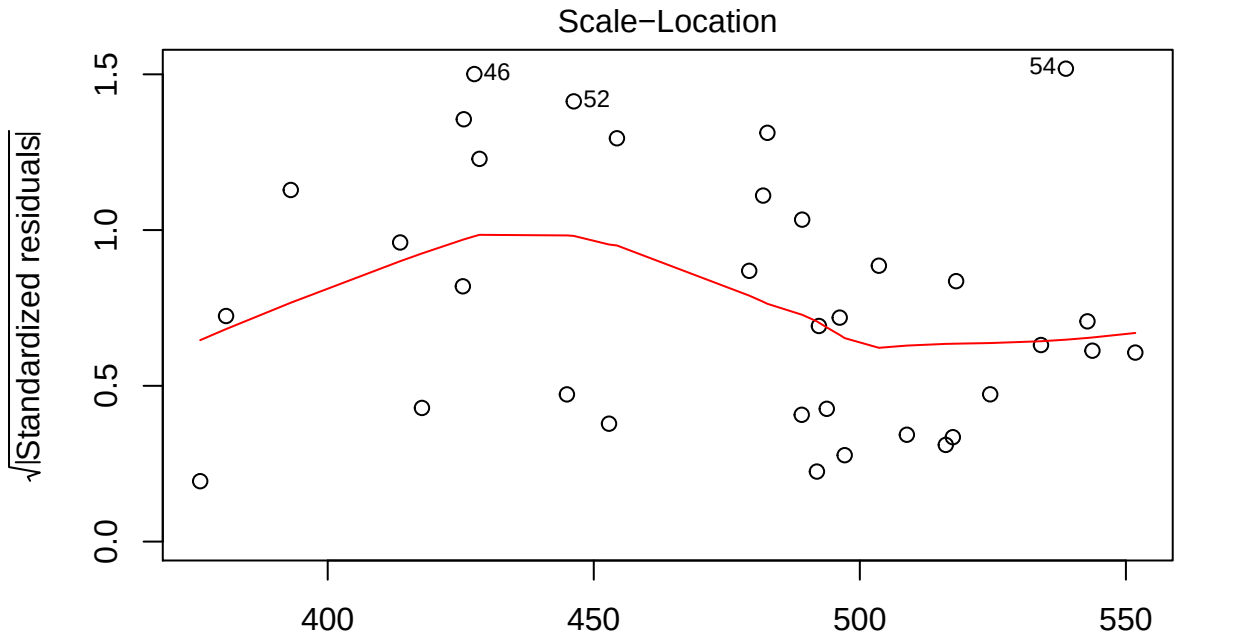
	Dependent variable:					
	math		reading		ps	
	(1)	(2)	(3)	(4)	(5)	(6)
log(GDPperc)	19.279*	−36.782***	17.940*	−25.810**	39.002***	2.227
	(10.881)	(12.139)	(9.098)	(10.973)	(12.141)	(19.536)
log(pop)	6.384	12.210**	6.813	11.460**	6.484	8.256
	(6.875)	(4.564)	(5.749)	(4.125)	(5.445)	(4.981)
popd	0.005	0.007*	0.003	0.005	0.001	0.003
	(0.003)	(0.003)	(0.002)	(0.003)	(0.002)	(0.004)
rteacher	−3.690	−2.585	−2.455	−1.519	−2.135	−0.441
	(2.912)	(1.959)	(2.435)	(1.771)	(1.856)	(1.828)
eyear	−7.555	−6.329	−6.301	−5.295	−11.611**	−8.845*
	(5.896)	(3.836)	(4.930)	(3.467)	(4.802)	(4.473)
expend	8.809	−3.060	10.894*	1.507	−8.131	−6.580
	(7.507)	(5.219)	(6.277)	(4.717)	(7.474)	(6.747)
internet		3.699***		2.901***		2.230**
		(0.614)		(0.555)		(0.944)
mobile		0.274		0.252		0.144
		(0.282)		(0.255)		(0.251)
(intercept)	267.479	471.732***	233.204	384.247***	181.296	279.162*
	(166.564)	(139.107)	(139.280)	(125.744)	(126.734)	(146.168)
Observations	34	34	34	34	23	23
R^2	0.404	0.769	0.407	0.731	0.705	0.793
Adjusted R^2	0.272	0.694	0.276	0.645	0.594	0.674
Residual Std. Error	47.712 (df = 27)	30.906 (df = 25)	39.896 (df = 27)	27.937 (df = 25)	27.096 (df = 16)	24.289 (df = 14)
F Statistic	3.052** (df = 6; 27)	10.374*** (df = 8; 25)	3.095** (df = 6; 27)	8.492*** (df = 8; 25)	6.371*** (df = 6; 16)	6.687*** (df = 8; 1
Note:						*p<0.1; **p<0.05; ***p<0.

