

Name: 林成鏞 Julius Ling Sing Kai

Department: Psychology I12

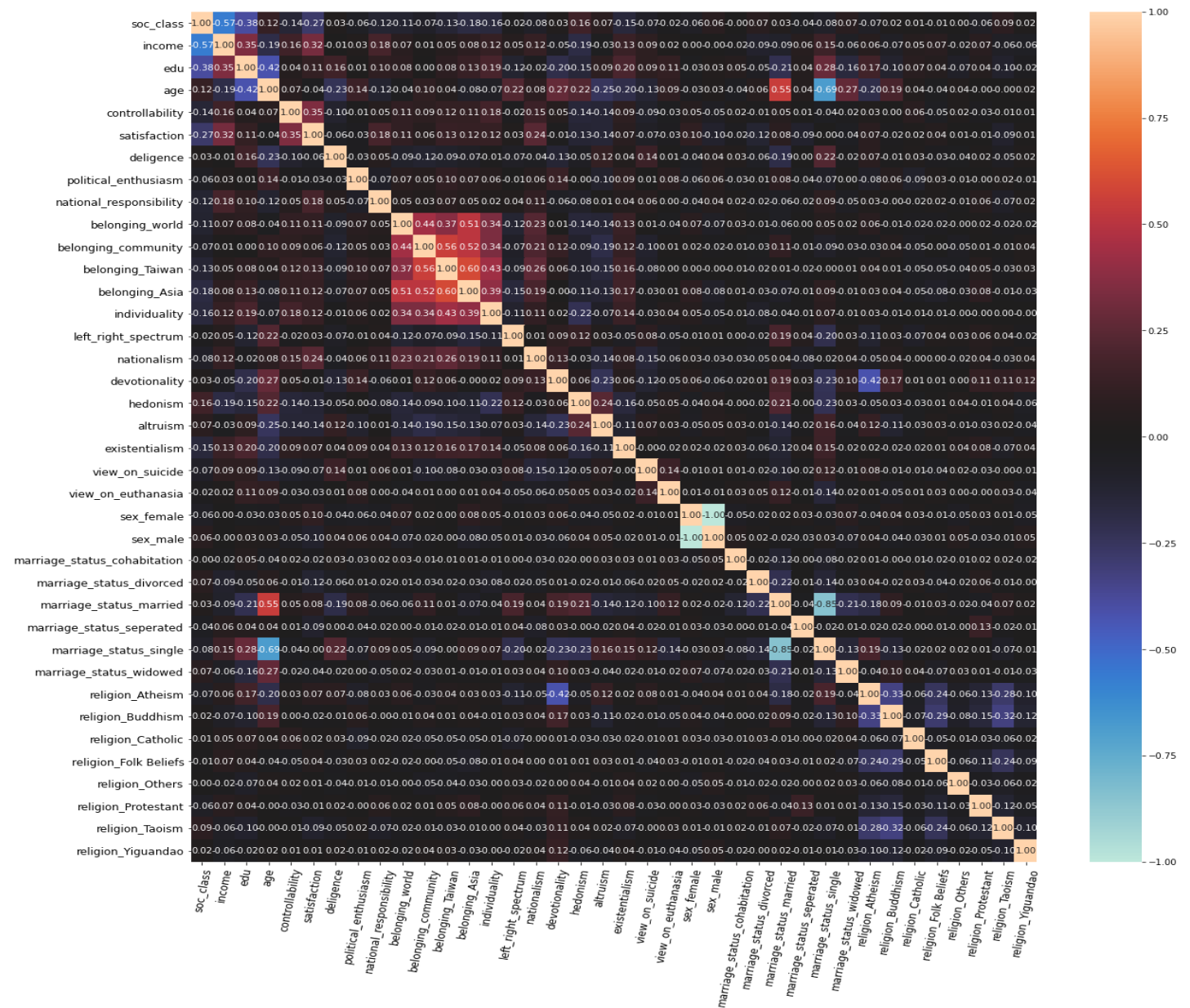
Student ID: D84085360

## Homework 3 - Moderation & Mediation Analysis

Dataset: 2010 World Values Survey, Taiwan (2010 世界價值觀調查-台灣)

Source: 中研院 SRDA [https://srda.sinica.edu.tw/datasearch\\_detail.php?id=I220](https://srda.sinica.edu.tw/datasearch_detail.php?id=I220)

The dataset contains 1238 observations and a total of 308 columns with the respective response on the questionnaire. The variables of interest are being subsetting and to be plotted a correlation heatmap to view the inner structure of the data. A total of **961 samples** are left after handling with the missing data.



It's notable that *Satisfaction* correlates with *Income*, *Social\_class*, *Controllability* in an unignorable effect (-0.27, 0.32, 0.35 respectively). Hence there's a question on the Mediation process of one's overall satisfaction.

Variables of Interest:

*Satisfaction*: min = 1, max = 10. The higher the greater satisfaction (生活滿意度)

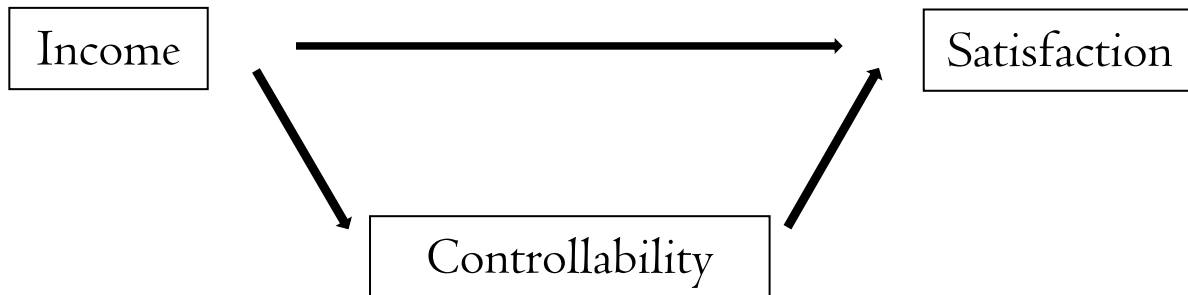
*Income*: min = 1, max = 10, representing the ratio of household income (eg. 1 = the bottom 10%, 10 = the top 10% etc.). The higher the greater household income (家庭收入)

