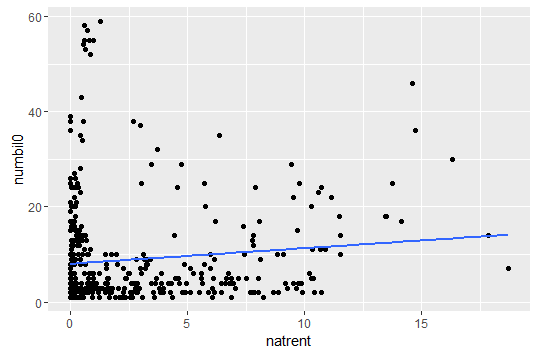
Q1.

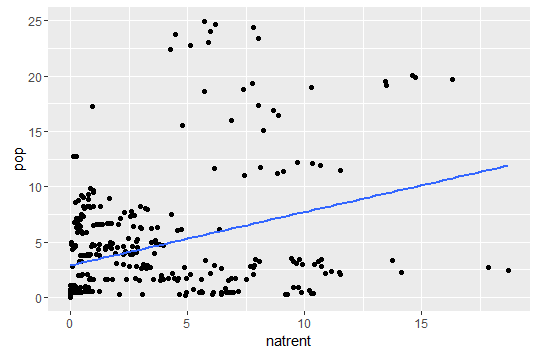
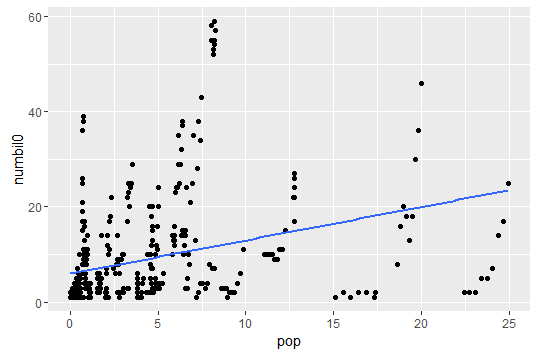
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Numbil0 | natrent | pop | gdppc | roflaw |
| Min | 1 | 0 | 33808 | 0.692 | 0.089 |
| 1st Quantile | 2 |  | 7173727 | 8.133 | 0.442 |
| Median | 4 |  | 20762750 | 24.330 | 0.713 |
| Mean | 8.851 |  | 41133354 | 29.426 | 0.671 |
| 3rd Quantile | 11 |  | 59140875 | 44.339 | 0.923 |
| Max | 59 |  | 249865520 | 193.892 | 1.000 |

What does a typical country look like?

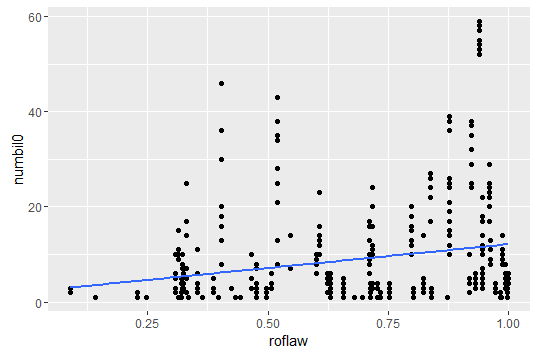
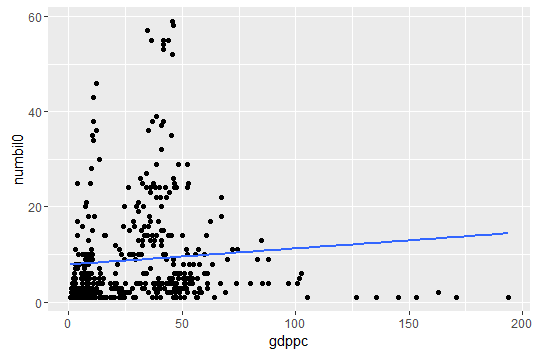
The variable “natrent” and “pop” need to rescale because the their number is too big, if we keep the original number their coefficient would be very smaller which is not good for linear regression.