Table 8: Table:Regression Estimates of Three Modeles

	Dependent variable:		
	math	reading	ps
	(1)	(2)	(3)
log(GDPperc)	−19.929*	−14.930	−6.536
	(10.282)	(9.516)	(17.757)
log(pop)	7.355**	8.735***	6.960
	(3.365)	(3.114)	(5.041)
popd	0.008**	0.007**	0.007*
	(0.003)	(0.003)	(0.004)
rteacher	−3.329**	−3.151**	−2.051
	(1.540)	(1.425)	(1.808)
eyear	−1.067	−1.551	−0.202
	(3.183)	(2.946)	(4.172)
expend	2.570***	2.154***	1.938**
	(0.502)	(0.464)	(0.875)
internet	−0.009	−0.169	−0.137
	(0.221)	(0.205)	(0.276)
mobile	437.763***	417.458***	345.194**
	(105.489)	(97.636)	(151.692)
Observations	50	50	32
R^2	0.635	0.586	0.556
Adjusted R^2	0.574	0.517	0.426
Residual Std. Error	33.031 (df = 42)	30.572 (df = 42)	31.048 (df = 24)
F Statistic	10.433*** (df = 7; 42)	8.494*** (df = 7; 42)	4.287*** (df = 7; 24)
Note:			*p<0.1; **p<0.05; ***p<0.01

Diagnose


Residuals vs Fitted
Residuals
60
40
20
0
−20
−60
52
46
54
400
450
500
550
Fitted values
lm(math ~ log(GDPperc) + log(pop) + popd + rteacher + eyear + expend + inte ...
Normal Q−Q
Standardized residuals
2
1
0
−1
−2
52
54
46
−2
−1
0
1
2
Theoretical Quantiles
lm(math ~ log(GDPperc) + log(pop) + popd + rteacher + eyear + expend + inte ...