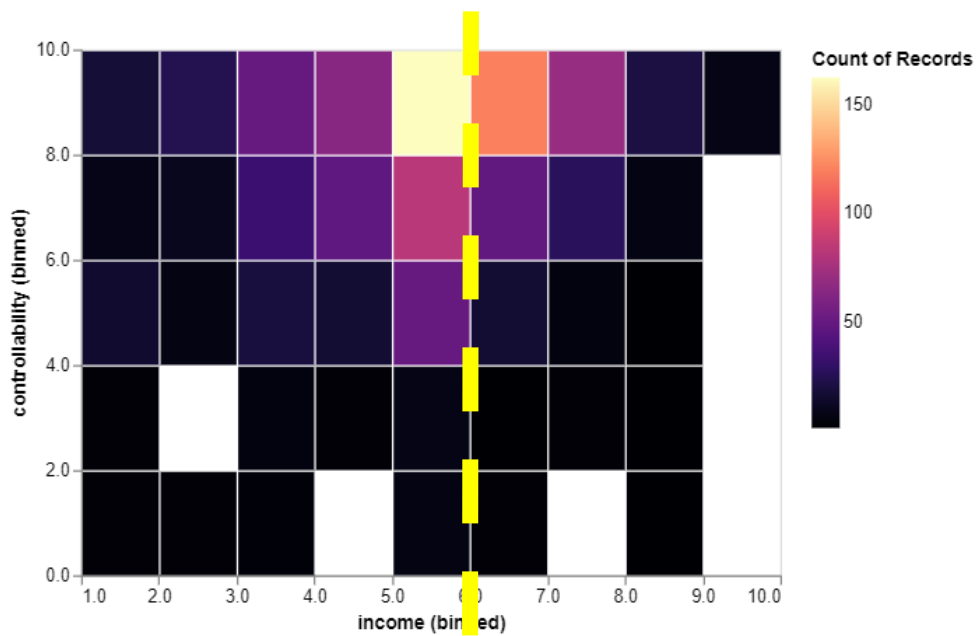
*Controllability*: min = 1, max = 10. The higher the greater controllability (生活可控感)

The analysis suggests that:

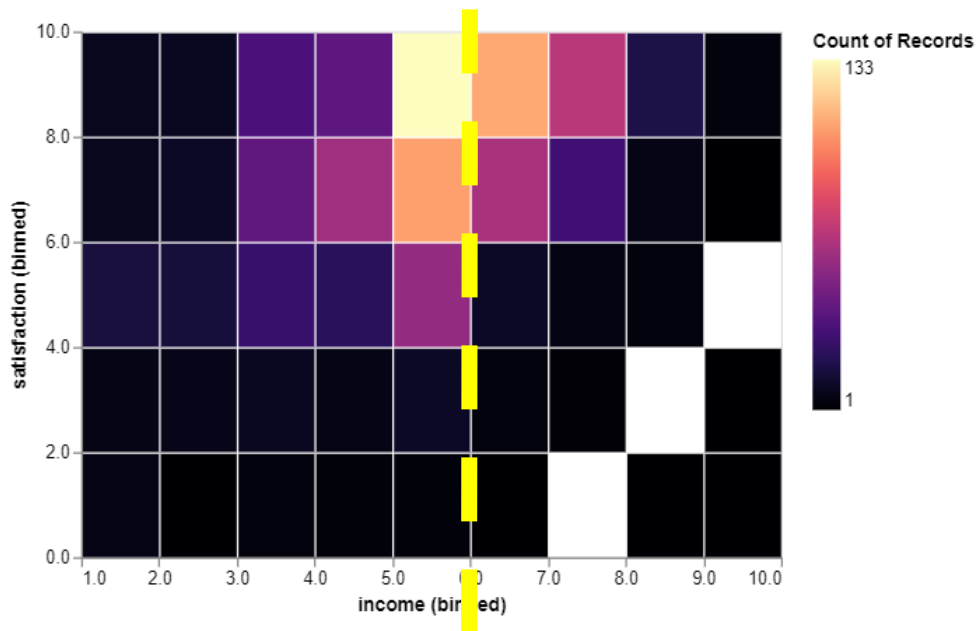
The **Direct Effect** of Income on the Satisfaction = 0.39

The **Indirect Effect** of Income on the Satisfaction, through the path of Controllability = 0.06 and statistical significance are shown.

