Part 3: Report

Echo Chen, Andrew Kroening, Pooja Kabber, Dingkun Yang

November 22nd, 2022

Report Draft

We need to clean data, due to limitation of confusion matrix

df_life_expectancy_2014_cleaned

Full Model

WORDS PLACEHOLDER

Table 1: Logistic Regression Model

	Dependent variable:
	Status_num
Life.expectancy	−0.07 (0.12)
	p = 0.57
percentage.expenditure	−0.0002 (0.0005)
	p = 0.62
BMI	−0.04 (0.03)
	p = 0.14
Total.expenditure	0.22 (0.18)
	p = 0.22
HIV.AIDS	−152.86 (23,893.84)
	p = 1.00
GDP	0.0000 (0.0001)
	p = 0.84
Income.composition.of.resources	41.25 (15.60)
	p = 0.01***
Schooling	−0.06 (0.39)
	p = 0.89
Constant	−12.15 (2,389.40)
	p = 1.00
Observations	131
Log Likelihood	−20.18
Akaike Inf. Crit.	58.36
Note:	*p<0.1; **p<0.05; ***p<0.01

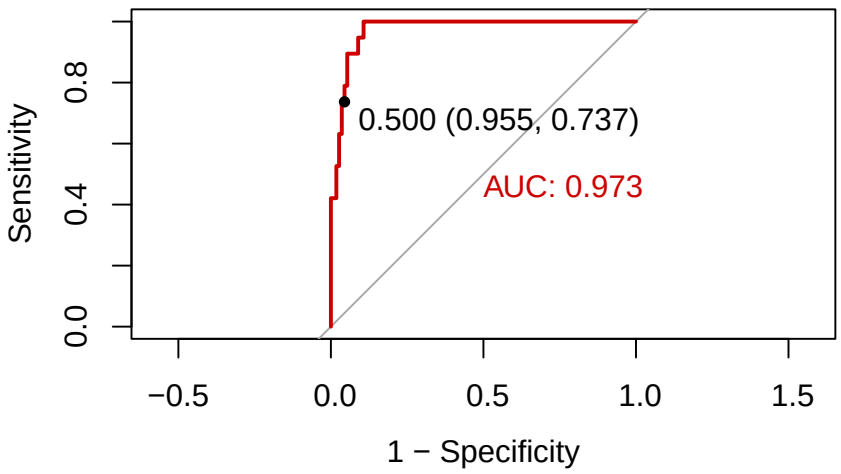
Percentage expenditure and GDP are the ones have higher VIF being over 10, WE NEED TO DECIDE WHICH ONE TO INCLUDE IN THE MODEL

Table 2: Variance Inflation Factors

Life.expectancy	percentage.expenditure	BMI	Total.expenditure	HIV.AIDS	GDP	Income.composition.of.resources	Scho
2.04	10.61	1.29	1.20	1.00	10.56	4.23	2.

Table 3: Confusion Matrix for Full Model

	True Developed	True Developing
Predicted Developed	14	5
Predicted Developing	5	107



Without percentage expenditure

Model 1

Table 4: Logistic Regression Model

	Dependent variable:
	Status_num
Life.expectancy	−0.07 (0.12)
	p = 0.53
BMI	−0.04 (0.03)
	p = 0.14
Total.expenditure	0.20 (0.17)
	p = 0.24
HIV.AIDS	−153.11 (23,950.77)
	p = 1.00
GDP	−0.0000 (0.0000)
	p = 0.52
Income.composition.of.resources	40.86 (15.68)
	p = 0.01***
Schooling	−0.03 (0.40)
	p = 0.95
Constant	−11.67 (2,395.09)
	p = 1.00
Observations	131
Log Likelihood	−20.32
Akaike Inf. Crit.	56.64
Note:	*p<0.1; **p<0.05; ***p<0.01