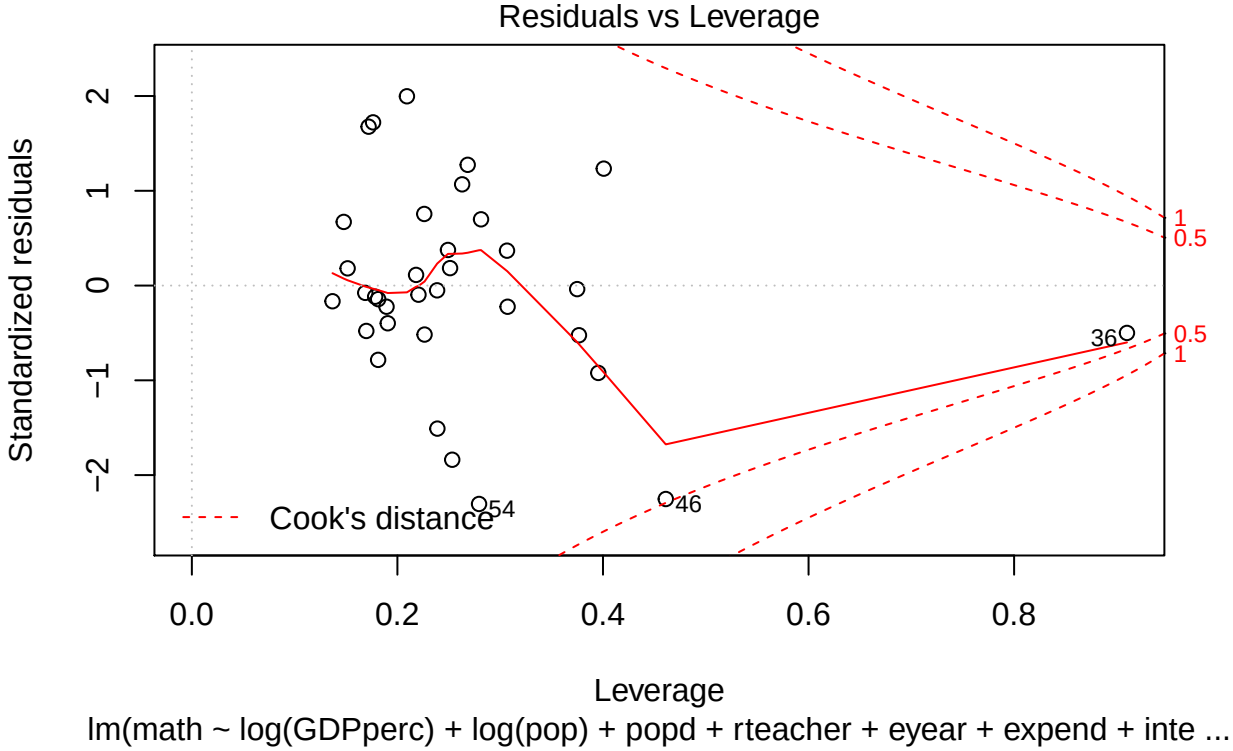


Table 9: Variance Inflation Factors for Math

log(GDPperc)	log(pop)	popd	rteacher	eyear	expend	internet	mobile
4.999	1.729	4.718	1.912	1.113	1.363	4.887	4.028

Table 10: Variance Inflation Factors for Reading

log(GDPperc)	log(pop)	popd	rteacher	eyear	expend	internet	mobile
4.999	1.729	4.718	1.912	1.113	1.363	4.887	4.028

Table 11: Variance Inflation Factors for Problem Soving

log(GDPperc)	log(pop)	popd	rteacher	eyear	expend	internet	mobile
7.523	2.222	8.105	1.938	1.536	2.043	6.617	4.625

Analysis

According to the result, in the secound regressio, these variables are statistically significant. - Math:GDP per capita - Reading: GDP per capita, expenditure - Problem solving: GDP per capita, population, pupills-teacher rate and duraion 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.

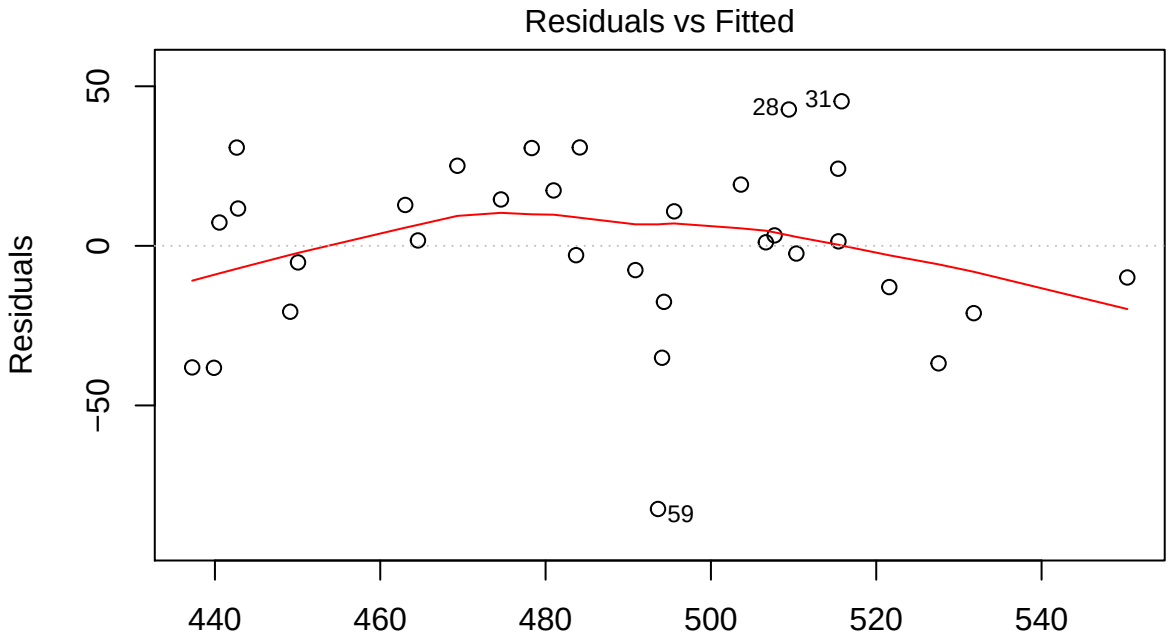
In addition, regarding GDP per capita of math ane reading, it bring the minues effect on the both scores after the variable internet and mobile are added, though it still bring the plus effect on problem solving. Even though it is not statistically significant for math and reading, but it might possible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