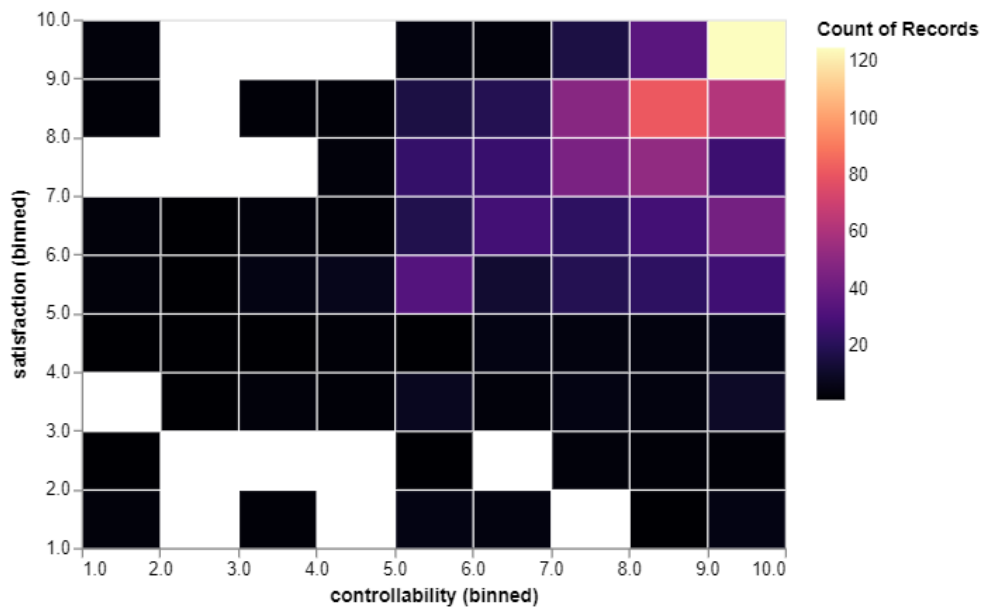




Count Heatmap:  
Income Vs. Controllability



Income Vs. Satisfaction



Controllability Vs. Satisfaction

(Sample size = 961)

```
Call:
lm(formula = satisfaction ~ income, data = dta)

Residuals:
    Min       1Q   Median       3Q      Max
-7.5283 -1.1381  0.2029  1.0324  4.5931

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.01670    0.19039   26.35  <2e-16 ***
income       0.39018    0.03711   10.52  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.905 on 959 degrees of freedom
Multiple R-squared:  0.1034, Adjusted R-squared:  0.1024
F-statistic: 110.6 on 1 and 959 DF, p-value: < 2.2e-16
```

```
Call:
lm(formula = controllability ~ income, data = dta)

Residuals:
    Min       1Q   Median       3Q      Max
-7.1275 -1.1667  0.2568  1.4490  3.2177

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.59012    0.19901   33.115  < 2e-16 ***
income       0.19218    0.03879    4.955 8.56e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.991 on 959 degrees of freedom
Multiple R-squared:  0.02496, Adjusted R-squared:  0.02394
F-statistic: 24.55 on 1 and 959 DF, p-value: 8.56e-07
```

```
Call:
lm(formula = satisfaction ~ income + controllability, data = dta)

Residuals:
    Min       1Q   Median       3Q      Max
-8.0451 -1.0754  0.2254  1.2019  5.0474

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.98966    0.26408   11.321  <2e-16 ***
income       0.33106    0.03560    9.299  <2e-16 ***
controllability 0.30759    0.02927   10.509  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.804 on 958 degrees of freedom
Multiple R-squared:  0.1961, Adjusted R-squared:  0.1944
F-statistic: 116.8 on 2 and 958 DF, p-value: < 2.2e-16
```

Sobel's Test:

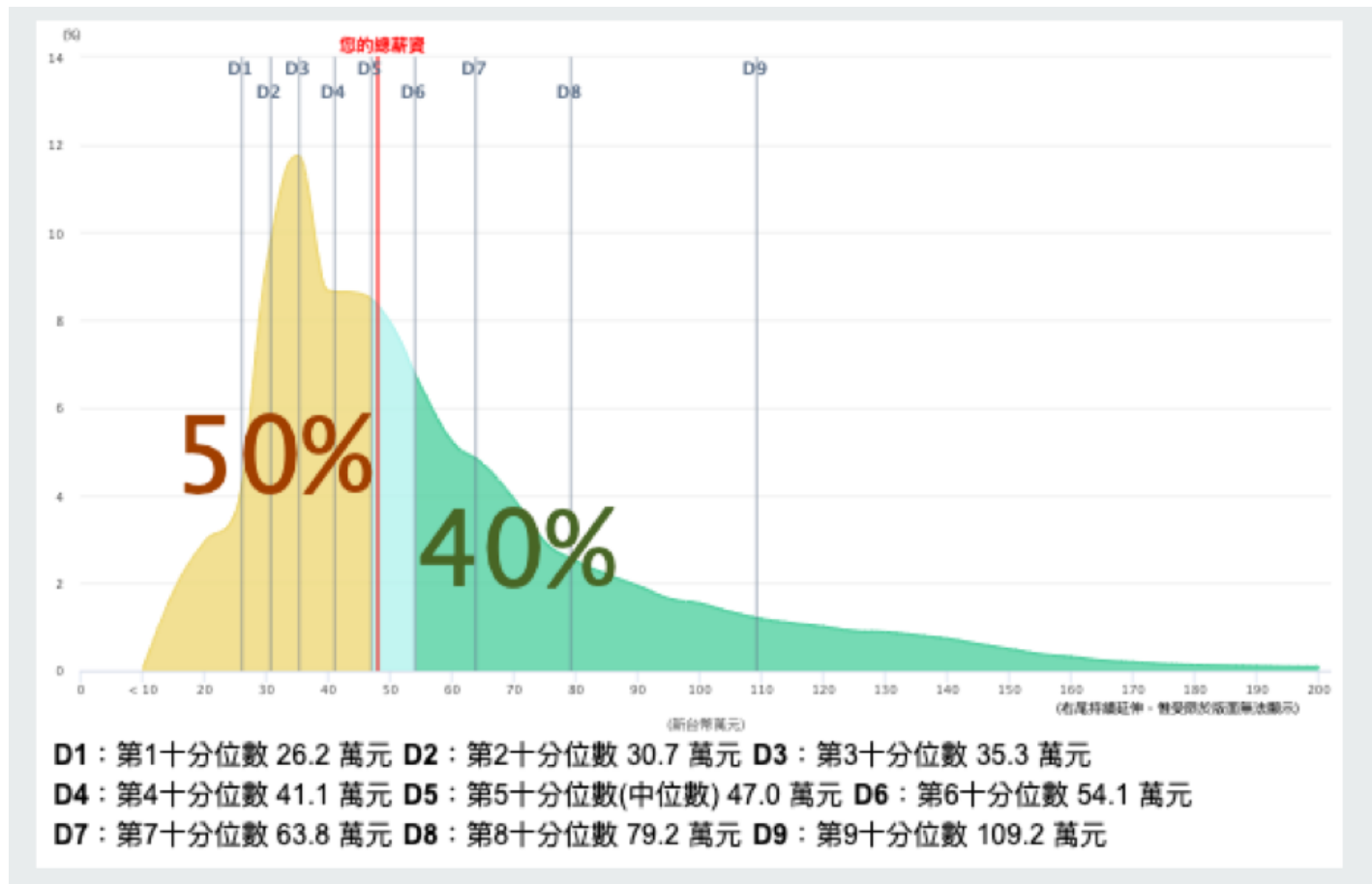
	Est	z_ab.Est	pz_ab.Est	
	5.911164e-02	4.481600e+00	7.408538e-06	(Statistical significance)

Note that **the coefficient of Income does decrease** and the P-value of the Sobel's Indirect Effect Test is below 0.05 (statistically significant). Thus, we are prone to believe that **the moderation effect of Controllability holds partially**:

In general the higher the Income, the higher the one's overall Satisfaction. Apart from that, the higher Income does increase one's sense of Controllability thus further driving up their Satisfaction.

