**HUMAN TRAFFICKING IN ROMANIA**

Authors: Boros Ștefan, Gherman Andrei, Drăgușin Cristian

Year: 2022

**INTRODUCTION:**

We studied the phenomenon of human trafficking in Romania by analyzing the dataset composed of all the human trafficking victims that have been rescued in Romania in 2019 (N = 697).

We have attempted to find, based on the sample, information about the age of all the human trafficking victims in the Romanian population, their gender distribution, the relationship between the age of the victims and the region of the county in which they were found in or the method of trafficking, as well as developing two regression models that try to predict the number of victims found in a county based on certain numerical variables.

We chose to study this phenomenon due to the overwhelming lack of literature on the topic in Romania (other than yearly reports from government institutions) as well as due to how big of a problem it is thought to be in Romania specifically.

**LITERATURE REVIEW:**

A European Commission report published last October concluded that Romania had the highest rate of human trafficked victims per million inhabitants in the EU, at 74 per million, followed by Hungary, at 64, and Bulgaria, at 40. Most of the victims were women who were subjected to sexual exploitation.

The Organized Crime and Terrorism Investigation Directorate (DIICOT) and the Department for Combating Organized Crime (DCCO) were responsible for investigating and prosecuting trafficking cases. DIICOT and DCCO reported pandemic-related restrictions and infections among police and prosecutors constrained their ability to open new investigations and process ongoing cases.

In 2020, approximately 35 percent (210) of identified victims received assistance from public institutions, public-private partnerships, and NGOs, a decline from 49 percent in 2019 and 48 percent in 2018.

The government maintained insufficient protection efforts. Authorities used the existing national victim identification and referral mechanism to identify victims and refer them to care. While ANITP drafted a new referral mechanism in 2018 with the support of NGOs, the government did not implement it for the third straight year. In 2020, the government changed its reporting methodology under the National Mechanism for Identification and Referral of Victims to allow any entity, including diplomatic missions, NGOs, and other organizations, to identify victims. Authorities identified 596 victims (429 sex trafficking, 131 labor trafficking, and 36 victims of attempted or uncategorized trafficking), compared with 698 in 2019 and 497 in 2018. Of the 596 victims, 255 were children. Authorities identified one foreign victim (zero in 2019 and 2018), but observers estimated there were numerous cases, particularly among asylum-seekers. In 2020, approximately 35 percent (210) of identified victims received assistance from public institutions, public-private partnerships, and NGOs, a decline from 49 percent in 2019 and 48 percent in 2018.

The government’s annual report of human trafficking in Romania identified an increase of 40% of victims found since the previous year. There are no specific trends for 2019 observed in the population of identified victims compared to previous years. The next year (2020) also showed a decrease in victims, such that the year 2019 was the one with the most victims found. We wonder if this is because it was the year in which the COVID-19 pandemic started, and whether this actually increased the human trafficking phenomenon or whether it actually made the already-existing victims easier to rescue. A few findings of their report are:

- The higher the age of the victims, the higher the likelihood of victims being trafficked externally.

- The likelihood of being trafficked internally is higher in the case of female victims.

- The lower the age, the higher the likelihood of sexual exploitation.

- The lower the age of the victims, the more likely they are to be recruited by strangers.

- The average period of time a victim is trafficked is 1 year, with a minimum of a few days and a maximum of 12 and a half years.

**DATA USED:**

Our data includes a sample of N = 697 victims that have been found and rescued in Romania in 2019, with information about their age, gender, citizenship, destination of the country they were (about to be) sent in, education level, urban/rural split, recruitment, exploitation method and the relationship with the recruiter. Source: <https://data.gov.ro/dataset/victime-trafic-de-persoane-2019>

We also gathered data about economic indicators split by county in Romania (in 2019): GDP/capital, average net salary, unemployment rate, number of retired people per employee. Sources:

https://panorama.ro/interactiv-economia-judetelor-romaniei/

https://www.eesc.europa.eu/sites/default/files/resources/docs/qe-01-15-435-ro-n.pdf

https://economie.hotnews.ro/stiri-finante\_banci-24269496-trei-indicatori-care-arata-unde-mai-buna-viata-salariatior-din-romania-topul-judetelor-vulnerabile.htm

**RESEARCH METHODOLOGY:**

We divided up the victims into 5 strata depending on the method of their exploitation: begging, sexual exploitation, work exploitation, theft or a failed attempt at exploiting them. We computed the average age of the victims in each of those strata and then applied the method of **stratified random sampling without replacement** in order to find out the average age of all the victims in the population, with a Cl of either 95% of 98%.