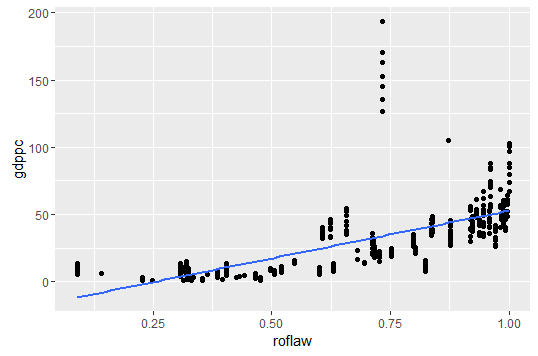
Q2.





Q3.





Since the ‘roflaw’ has relationships with both ‘gdppc’ and ‘numbil0’ which is the dependent and independent variable of the simple linear regression. We can assume that the ‘roflaw’ is a omitted variable in the simple linear regression. As the relationship between ‘roflaw’ and ‘numbil0’ is positive linear, then the coefficient or slope of ‘gdppc’ should be smaller. (解释一下怎么影响的)

Q4.

The table is shown below.

As more and more variable added in the regression, the R and adjusted R was getting larger, which means our regression was getting more suitable for estimating the number of billionaires. Every time the new variable added in the regression, the old variables’ coefficients will have a big change. This shows that the new variable is useful for the regression in the other way.

When the ‘pop’ and ‘roflaw’ added in the regression, the coefficient of ‘natrent’ will have a large decrease and increase. This also means those two variables are the omitted variable for the regression which ‘natrent’ is the dependent variable and ‘numbil0’ is the independent variable.

In the regression, every 0.1 unit of ‘roflaw’ increasement will make the ‘numbilo’ increase about 2 units holding the other variables fixed. .

Like we talked about in Q3, the ‘roflaw’ have omitted variable bias in the regression impacted with the coefficient of ‘gdppc’.

We can see that the dummy variable d2013 has extremely high coefficient and small p-value compared with other dummy variables. This could have two explanation:

1. Most of the countries are always developing. The countries which were observed before 2013 might have more billionaires in 2013. It’s not fair for the countries which observed before 2013.
2. 2013 has more countries than any other years.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | | | | | |
|  | *Dependent variable:* | | | | |
|  |  | | | | |
|  | *numbil0* | | | | |
|  | *(1)* | *(2)* | *(3)* | *(4)* | *(5)* |
|  | | | | | |
| *natrent* | *0.326* | *-0.014* | *-0.046* | *0.240* | *0.216* |
|  | *(0.146)* | *(0.147)* | *(0.140)* | *(0.110)* | *(0.108)* |
|  |  |  |  |  |  |
| *pop* |  | *0.703* | *0.905* | *1.064* | *1.077* |
|  |  | *(0.133)* | *(0.149)* | *(0.143)* | *(0.140)* |
|  |  |  |  |  |  |
| *gdppc* |  |  | *0.096* | *-0.028* | *-0.032* |
|  |  |  | *(0.024)* | *(0.010)* | *(0.010)* |
|  |  |  |  |  |  |
| *roflaw* |  |  |  | *21.966* | *22.691* |
|  |  |  |  | *(2.270)* | *(2.291)* |
|  |  |  |  |  |  |
| *d2006* |  |  |  |  | *0.862* |
|  |  |  |  |  | *(1.746)* |
|  |  |  |  |  |  |
| *d2007* |  |  |  |  | *1.932* |
|  |  |  |  |  | *(1.721)* |
|  |  |  |  |  |  |
| *d2008* |  |  |  |  | *2.586* |
|  |  |  |  |  | *(1.834)* |
|  |  |  |  |  |  |
| *d2009* |  |  |  |  | *0.443* |
|  |  |  |  |  | *(1.658)* |
|  |  |  |  |  |  |
| *d2010* |  |  |  |  | *1.420* |
|  |  |  |  |  | *(1.722)* |
|  |  |  |  |  |  |
| *d2011* |  |  |  |  | *3.387* |
|  |  |  |  |  | *(1.834)* |
|  |  |  |  |  |  |
| *d2012* |  |  |  |  | *3.873* |
|  |  |  |  |  | *(1.802)* |
|  |  |  |  |  |  |
| *d2013* |  |  |  |  | *6.101* |
|  |  |  |  |  | *(1.852)* |
|  |  |  |  |  |  |
| *Constant* | *8.030* | *5.992* | *2.406* | *-10.038* | *-12.814* |
|  | *(0.653)* | *(0.550)* | *(0.863)* | *(1.547)* | *(1.933)* |
|  |  |  |  |  |  |
|  | | | | | |
| *Observations* | *444* | *444* | *444* | *444* | *444* |
| *R2* | *0.012* | *0.107* | *0.157* | *0.289* | *0.318* |
| *Adjusted R2* | *0.009* | *0.103* | *0.152* | *0.282* | *0.299* |
| *Residual Std. Error* | *10.709 (df = 442)* | *10.189 (df = 441)* | *9.910 (df = 440)* | *9.116 (df = 439)* | *9.006 (df = 431)* |
| *F Statistic* | *5.004\*\* (df = 1; 442)* | *16.478\*\*\* (df = 2; 441)* | *15.035\*\*\* (df = 3; 440)* | *31.309\*\*\* (df = 4; 439)* | *12.383\*\*\* (df = 12; 431)* |
|  | | | | | |
| *Note:* | *\*p<0.1; \*\*p<0.05; \*\*\*p<0.01* | | | | |