However,

the distribution of household income of Taiwan is usually positively skewed (shown as below).



Hence, I'm subsetting the data regarding the **Income** to the Relatively-Low-Income (Income  $\leq 5$ ) group and the Relatively-High-Income (Income  $> 5$ ) group. Then I'm to perform the moderation analysis again.

Here's just a remark on my previous try validating my hypothesis by not conducting a spline regression. I think it's worth retaining as a remind of such "misbehaving". Broken, but that's where the light gets in.

The Relatively-High-Income group (Sample size = 327):

```
Call:
lm(formula = satisfaction ~ income, data = dta_h)

Residuals:
    Min       1Q   Median       3Q      Max
-6.6037 -0.6037  0.3963  1.3963  2.4263

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  7.648645   0.776875   9.845  <2e-16 ***
income      -0.007499   0.117445  -0.064   0.949
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.653 on 325 degrees of freedom
Multiple R-squared:  1.254e-05,    Adjusted R-squared:  -0.003064
F-statistic: 0.004077 on 1 and 325 DF,  p-value: 0.9491
```

```
Call:
lm(formula = controllability ~ income, data = dta_h)

Residuals:
    Min       1Q   Median       3Q      Max
-7.2058 -0.7857  0.2143  1.2143  2.2143

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.52541    0.77894   8.267 3.5e-15 ***
income       0.21101    0.12961   1.760  0.07931 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.68 on 325 degrees of freedom
Multiple R-squared:  0.00943, Adjusted R-squared:  0.006396
F-statistic: 0.00943 on 1 and 325 DF,  p-value: 0.07931
```

Here's just a remark on my previous try validating my hypothesis by not conducting a spline regression. I think it's worth retaining as a remind of such "misbehaving". Broken, but that's where the light gets in.

```
Call:
lm(formula = satisfaction ~ income + controllability, data = dta_h)

Residuals:
    Min       1Q   Median       3Q      Max
-6.9764 -0.6571  0.3429  0.9936  4.2094

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.02040   0.82801   7.271 2.70e-12 ***
income      -0.05991   0.11432  -0.524   0.601
controllability 0.24952   0.05289   4.718 3.55e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.602 on 324 degrees of freedom
Multiple R-squared:  0.06429, Adjusted R-squared:  0.05852
F-statistic: 11.13 on 2 and 324 DF,  p-value: 2.112e-05
```

Sobel's Test:

Est	z_ab.Est	pz_ab.Est	
0.05241364	1.64917161	0.09911248	(Statistical Insignificance)

The Relatively-Low-Income group (Sample size = 634):

```
Call:
lm(formula = satisfaction ~ income, data = dta_1)

Residuals:
    Min       1Q   Median       3Q      Max
-6.0223 -1.1156 -0.0223  0.9777  4.7912

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.75542    0.26705   17.807 < 2e-16 ***
income       0.45338    0.06414    7.069 4.15e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.01 on 632 degrees of freedom
Multiple R-squared:  0.07327, Adjusted R-squared:  0.0718
F-statistic: 49.97 on 1 and 632 DF, p-value: 4.149e-12
```

```
Call:
lm(formula = controllability ~ income, data = dta_1)

Residuals:
    Min       1Q   Median       3Q      Max
-6.4734 -1.4376  0.5266  1.6697  3.0988

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.75814    0.21352   31.837 < 2e-16 ***
income       0.14505    0.06806    2.119 0.03606 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.134 on 632 degrees of freedom
Multiple R-squared:  0.00634, Adjusted R-squared:  0.005363
F-statistic: 4.425 on 1 and 632 DF, p-value: 0.03606
```

Here's just a remark on my previous try validating my hypothesis by not conducting a spline regression. I think it's worth retaining as a remind of such "misbehaving". Broken, but that's where the light gets in.

```
Call:
lm(formula = satisfaction ~ income + controllability, data = dta_1)

Residuals:
    Min       1Q   Median       3Q      Max
-6.4442 -1.1328  0.1493  1.2115  5.1001

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.53967    0.34527    7.356 5.93e-13 ***
income       0.40648    0.06038    6.732 3.78e-11 ***
controllability 0.32786    0.03515    9.327 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.886 on 631 degrees of freedom
Multiple R-squared:  0.1856, Adjusted R-squared:  0.183
F-statistic: 71.88 on 2 and 631 DF, p-value: < 2.2e-16
```

Sobel's Test:

Est	z_ab.Est	pz_ab.Est	
0.04690019	2.04939420	0.04042358	(Statistical significance)

The effect of higher *Income* on one's overall *Satisfaction* seems hold in the Relatively-Low-Income group while it shows no significant effect on the Relatively-High-Income group (it even shows a negative correlation!)

On the other hand, the sense of *Controllability* seems to affect one's overall *Satisfaction* in both group while the moderation effect of *Controllability* in the Relatively-High-Income group is not significant.

I'm believing that *Income* does affect one's overall *Satisfaction* by offering the sense of *Controllability* in the Relatively-Low-Income group but its limitation meets as the *Income* drives up, and it no longer offers the sense of *Controllability* to the Relatively-High-Income group.

I'm conducting a **Spline Regression** to tell one different story each for the two groups, where an apparent disparity, in terms of their income, lies.

The best way to differentiate the group based on income is at *Income* = 6 (justified with the highest-coefficient-of-determination model selection approach). The results is shown below.