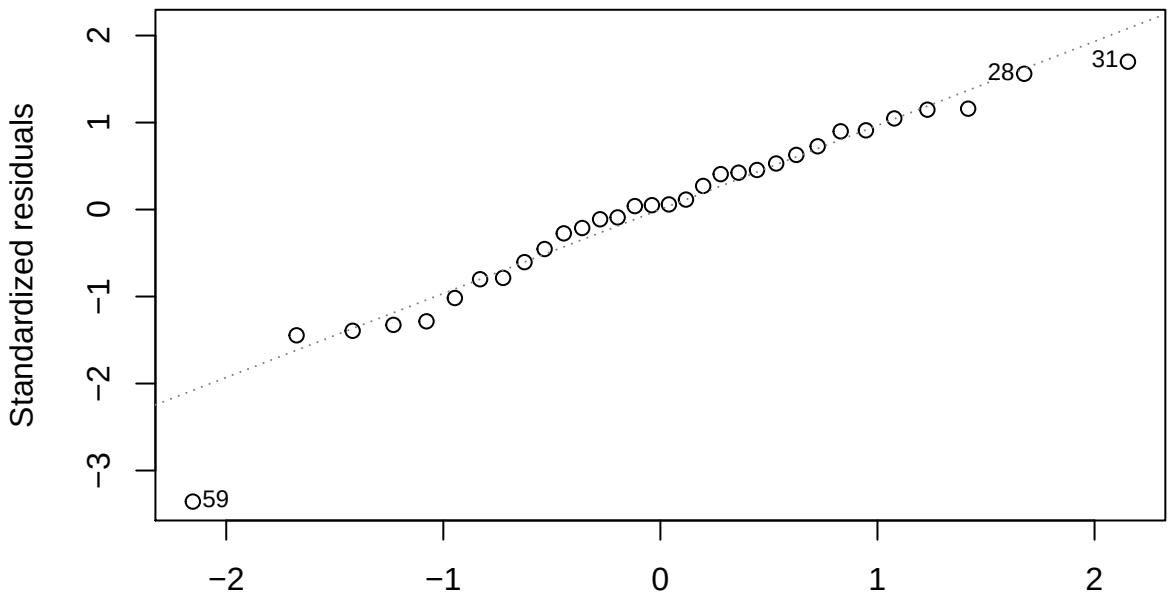
Also, we noticed that about the academic ability (math and reading,) internet accessibility would bring the some impact on the scores. However, it would bring less impact on the problem solving ability.