Q5

In first regression, as one unit of log(pop) increases, the number of billionaires would increase by 3.866, holding the other variables fixed.

In second regression, as one unit of pop increases, the number of log(billionaires) would increase by 0.099, holding the other variables fixed.

In third regression, as 1% of log(pop) increases, the number of log(billionaires) would increase by 0.442%, holding the other variables fixed.

In the all three regressions, for testing the pop or log(pop) equal zeros, the p-values are all below 5% significant level, therefore the pop or log(pop) impacts the numbil0 in 5% significant level.

|  |  |  |  |
| --- | --- | --- | --- |
|  | | | |
|  | *Dependent variable:* | | |
|  |  | | |
|  | numbil0 | log(numbil0) | |
|  | (1) | (2) | (3) |
|  | | | |
| log(pop) | 3.866\*\*\* |  | 0.442\*\*\* |
|  | (0.442) |  | (0.029) |
|  |  |  |  |
| pop |  | 0.099\*\*\* |  |
|  |  | (0.012) |  |
|  |  |  |  |
| natrent | 0.157 | 0.053\*\*\* | 0.035\*\*\* |
|  | (0.128) | (0.009) | (0.009) |
|  |  |  |  |
| gdppc | 0.084\*\*\* | -0.005\*\*\* | 0.008\*\*\* |
|  | (0.021) | (0.002) | (0.002) |
|  |  |  |  |
| roflaw | 13.781\*\*\* | 2.459\*\*\* | 1.622\*\*\* |
|  | (1.865) | (0.201) | (0.178) |
|  |  |  |  |
| Constant | -5.387\*\*\* | -0.423\*\*\* | -0.057 |
|  | (1.303) | (0.140) | (0.127) |
|  |  |  |  |
|  | | | |
| Observations | 444 | 444 | 444 |
| R2 | 0.320 | 0.323 | 0.450 |
| Adjusted R2 | 0.314 | 0.316 | 0.445 |
| Residual Std. Error (df = 439) | 8.914 | 0.877 | 0.790 |
| F Statistic (df = 4; 439) | 51.596\*\*\* | 52.280\*\*\* | 89.662\*\*\* |
|  | | | |
| *Note:* | \*p<0.1; \*\*p<0.05; \*\*\*p<0.01 | | |

Q6

R output：

Call:

lm(formula = log(numbil0) ~ log(pop) \* d2006 + log(pop) \* d2007 +

log(pop) \* d2008 + log(pop) \* d2009 + log(pop) \* d2010 +

log(pop) \* d2011 + log(pop) \* d2012 + log(pop) \* d2013 +

natrent + gdppc, data = bil\_data)

Residuals:

Min 1Q Median 3Q Max

-2.16038 -0.54226 0.04721 0.54029 2.22842

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 0.587402 0.146624 4.006 7.29e-05 \*\*\*

log(pop) 0.466718 0.071441 6.533 1.85e-10 \*\*\*

d2006 0.040324 0.186143 0.217 0.82860

d2007 0.078915 0.180811 0.436 0.66273

d2008 0.185641 0.179841 1.032 0.30254

d2009 -0.016032 0.182998 -0.088 0.93023

d2010 0.081989 0.182334 0.450 0.65318

d2011 0.244905 0.183061 1.338 0.18167

d2012 0.363769 0.180842 2.012 0.04490 \*

d2013 0.479885 0.177836 2.698 0.00724 \*\*

natrent 0.002921 0.011846 0.247 0.80534

gdppc 0.019879 0.001871 10.622 < 2e-16 \*\*\*

log(pop):d2006 0.032044 0.102866 0.312 0.75557

log(pop):d2007 0.072587 0.100424 0.723 0.47020

log(pop):d2008 0.025129 0.096866 0.259 0.79544

log(pop):d2009 0.002172 0.100527 0.022 0.98277

log(pop):d2010 -0.015778 0.099336 -0.159 0.87388

log(pop):d2011 0.044585 0.100489 0.444 0.65750

log(pop):d2012 -0.036743 0.101898 -0.361 0.71859

log(pop):d2013 -0.027485 0.099944 -0.275 0.78345

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.84 on 424 degrees of freedom

Multiple R-squared: 0.3994, Adjusted R-squared: 0.3725

F-statistic: 14.84 on 19 and 424 DF, p-value: < 2.2e-16

From the basic year which is 2005, every 1 unit of log(pop) increase 0.467 unit of log(numnbil0). And for the other years from 2006 to 2013, 1 unit of log(pop) increase will make log(numbil0) increase by 0.467+0.032, 0.467+0.073, 0.467+0.025, 0.467+0.002, 0.467-0.015, 0.467+0.046, 0.467-0.036, 0.467-0.027 unit increase separately holding the other variables fixed. The statistical significance of log(pop) and d2013 are quite good, which are lower than 5%. But the other dummy year variables’ significance are bad, they are all higher than 10%.

1% extra pop is associate with and change in bilnum0 by 0.439% in 2013 holding the other variable fixed. The significance of interaction between log(pop) and d2013 is high which is within 5% significance level.

Because the sum of the coefficients of year variables will equal 1, so the year variables will be dummy variables. We avoid the dummy variable trap by use the d2005 as the base case or omitted category in the regression. (THE OTHER WAY) ~~Otherwise, we add +0 in the end of the regression code, like:~~

~~reg9 =lm(log(numbil0)~log(pop)\*d2006+log(pop)\*d2007+log(pop)\*d2008+log(pop)\*d2009~~

~~+log(pop)\*d2010+log(pop)\*d2011+log(pop)\*d2012+log(pop)\*d2013+~~

~~natrent+gdppc,data = bil\_data)~~

Q7.

Since the F value is 0.2214 and p-value is 0.987, therefore we have strong evidence to show that we cannot reject the null hypothesis, that means there is no interaction between the dummy year variables and log(pop) variable.

Q8.

Output:

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -1.02172 0.30536 -3.346 0.000891 \*\*\*

log(gdppc) 0.68241 0.11682 5.841 1.01e-08 \*\*\*

roflaw 1.88462 0.65210 2.890 0.004043 \*\*

log(pop) 0.45891 0.02627 17.472 < 2e-16 \*\*\*

natrent 0.02337 0.01155 2.023 0.043716 \*

log(gdppc):roflaw -0.42168 0.18639 -2.262 0.024163 \*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

From the output, log(gdppc), roflaw and log(gdppc):roflaw’s p-value is lower than 0.05, we cannot reject the null hypothesis in 5% significant level. So that the coefficient of log(gdppc), roflaw and log(gdppc):roflaw is not 0. (interception the sign of 3 variables)

Partial effects of GDP as a function of roflaw:

1 unit of log(gdppc) is associated with an increase in log(numbil0) of: 0.682-0.422\*roflaw of number of bilnum0 per unit of log(gdppc).

Q9

|  |  |  |  |
| --- | --- | --- | --- |
| Roflaw | Elasticity | F-statistic | p-value |
| Roflaw = 0.1 |  |  |  |
| Roflaw = 0.2 |  |  |  |
| Roflaw = 0.3 |  |  |  |
| Roflaw = 0.4 |  |  |  |
| Roflaw = 0.5 |  |  |  |
| Roflaw = 0.6 |  |  |  |
| Roflaw = 0.7 |  |  |  |
| Roflaw = 0.8 |  |  |  |
| Roflaw = 0.9 |  |  |  |
| Roflaw = 1.0 |  |  |  |