



Predicting Life Satisfaction in Ecuador Using Logistic Regression: Insights from the World Values Survey Wave 7 (2017-2022)

Presented By:

Name	Code
Kerlous Marzouk Rady	5210853
Verina Alfred Fahmy	5210407
Farah Hossamedin Mustafa	5210468

Supervision: Dr. Laila El Zeini

Introduction:

The World Values Survey (WVS) is a renowned global project that explores people's values, beliefs, and cultural habits across different countries. Wave 7 of the survey (2017-2022) provides a vast amount of data on various aspects of life, such as life satisfaction, across diverse societies. Ecuador's participation in this wave allows us to gain a better understanding of the factors that affect life satisfaction among its citizens.

For our analysis, we conducted a logistic regression model to explore the factors influencing life satisfaction in Ecuador. We selected variables that we believed to be strongly related to life satisfaction, including education level, gender, marital status, employment status, age, urban vs. rural location, trust in family, freedom, faith, income, belief in success, and political interest.

As part of our exploration, we tested an interaction term between **gender** and **employment status**, hypothesizing that the relationship between employment and life satisfaction might vary by gender. The coefficient for this interaction term was found to be statistically significant, indicating that gender and employment status interact in influencing life satisfaction.

Next, we applied **stepwise regression** using both **backward** and **forward** selection methods to refine the model. Only the significant variables were retained in the final model.

We finalized the model based on the last iteration from the stepwise regression process. Subsequently, we tested the key assumptions of the model, such as **multicollinearity**, and confirmed that these assumptions were satisfied. To assess the model's goodness of fit, we calculated the **Area Under the ROC Curve**, which was 64%. Additionally, we conducted other goodness-of-fit tests, including the **likelihood ratio** and **classification table**, all of which produced satisfactory results, confirming that the model fit the data well.

Final model:

coefficient	Estimate	Odds ratio	Standard error	Z-value	P-value
Intercept	1.6589	5.253747	0.2916	5.875	4.22e-09 ***
Education Middle	0.1647	1.1791	0.1823	0.906	0.364686
Education High	0.6651	1.9446	0.2366	2.829	0.004665 **
Female	-0.3755	0.6869	0.2467	-1.536	0.124645
Unemployed	-0.8685	0.4195	0.2440	-3.569	0.000358 ***
Age (30:49)	-0.4857	0.6152	0.2024	-2.411	0.015897 *
Age (50 and more)	-0.1761	0.8385	0.2218	-0.801	0.423281
Not trust my family	-0.5590	0.5717	0.1951	-2.881	0.003968 **
High Income	0.5784	1.7908	0.1581	3.694	0.000220 ***
Success by luck	0.5784	1.7832	0.1862	3.109	0.001874 **
Female*unemp	0.6695	1.998	0.3287	2.041	0.041233 *

Table 1 summary of final model

- **Age:** Categorized into three intervals: "young" (reference category), "middle-aged", and "older". "Young" serves as the baseline to analyze age's impact on life satisfaction.
- **Education:** Divided into three levels, with "low" as the reference category, reflecting the individual's educational attainment.
- **Income:** Categorized into two levels, with "low" as the reference category, indicating the income group.
- **Success:** This variable has two levels, with the reference category being "success through hard work". It reflects individuals' beliefs about the source of success, whether they attribute it to hard work and effort or to luck and external factors.
- **Trust in Family:** Two levels, with the reference category being "trust in family", examining the role of family trust in life satisfaction.
- **Employment:** Two levels, with "employees" as the reference category, distinguishing between employed and non-employed individuals.

Interpretations:

- **Θ^1 : 1.1791**
Individuals with a middle education level are estimated to have 18% higher odds of experiencing greater life satisfaction compared to those with a low education level, accounting for other factors such as sex, income, and employment status. Although this coefficient is **insignificant**, it represents a dummy variable for education, which has other significant dummies in the model. Therefore, it is retained in the analysis to provide a complete interpretation of the education variable.
- **Θ^2 : 1.9446**
Individuals with a high education level are estimated to have 94% higher odds of experiencing greater life satisfaction compared to those with a low education level, while controlling for variables like sex, income, and employment status.
- **Θ^3 : 0.6869**
Being female is associated with 31% lower odds of higher life satisfaction compared to being male, assuming employment status is constant and without considering the interaction effect with employment. However, this effect is **insignificant**, but as a researcher, it is important to retain the main effect, as it is part of an interaction term in the model.
- **Θ^4 : 0.4195**
Unemployed individuals are estimated to have 58% lower odds of experiencing higher life satisfaction compared to employed individuals, holding gender constant and without accounting for the interaction effect with gender.
- **Θ^5 : 0.6152**
Middle-aged individuals are estimated to have 38% lower odds of higher life satisfaction compared to younger individuals, considering other factors like education, income, and gender.
- **Θ^6 : 0.8385**
Older individuals are estimated to have 16% lower odds of experiencing higher life satisfaction compared to younger individuals, controlling for other factors such as education, income, and gender. Although this effect is **insignificant**, it represents a coefficient for a dummy variable, and the other dummies in the model are significant. Therefore, it is retained in the analysis to provide a complete interpretation of the age variable.
- **Θ^7 : 0.5717**
Individuals who lack trust in their families are estimated to have 43% lower odds of greater life satisfaction compared to those who trust their families, while controlling for factors like education, income, and gender.
- **Θ^8 : 1.7908**
Individuals with a high-income level are estimated to have 79% higher odds of

experiencing greater life satisfaction compared to those with a low-income level, accounting for variables such as education, age, and gender.

- **Θ^9 : 1.7832**

Believing that success is determined by luck is associated with 78% higher odds of greater life satisfaction compared to believing that success is determined by hard work, controlling for factors like education, income, and gender.