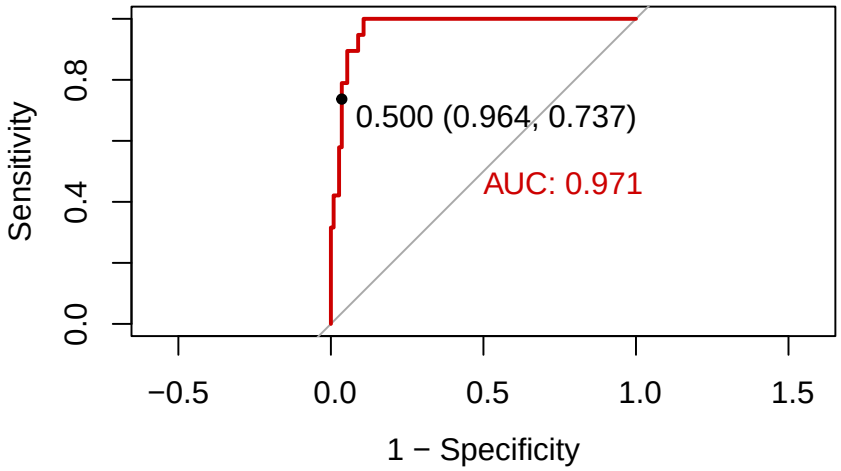
Table 5: Variance Inflation Factors

Life.expectancy	BMI	Total.expenditure	HIV.AIDS	GDP	Income.composition.of.resources	Schooling
2.03	1.28	1.09	1.00	1.62	4.34	2.07

Income composition of resources VIF is near 5, with p-value less than 0.05 (CANNOT DELETE, but maybe delete total expenditure)

Table 6: Confusion Matrix for Model 1

	True Developed	True Developing
Predicted Developed	14	4
Predicted Developing	5	108



Full Model vs Model 1

Table 7: Analysis of Deviance: Full Model vs Model 1

	Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
1	122	40.36			
2	123	40.64	-1	-0.28	0.59

Without GDP

Model 2

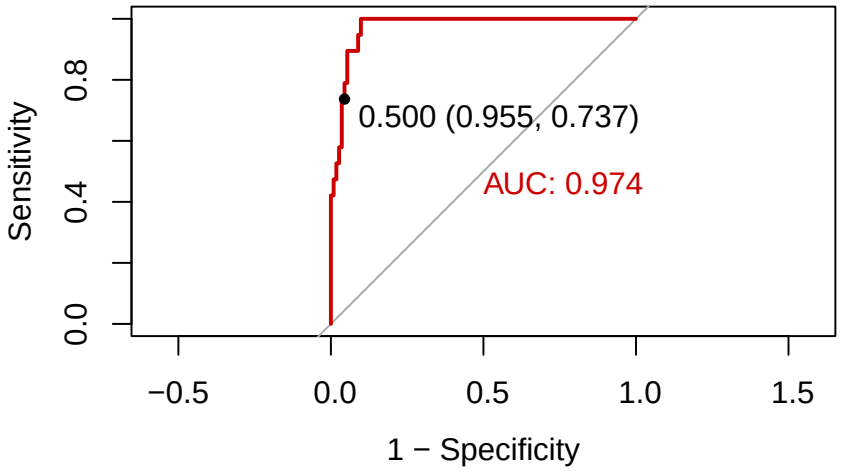
Table 8: Logistic Regression Model

	Dependent variable:
	Status_num
Life.expectancy	−0.07 (0.12)
	p = 0.55
BMI	−0.04 (0.03)
	p = 0.13
percentage.expenditure	−0.0001 (0.0002)
	p = 0.41
Total.expenditure	0.22 (0.17)
	p = 0.22
HIV.AIDS	−153.15 (23,916.24)
	p = 1.00
Income.composition.of.resources	41.73 (15.53)
	p = 0.01***
Schooling	−0.05 (0.39)
	p = 0.89
Constant	−12.26 (2,391.64)
	p = 1.00
Observations	131
Log Likelihood	−20.20
Akaike Inf. Crit.	56.40
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 9: Variance Inflation Factors

Life.expectancy	BMI	percentage.expenditure	Total.expenditure	HIV.AIDS	Income.composition.of.resources	Schooling
2.02	1.31	1.63	1.14	1.00	4.21	2.08

% Error: Unrecognized object type.