We tested two hypotheses: that the average age of a victim in the population is less than 20 (one-tailed test) and that 80% of victims in the population are women (two-tailed test). We used the student’s t-distribution to make a t-test testing the validity of each of those hypotheses (Cl = 95%). In the two-tailed test, the gender variable was a Boolean one, where we assigned 1 for female and 0 for male in order to compute variances and averages.

We applied the one-way ANOVA method, two times, to test whether there are statistically significant differences in the variation of the victim’s age in a county between the regions of each county, as well as to test whether there are statistically significant differences in the variation of a victim’s age between the methods of exploitation of each victim.

We have created a simple linear regression model, using the method of the sum of least squares (in Excel), to attempt to predict the number of victims found in a county (relative to its population) based on that county’s GDP/capital.

We have created a multiple linear regression model to predict the number of victims found in a county (relative to its population) based on that county’s GDP/capital, average net salary, unemployment rate, number of retirees per employer, and the dummy variable of whether the county is in the Eastern region of Romania or not, which interacts with the GDP/capital. We chose East from the popular belief that the Eastern region is more likely to be “a paradise for human traffickers”.

**RESULTS:**

**1). Stratified random sampling:**

We can say with 95% certainty that the average age of victims in the population in the interval [20.38; 21.56], and with 98% certainty that it is in the interval [20.33; 21.61].