- **Θ^{10} : 1.998**

The interaction term indicates that the combined effect of being female and unemployed results in nearly double the odds of higher life satisfaction compared to male employees. This suggests that the relationship between gender and life satisfaction is influenced by employment status, and vice versa.

Model diagnostics and assumptions:

This method run different models which contain in each model one of the explanatory variables as a response and study the association between those variable and other explanatory variables, if the GVIF is greater than 10, then there is multicollinearity among the variables and otherwise there is no multicollinearity among the variables.

Table 2:Checking Multicollinearity

variable	GVIF
edu	1.23662151
sex	2.50222431
mar1	1.35349107
empl	2.54942851
age	1.59703462
ur	1.04407395
tr1	1.04659275
free	1.02637607
faith	1.04410061
inc	1.02411484
suc1	1.01225091
chief	1.29498619
sex:empl	4.14317148

So, from this table we can conclude that there is no multicollinearity among the explanatory variables as all GVIF values less than 10.

Goodness of fit:

1] Likelihood Ratio test (Deviance).

The Final model after stepwise is nested within the full model.

The hypotheses:

H₀:The model is True. vs **H₁:**The model is not true.

Deviance	3.0091
P-Value	0.6986

SO At significance level $\alpha = 0.05$, we can not reject H₀ So the model is true.

2] Hosmer-Lem show test.

As the variable age is continuous variable so we used Hosmer-lem show test

The hypotheses:

H₀:The model is true vs **H₁:**The model is not true.

Chi-square value (χ^2)	4.6877
Degrees of freedom	8
P-value	0.7904

SO At significance level $\alpha = 0.05$, we can not reject H₀ So the model is true.

3] McFadden's Pseudo R squared.

- Final Model = 0.06290107

Based on McFadden R squared, it is considered a good model as the values of McFadden R squared with logistic regression takes small values than adjusted R squared in linear regression.

4] Classification table.

The corresponding classification table will be: (with cutoff point 0.8)

Classified	True		
	Yes	No	Total
Yes	731	98	829
No	259	112	371
Total	990	210	1200

The proportion of correct classification to be satisfied with your life

(Sensitivity) = 88.17%.

The proportion of correct classification not to be satisfied

(Specificity) = 30.18%.

The proportion of overall correct classification = 70.25%.

Since the overall correct classification of our final model equal to almost 70% and this is an indication that the model is good in classify the persons in the sample satisfied with their life or not since the overall correct classification is greater than 60% and the final model has acceptable predictive power.

5] Roc curve.

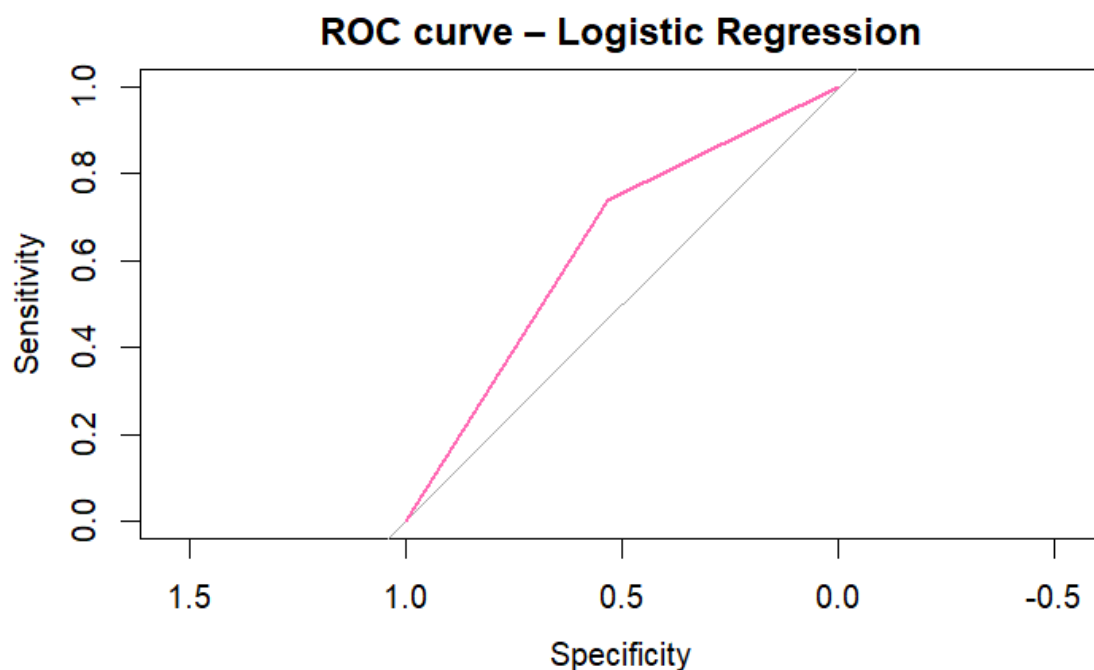


Figure 1: shows the ROC curve and the Area under the curve.

From figure 1, the area under the curve (AUC) equal 0.64, then AUC is significantly greater than 0.5 which is better than the random assignment model (null model).

Conclusion:

The analysis of the World Values Survey data for Ecuador showed some interesting results about life satisfaction. People with higher education, better jobs, and higher income reported greater life satisfaction. On the other hand, unemployment and low trust in family had a negative effect. Surprisingly, those who believed that success comes from luck rather than hard work were more likely to feel satisfied with life. We also noticed that unemployed women had higher life satisfaction than unemployed men.

The model we used to study this data worked well, predicting life satisfaction with 70% accuracy. These findings can help policymakers and researchers understand what makes life more satisfying in Ecuador.