(Cutoff point: *Income* = 6)

```
lm(formula = satisfaction ~ income + cp_income, data = dta)
```

Residuals:

Min	1Q	Median	3Q	Max
-7.1069	-1.1084	0.3536	1.1427	4.8409

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	4.67177	0.24988	18.696	< 2e-16 ***
income	0.48733	0.05882	8.284	3.97e-16 ***
cp_income	-0.23770	0.11181	-2.126	0.0338 *

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

Residual standard error: 1.901 on 958 degrees of freedom  
Multiple R-squared: 0.1076, Adjusted R-squared: 0.1057  
F-statistic: 57.74 on 2 and 958 DF, p-value: < 2.2e-16

```
lm(formula = controllability ~ income + cp_income, data = dta)
```

Residuals:

Min	1Q	Median	3Q	Max
-7.2525	-1.1904	0.2569	1.5117	3.1074

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	6.74361	0.26170	25.769	<2e-16 ***
income	0.14894	0.06161	2.418	0.0158 *
cp_income	0.10577	0.11709	0.903	0.3666

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

Residual standard error: 1.991 on 958 degrees of freedom  
Multiple R-squared: 0.02579, Adjusted R-squared: 0.02376  
F-statistic: 12.68 on 2 and 958 DF, p-value: 3.67e-06



```
lm(formula = satisfaction ~ income + controllability + cp_income +
    cp_controllability, data = dta)
```

Residuals:

Min	1Q	Median	3Q	Max
-7.3124	-0.9415	0.2130	1.1072	4.8768

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	2.71144	0.31789	8.530	< 2e-16 ***
income	0.42274	0.05699	7.418	2.62e-13 ***
controllability	0.29807	0.03008	9.909	< 2e-16 ***
cp_income	-0.41171	0.13799	-2.984	0.00292 **
cp_controllability	0.04624	0.02901	1.594	0.11125

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.798 on 956 degrees of freedom  
Multiple R-squared: 0.2036, Adjusted R-squared: 0.2003  
F-statistic: 61.11 on 4 and 956 DF, p-value: < 2.2e-16

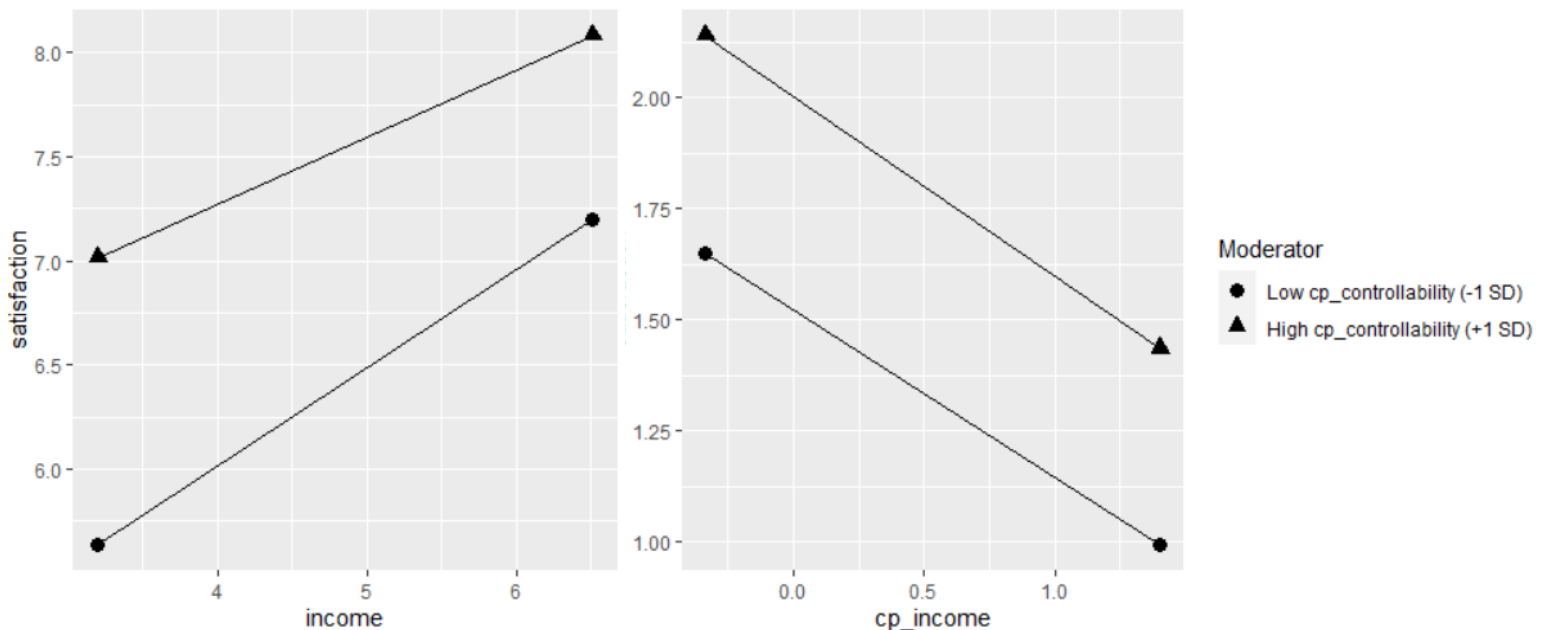
Sobel Test:

(*Income* < 6):

Est	z_ab.Est	pz_ab.Est	(Statistical Significance)
0.04439647	2.34880278	0.01883388	

(*Income* >= 6):

Est	z_ab.Est	pz_ab.Est	(Statistical Insignificance)
0.004891032	0.785910625	0.431919849	