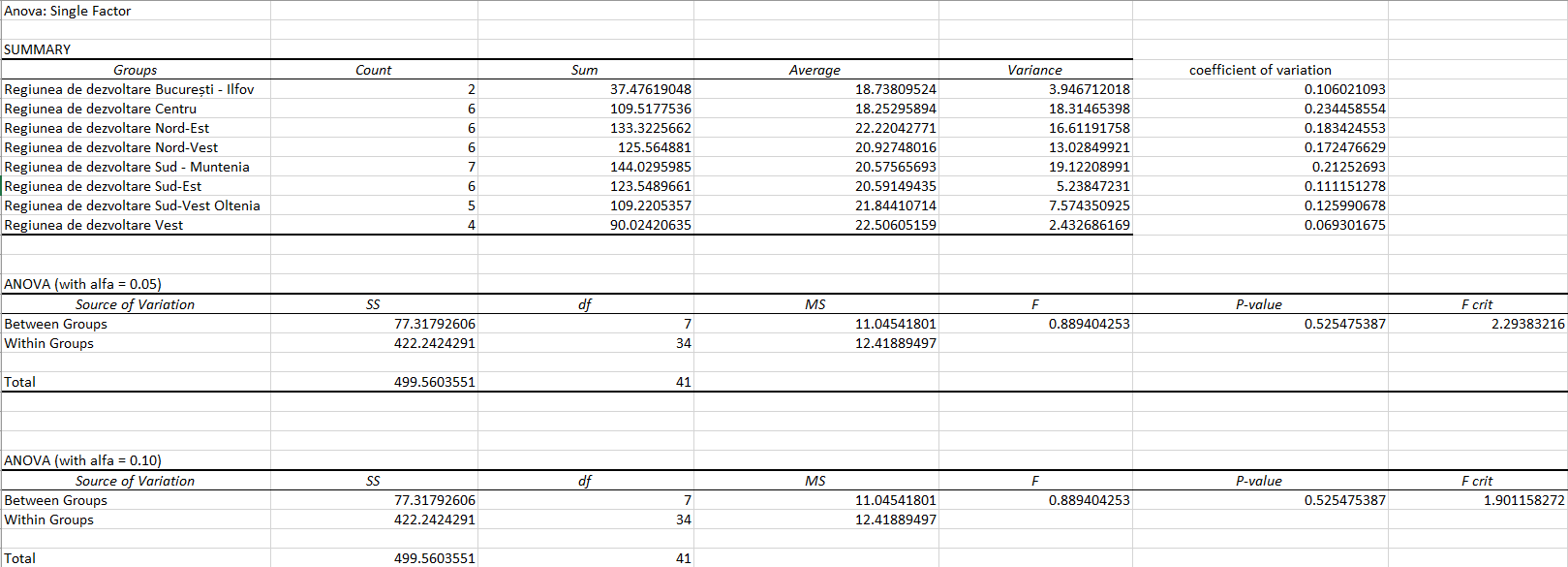
**2). Hypothesis testing:**

The hypothesis that the average age of a victim is the population is less than 20 is REJECTED at the Cl of 95% (i.e.: there is more of a 5% chance that it is greater than 20), since our t-value is 3.088613 which is greater than t alpha when alpha = 0.05 (1.647). 20.97 was the average of a victim in our sample.

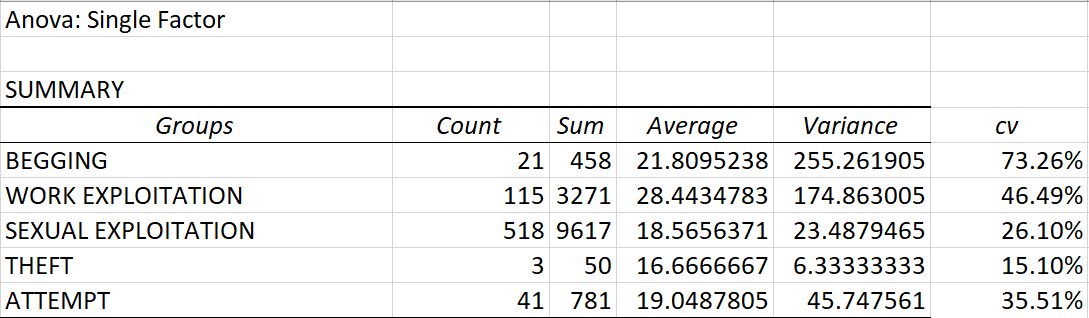
The hypothesis that 80% of the victims in the population are women is REJECTED at the Cl of 95% (i.e.: there is more of a 5% chance that the percentage of women is not 80%), since our t-value is 2.37795382, which is greater than t alpha/2 when alpha = 0.05 (1.96). 83.35% of victims in our sample were women.