Sensitivity
0.0
0.4
0.8
0.500 (0.955, 0.737)
AUC: 0.974
−0.5
0.0
0.5
1.0
1.5
1 − Specificity

Model 1 vs Model 2

Table 10: Analysis of Deviance: Model 1 (W/O percentage expenditure) vs Model 2 (W/O GDP)

	Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
1	123	40.64			
2	123	40.40	0	0.24	

Full Model vs Model 2

Table 11: Analysis of Deviance: Full Model vs Model 2

	Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
1	122	40.36			
2	123	40.40	-1	-0.04	0.83

Using Model 1 (w/o percentage expenditure) to predict out-of-sample (Year 2013) probabilities

Table 12: Confusion Matrix 4

	True Developed	True Developing
Predicted Developed	14	6
Predicted Developing	5	105

Using Model 2 (w/o GDP) to predict out-of-sampl e(Year 2013) probabilities

Table 13: Confusion Matrix 5

	True Developed	True Developing
Predicted Developed	14	5
Predicted Developing	5	106

Model 2 is slightly better, numerical wise. But not statistically different than Model 1

Tested, not good

Without Total Expenditure and percentage expenditure.

Model 3

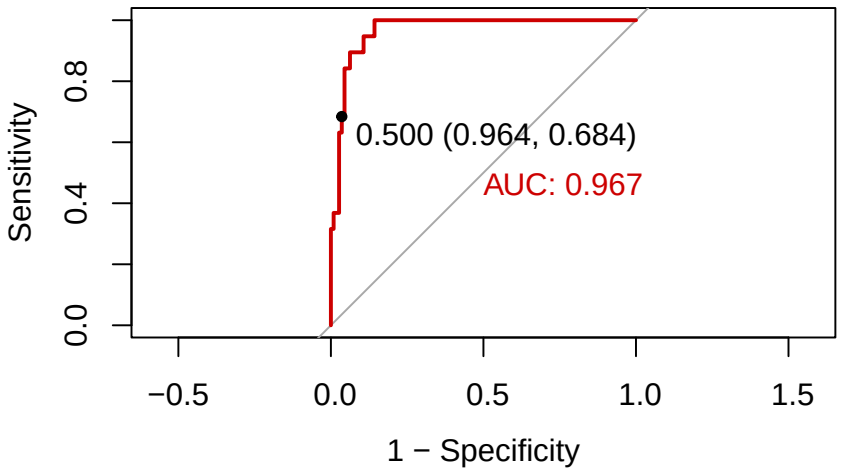
Table 14: Logistic Regression Model

	Dependent variable:
	Status_num
Life.expectancy	−0.04 (0.11)
	p = 0.74
BMI	−0.03 (0.03)
	p = 0.19
HIV.AIDS	−152.18 (24,634.34)
	p = 1.00
Income.composition.of.resources	35.30 (13.99)
	p = 0.02**
Schooling	0.01 (0.39)
	p = 0.98
Constant	−9.94 (2,463.44)
	p = 1.00
Observations	131
Log Likelihood	−21.27
Akaike Inf. Crit.	54.54
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 15: Variance Inflation Factors

Life.expectancy	BMI	HIV.AIDS	Income.composition.of.resources	Schooling
2.01	1.19	1.00	3.62	2.13

% Error: Unrecognized object type.



Model 1 vs Model 3

Table 16: Analysis of Deviance: Model 1 vs Model 3

	Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
1	123	40.64			
2	125	42.54	-2	-1.90	0.39

Full Model vs Model 3

Table 17: Analysis of Deviance: Full Model vs Model 3

	Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
1	122	40.36			
2	125	42.54	-3	-2.18	0.54