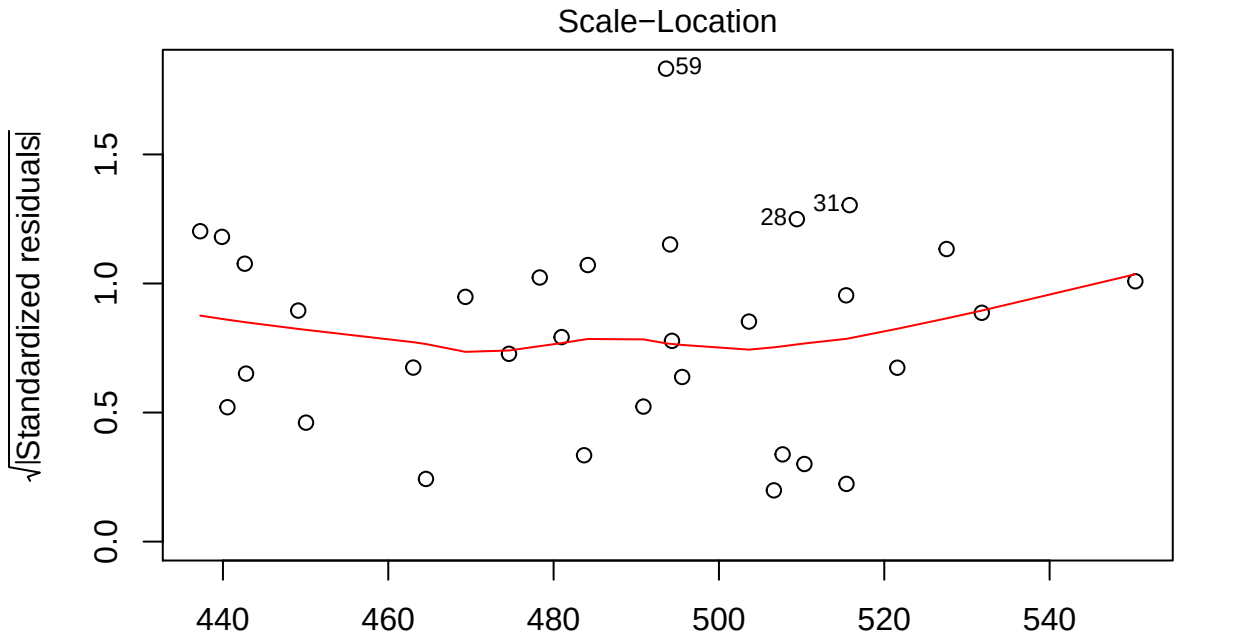
Actually, there is some discussion that internet accessibility might bring reverse effect on student's thinkig ability. Therefore, next we focus on the problem solving score.

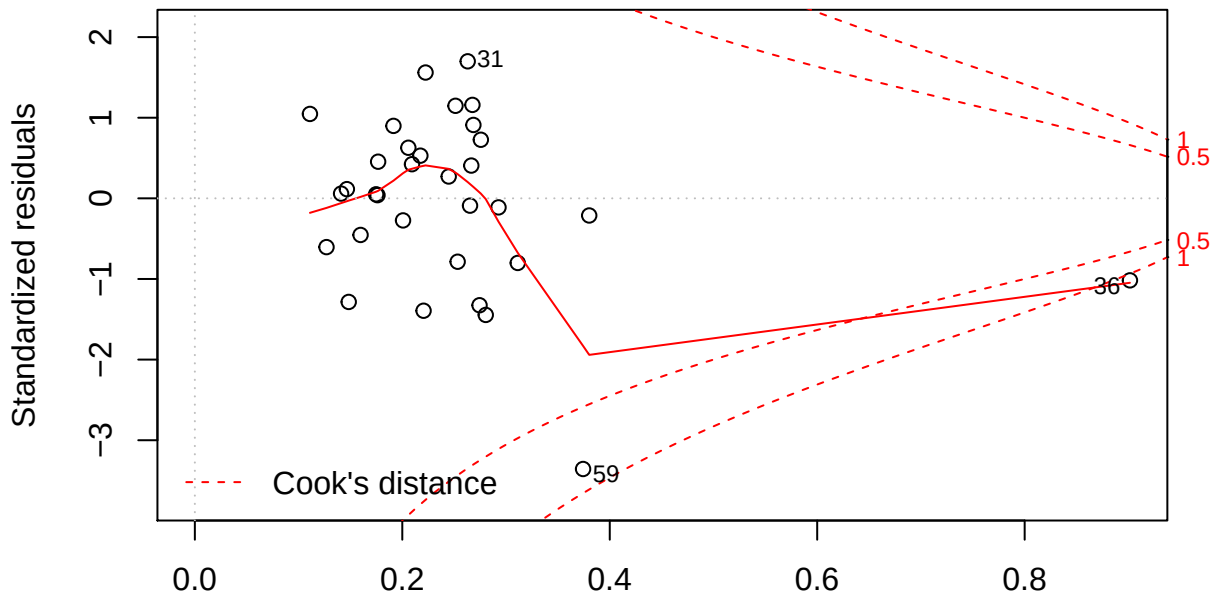
Correlation between Problem Solving and Internet

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


Residuals vs Fitted
Residuals
50
0
−50
28
31
59
440
460
480
500
520
540
Fitted values
lm(ps ~ log(GDPperc) + log(pop) + popd + rteacher + eyear + internet + mobi ...