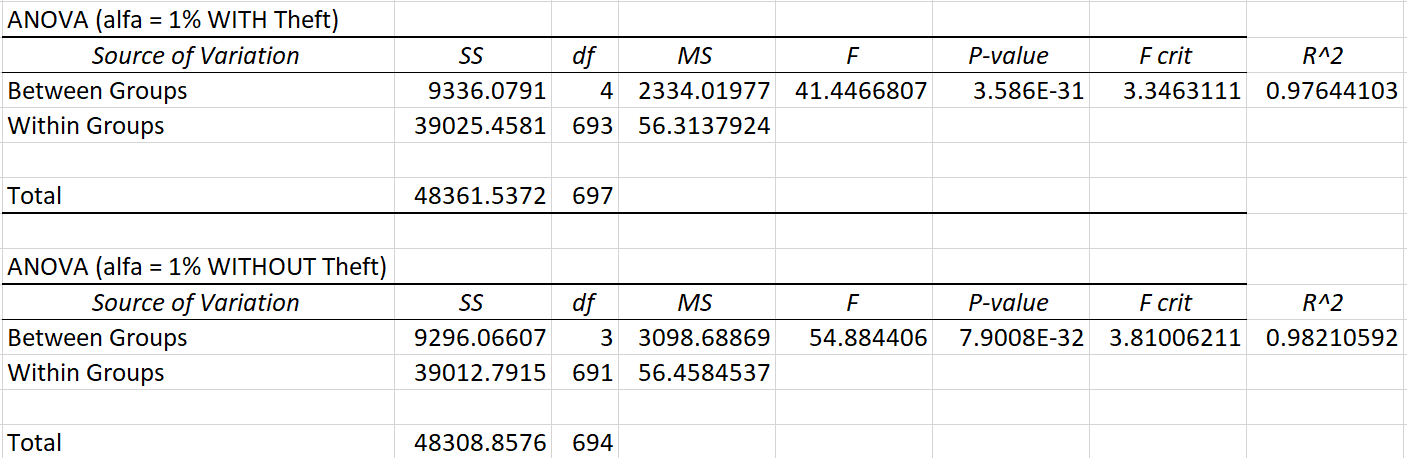
**3). ANOVA**

With an F-value greater than F-crit at both a 95% and a 90% Cl, as well as a p value of 52.54%, we can safely say that the variation of age between regions of the country is statistically insignificant:



In the case of the exploitation methods however, there were statistically significant differences in the variation of the age of the victims found, even at confidence levels of 99%:

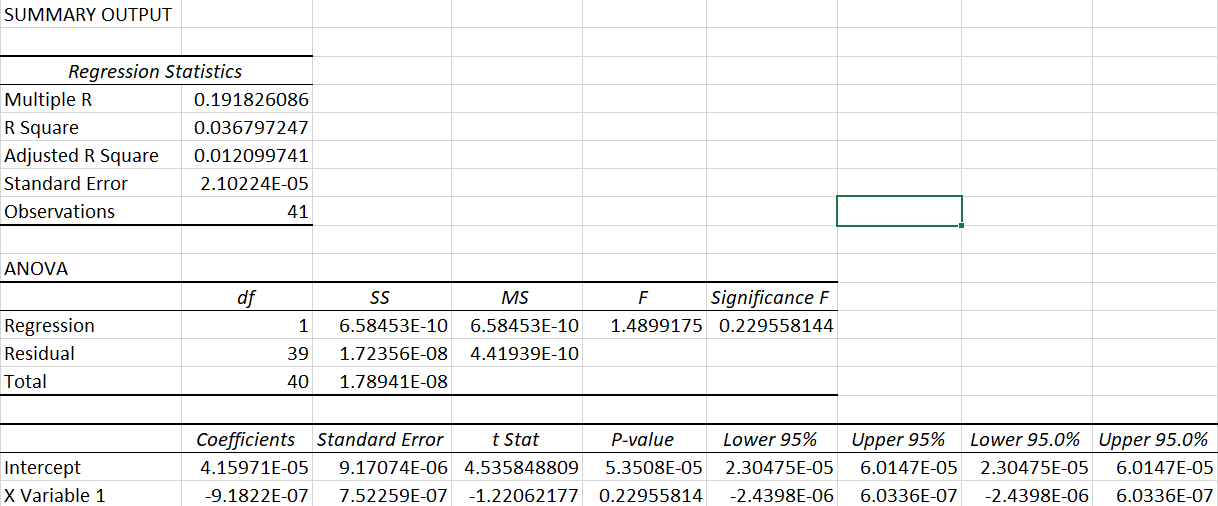




Since we only have 3 victims who have been exploited by theft, we have computed the ANOVA method with and also without taking them into account. If we take the 3 victims exploited by theft into account, the exploitation method explains 97.64% of the variation in age. If we don’t, it explains 98.21%. F is bigger than F-crit and the p-value is almost infinitesimally small in both cases, so the differences are statistically significant.