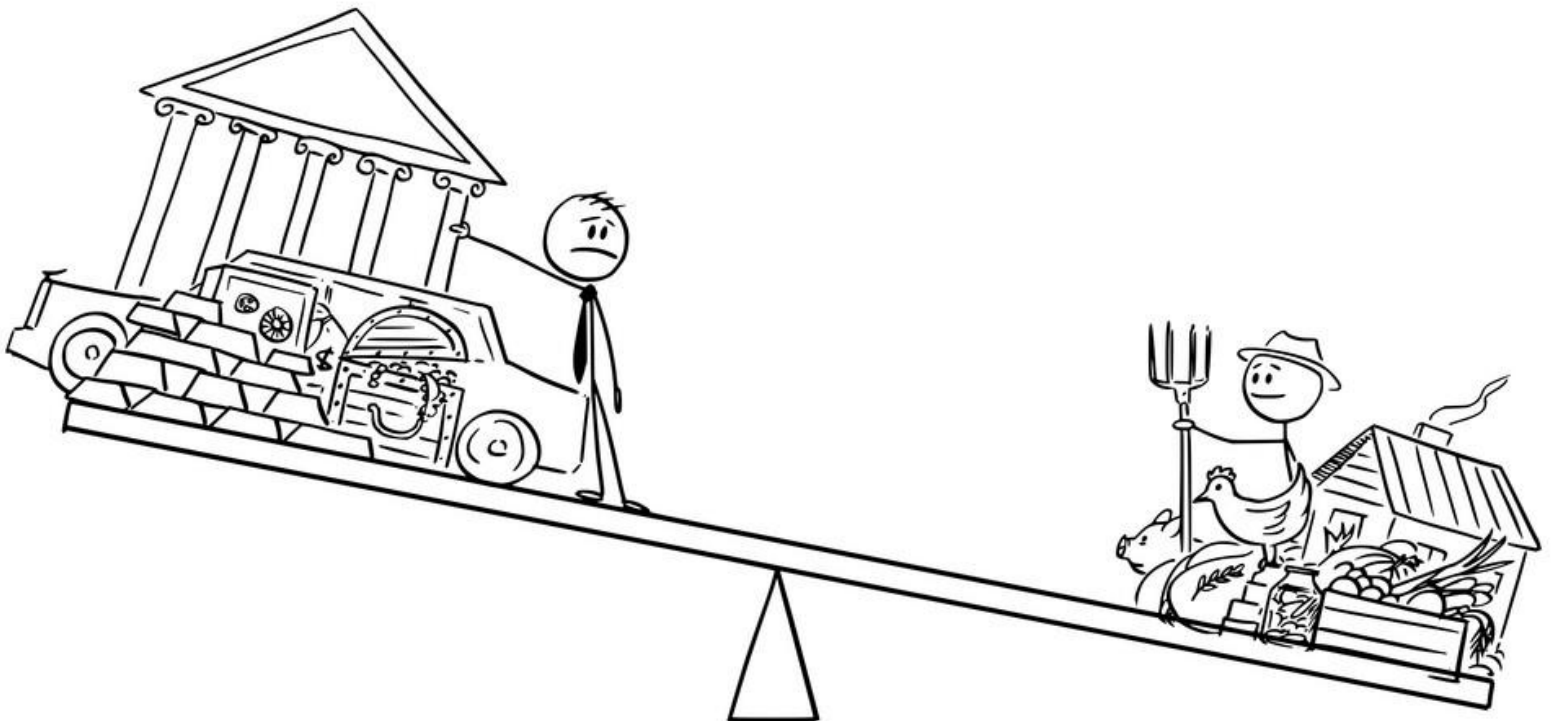


Comparison Between *Income* < 6 group and *Income* ≥ 6 group:

The result of analyses supports the heterogeneity of the moderation process of satisfaction between the lower- and higher-income group. For the lower-income group, the *Income* correlates positively with *Satisfaction* while the thing seems so different for the higher-income group, the *Income* slightly correlates negatively with *Satisfaction*.

For the lower-income group, higher *Income* offers greater sense of *Controllability* while for the higher-income group, the extreme meets, higher *Income* doesn't contribute to higher sense of *Controllability*.

The moderation effect of Controllability on one's overall Satisfaction is confirmed in the lower-income group. The overall Satisfaction group doesn't seem to be controlled through such moderation pathway.



My second question is on the Moderation process of one's *Altruism*.

Variables of Interest:

*Altruism*: min = 1, max = 6. The higher the greater the willingness to be altruistic (利他性)

*Education* (in ordinal categories): The higher the more years spent on education.