Normal Q−Q
Standardized residuals
2
1
0
−1
−2
−3
28
31
59
−2
−1
0
1
2
Theoretical Quantiles
lm(ps ~ log(GDPperc) + log(pop) + popd + rteacher + eyear + internet + mobi ...





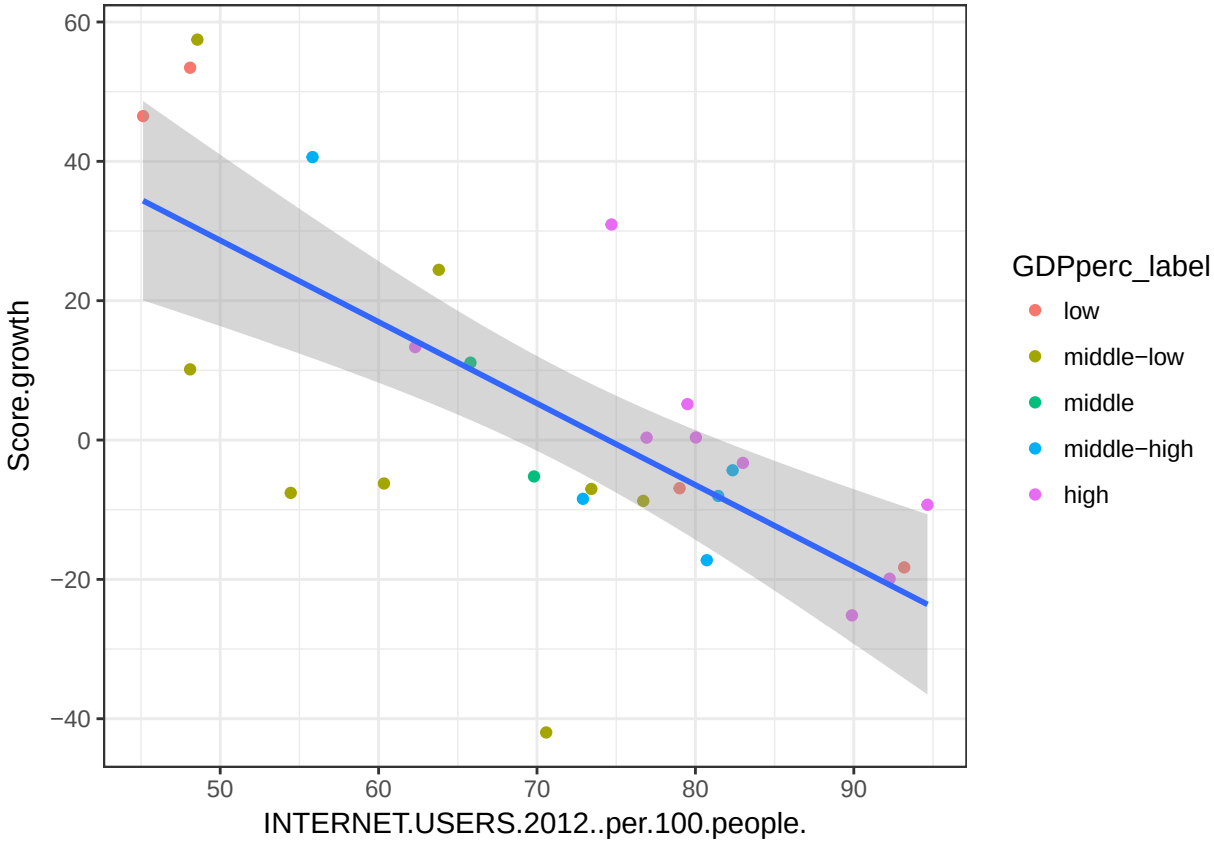
2.5 % 97.5 % (Intercept) 3.211699e+01 658.27002589 log(GDPperc) -4.318557e+01 30.11356920 log(pop) -3.443533e+00 17.36433190 popd -3.808321e-04 0.01522054 rteacher -5.782965e+00 1.68059501 eyear -8.812377e+00 8.40906601 internet 1.316384e-01 3.74374470 mobile -7.061893e-01 0.43261435 Min. 1st Qu. Median Mean 3rd Qu. Max. 1755 11500 22240 32610 47060 149200