**4). Simple linear regression**

For our simple regression model, we have only been able to predict R2 = 3.67% of the variation in number of victims relative to population between counties. While this means that the model is not very practically useful, F is bigger than F-crit, so it is a statistically significant model (it is better than an intercept-only model, than a “horizontal line”):



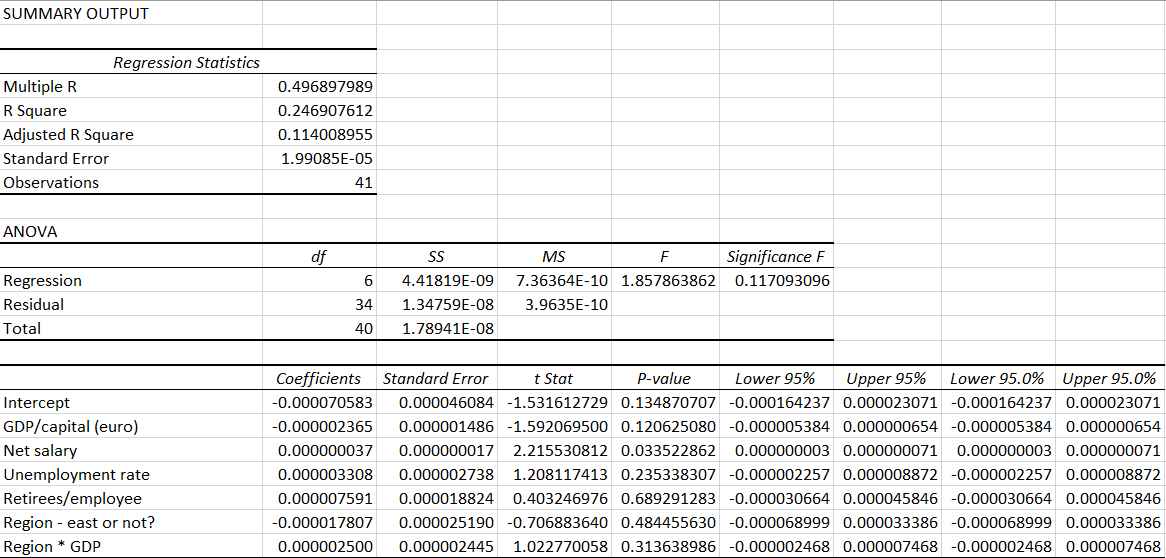
Here are the residuals, where we’ve also added the ratio between the residual and the predicted value (to show how much the model was “off” in each county):

|  |  |  |  |
| --- | --- | --- | --- |
| RESIDUAL OUTPUT |  |  |  |
|  |  |  |  |
| *Observation* | *Predicted Y* | *Residuals* | *Residual/Predicted Y* |
| 1 | 2.84665E-05 | -2.0479E-05 | -71.94% |
| 2 | 2.95684E-05 | -1.25777E-05 | -42.54% |
| 3 | 2.9752E-05 | 3.26594E-06 | 10.98% |
| 4 | 3.33331E-05 | 3.0159E-05 | 90.48% |
| 5 | 3.17721E-05 | 2.2788E-06 | 7.17% |
| 6 | 3.26903E-05 | -1.44032E-05 | -44.06% |
| 7 | 3.52613E-05 | 2.41883E-05 | 68.60% |
| 8 | 3.25985E-05 | 3.11951E-05 | 95.70% |
| 9 | 2.69055E-05 | 1.24594E-05 | 46.31% |
| 10 | 1.15712E-05 | 1.09874E-06 | 9.50% |
| 11 | 3.32412E-05 | -1.39238E-05 | -41.89% |
| 12 | 3.1313E-05 | 9.74109E-08 | 0.31% |
| 13 | 3.35167E-05 | -1.73694E-05 | -51.82% |
| 14 | 2.562E-05 | -1.60606E-05 | -62.69% |
| 15 | 2.58955E-05 | 6.76652E-06 | 26.13% |
| 16 | 3.2874E-05 | -2.8458E-05 | -86.57% |
| 17 | 3.33331E-05 | 1.21117E-06 | 3.63% |
| 18 | 3.15884E-05 | 6.55961E-05 | 207.66% |
| 19 | 3.32412E-05 | 2.58634E-05 | 77.81% |
| 20 | 3.2874E-05 | 2.96969E-05 | 90.34% |
| 21 | 2.98438E-05 | 1.7588E-05 | 58.93% |
| 22 | 3.29658E-05 | -2.99474E-05 | -90.84% |
| 23 | 3.18639E-05 | -1.00082E-05 | -31.41% |
| 24 | 3.22312E-05 | 1.83275E-05 | 56.86% |
| 25 | 2.93847E-05 | -1.33734E-05 | -45.51% |
| 26 | 3.27821E-05 | -2.12773E-05 | -64.91% |
| 27 | 3.41595E-05 | -5.54615E-06 | -16.24% |
| 28 | 3.22312E-05 | 1.67382E-05 | 51.93% |
| 29 | 3.39758E-05 | -1.46289E-05 | -43.06% |
| 30 | 3.40677E-05 | -1.95829E-06 | -5.75% |
| 31 | 2.84665E-05 | -8.29648E-06 | -29.14% |
| 32 | 3.14048E-05 | 3.40226E-05 | 108.34% |
| 33 | 3.2874E-05 | -2.77102E-05 | -84.29% |
| 34 | 2.76401E-05 | -4.11409E-06 | -14.88% |
| 35 | 3.45268E-05 | -2.66309E-05 | -77.13% |
| 36 | 3.42513E-05 | -1.0046E-05 | -29.33% |
| 37 | 2.59873E-05 | -7.41804E-06 | -28.54% |
| 38 | 3.16803E-05 | 1.47534E-05 | 46.57% |
| 39 | 3.20476E-05 | -1.44103E-05 | -44.97% |
| 40 | 3.5445E-05 | -1.15551E-06 | -3.26% |
| 41 | 3.37004E-05 | -1.55137E-05 | -46.03% |