1: < 6 years

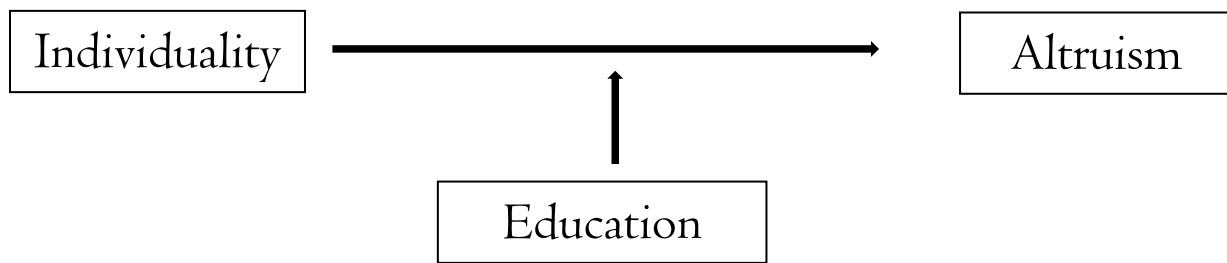
2: 6 – 9 years

3: 9 – 12 years

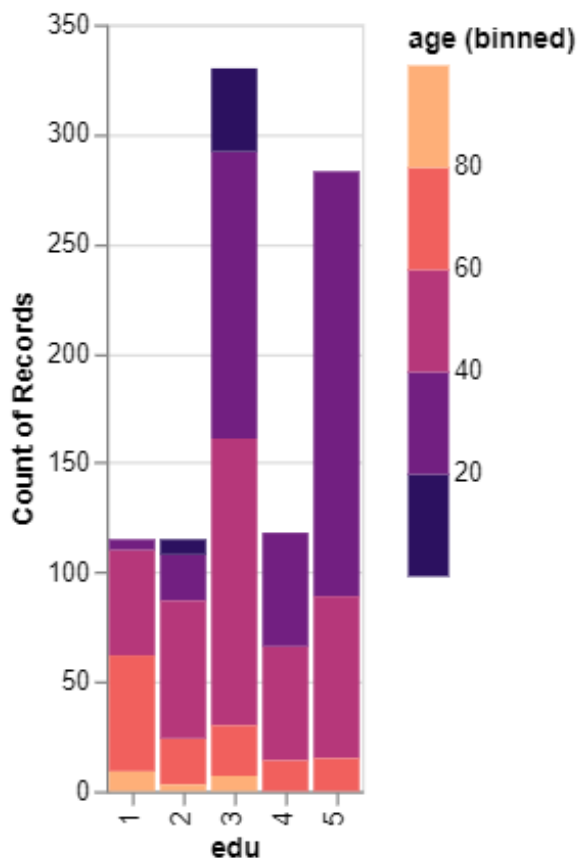
4: 12 – 14 years

5: 14 – 16 years

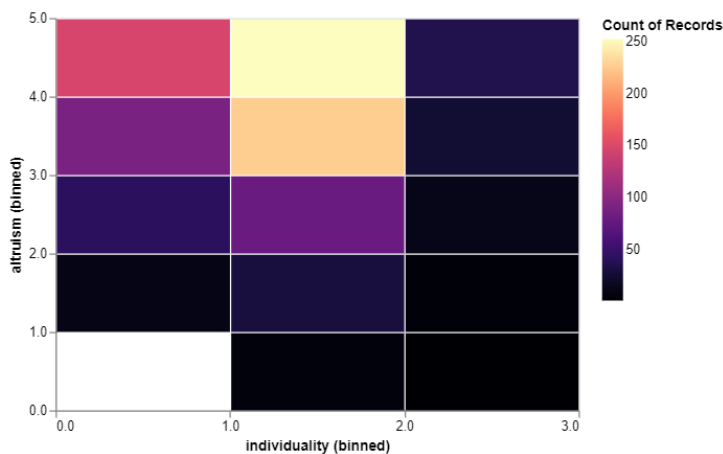
*Individuality*: min = 1, max = 4. The higher the greater sense of individuality (個體自主感)



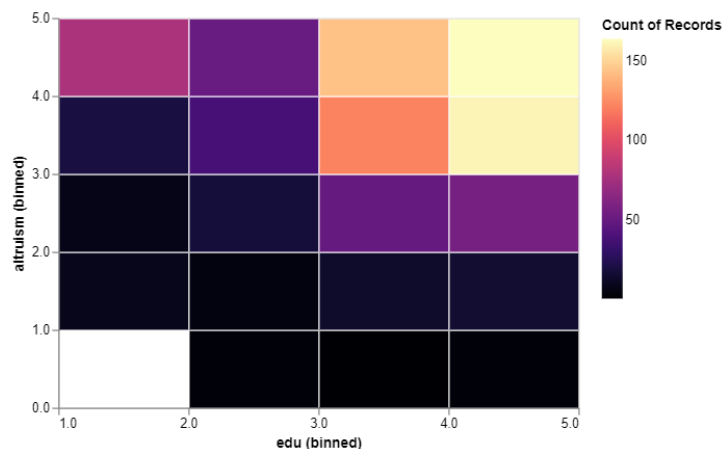
\*\*\* Control: Age



Bar Plot (Education)



Individuality Vs. Altruism



Education Vs. Altruism

```
Call:
lm(formula = altruism ~ age + individuality + edu, data = dta)
```

Residuals:

Min	1Q	Median	3Q	Max
-3.6677	-0.5138	0.0213	0.6984	2.1857

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	2.727150	0.169581	16.082	< 2e-16 ***
age	0.016508	0.002207	7.480	1.68e-13 ***
individuality	-0.156916	0.055510	-2.827	0.0048 **
edu	0.002633	0.027246	0.097	0.9230

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.004 on 957 degrees of freedom  
Multiple R-squared: 0.07041, Adjusted R-squared: 0.06749  
F-statistic: 24.16 on 3 and 957 DF, p-value: 4.474e-15

```
Call:
lm(formula = altruism ~ age + individuality + edu + individuality:edu,
    data = dta)
```

Residuals:

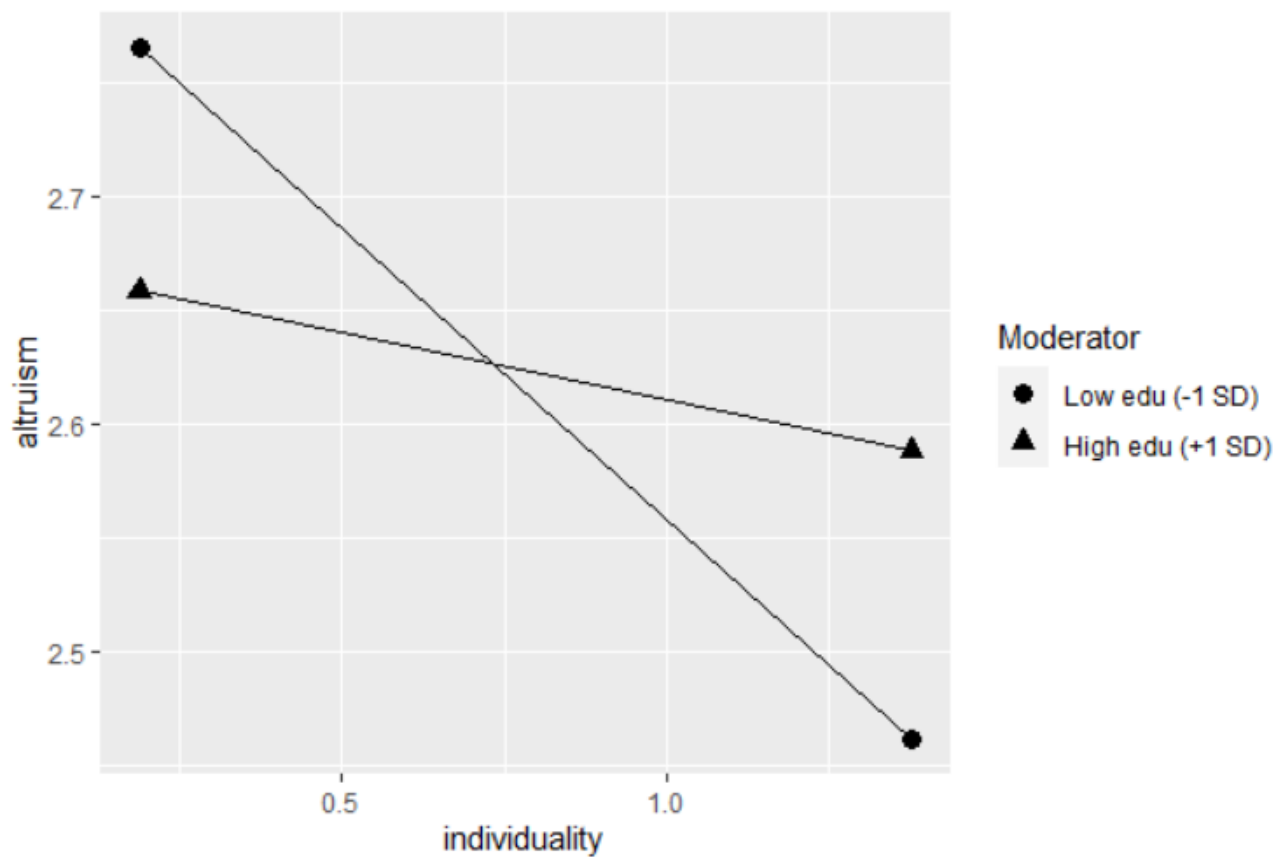
Min	1Q	Median	3Q	Max
-3.6760	-0.5064	-0.0022	0.7434	2.1231