Analysis about score growth

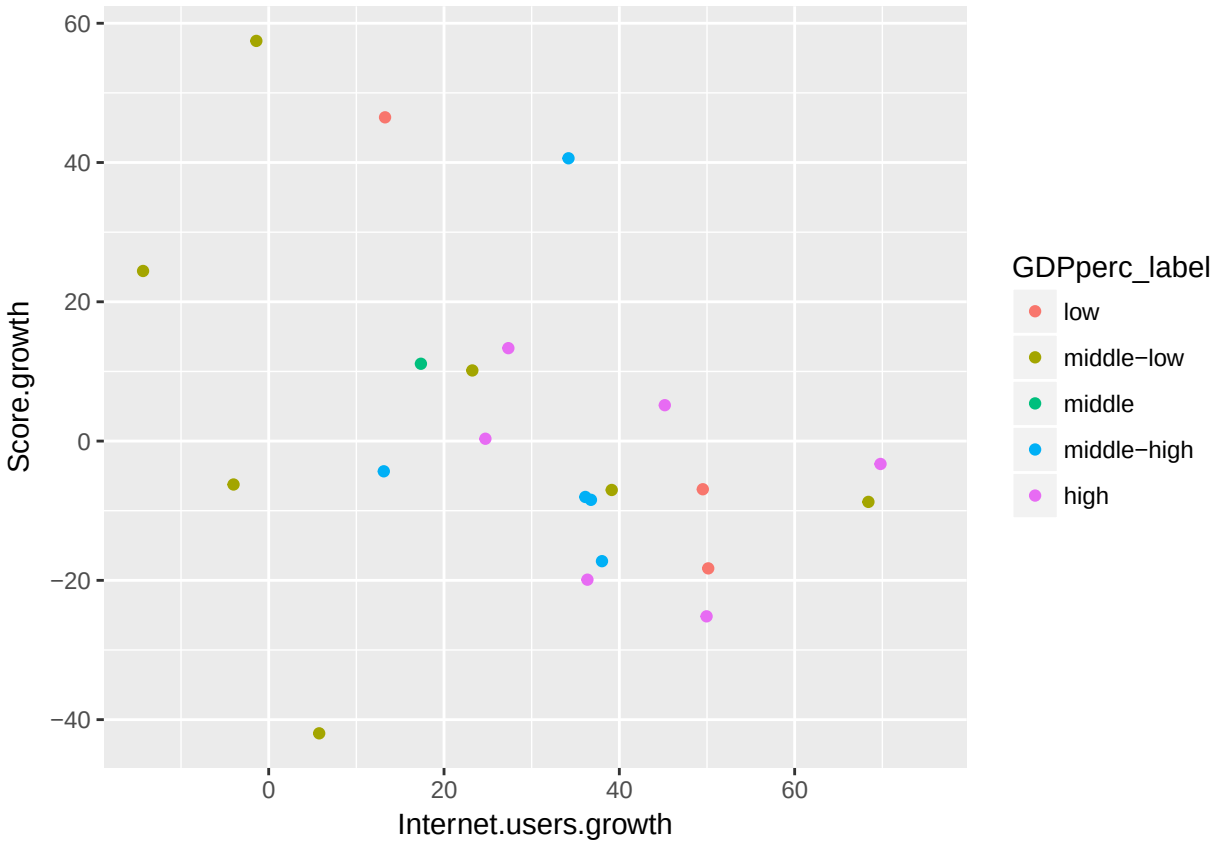
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    3989   13140   24450   29960   44730  101600
```

```
## 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).
```


60
40
20
0
−20
−40
Score.growth
0
20
40
60
Internet.users.growth
GDPperc_label
low
middle−low
middle
middle−high
high