We can conclude from the very small R2 value of 3.67%, from the residual output, and even visually, from the scatter-plot, that our simple linear regression model is not a very accurate model in practice.

**5). Multiple linear regression:**

These are the results of the multiple linear regression model:



The R2 value is 24.69%, so the model can predict 24.69% of the variation in the number of victims relative to a county’s population. However, the value we should pay more attention to is the *adjusted* R2, which is only 11.4%. The adjusted R2 is a value that is closer to the model’s predictive ability, since it adjusts the R2 value based on the number of parameters we have put in, in order to avoid overcomplicating a model (which very often runs into the issue of attempting to model the “random noise” in data by adding too many insignificant variables). The most insignificant parameter is the number of retirees per employee, with a p-value of 68.9%, with the most significant one being the net salary.

Regardless, the F-value is bigger than F-crit (1.85 > 0.11), hence the model being a statistically significant one (better than an intercept-only model).

The big surprise of this model came with the Boolean variable, which is the only negative parameter other than the GDP/capital. This means that Eastern counties tend to have less victims relative to their counties’ population. However, the p-value of this parameter is as big as 48.44%, so this conclusion should be taken with a grain of salt. It is safer to assume that the impact of region (eastern vs. non-eastern) on human trafficking is unknown or unpredictable.

**LIMITATIONS:**

Our data is extracted from the government’s official reporting of the number of victims saved and rescued, and therefore is a *biased* sample, arguments being possible to make as to how it is not random enough, and hence not representative of all the victims that have ever been trafficked. For example, maybe victims of a certain age/exploitation method are easier or harder to rescue than of other ages/methods/etc., or maybe victims in certain regions are easier to rescue. In this light, a point of improvement for our study would be an adjustment for this sampling bias.

Another limitation is that our data is only from 2019, which was also the start of the COVID-19 pandemic, and we have not controlled for that factor. A point of improvement for such research would be the analysis of panel data, taking into account *all* the data provided by the government from all years about the victims rescued.

**CONCLUSIONS:**

In terms of our analysis, it is safe to say that the phenomenon of human trafficking in Romania is hard to predict. Out of all our statistical tests, the only *certain* relationship that we can draw is between the age and the method of exploitation of a victim, with R2 values so high that they stand even in front of our inevitable sampling bias.

**BIBLIOGRAPHY:**

https://www.state.gov/reports/2021-trafficking-in-persons-report/romania/

<https://balkaninsight.com/2021/06/03/romania-scraps-statute-of-limitations-for-human-trafficking/>

https://anitp.mai.gov.ro/ro/docs/studii/Raport%20anual%20privind%20fenomenul%20traficului%20de%20persoane%20in%202019.pdf