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	2.923813	0.200557	14.578	< 2e-16 ***
age	0.016649	0.002206	7.548	1.03e-13 ***
individuality	-0.406083	0.146941	-2.764	0.00583 **
edu	-0.054256	0.041302	-1.314	0.18928
individuality:edu	0.074081	0.040459	1.831	0.06741 .

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

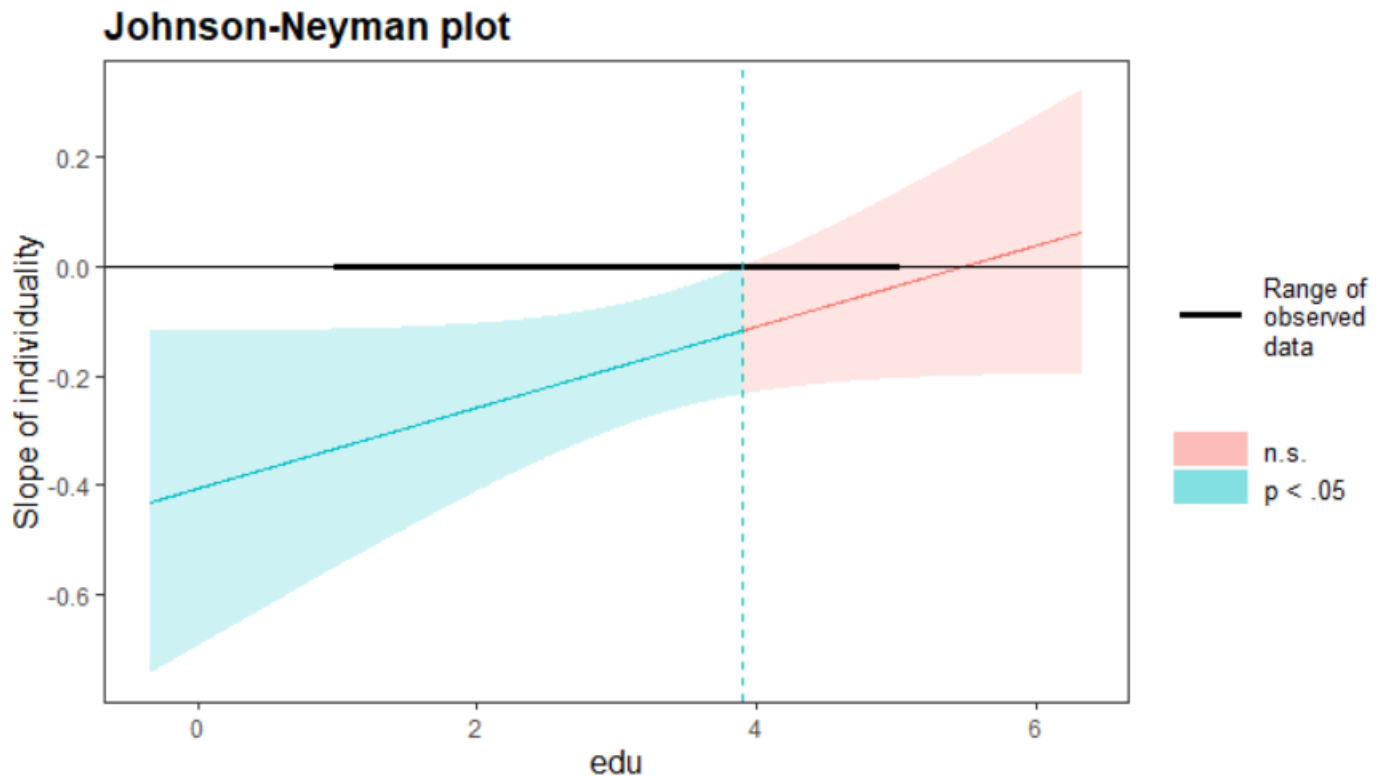
Residual standard error: 1.002 on 956 degrees of freedom  
Multiple R-squared: 0.07366, Adjusted R-squared: 0.06978  
F-statistic: 19 on 4 and 956 DF, p-value: 4.744e-15



\*\* simple slopes analysis ( df= 956 ) \*\*

	simple slope	standard error	t-value	p.value	
Low edu (-1 SD)	-0.2564	0.0776	-3.30	0.001	***
High edu (+1 SD)	-0.0590	0.0770	-0.77	0.444	

The Moderation Graph does indicate that those with higher sense of *Individuality* have lower *Altruistic* attitude. However, *Education* seems to “cushion” the depletion of *Altruistic* attitude as the increase of one’s sense of individuality.



### JOHNSON-NEYMAN INTERVAL

When *Individuality* is within the interval  $[-25.66, 3.90]$ , the p-value of the slope of *Education* is smaller than 0.05.

\*\*\*Note: The range of observed values of individuality is  $[1.00, 5.00]$

Thus, I prone to believe that *Education* moderates the depletion of *Altruistic* attitude as the sense of *Individuality* increases.

Those with higher educational background seem to retain their altruistic attitude better than those with lower educational background. It may be due to the impact of social expectation: Greater Power Comes with Greater Responsibility. Educational system “indoctrinates”, more or less.

With that, I rest my case this time.