Session:
May 2020

Mathematics HL
Internal Assessment

GROWTH MODELS

How can we use mathematical models for explaining population dynamics in various countries?

1. Introduction

Estimates say that the country where I live, Croatia, has lost about 500,000 inhabitants in the last 5 years due to socio-economic reasons. Precisely because I live in the country with negative demographic trends to which I am exposed, I have decided in this exploration to research mathematical models of population growth.

Demographic trends are one of the key drivers for the development of national economies. Demographic trends affect well-being, culture and ultimately the lives of all people. Therefore, population projections are quite interesting to me. For example, I am impressed by the movement of the population in India compared to Croatia, where the annual increase of India's population is equal to the three times the population of Croatia.

A better understanding of such models will certainly help me in later education and show me how to act in certain situations, especially in a world where the maximum capacity of people will be reached very quickly.

2. Aim

My goal would be to compare the results obtained from using two population growth models with the actual data, and then to determine which of the models used is closest to the actual data. Throughout my work, I will learn about the properties of the logistic and exponential function in the application of the Indian population. After that, I will apply the newly acquired knowledge to the data of the Croatian population. Finally, I will explain how the model fits with the real data and give the true historical and mathematical reasons for these agreements.

Therefore, my research question is:

How can we use mathematical models for explaining population dynamics in various countries?

To identify differences in the application of two models, I compared results with the real data obtained from the Indian Office of the Registrar General & Census Commissioner in the period from 1901. to 2011. Also, I did the same analysis with the data obtained from the Croatian Statistical Office encompassing the period from 1857. to 2011.

3. Investigation

To answer my research question, I will examine two models. Before their application, I need to get acquainted with their properties, learn the definitions and the assumptions under which the models are valid.

3.1. Malthusian law of population growth (exponential law)

The first assumption that must be made when dealing with the population of any species is that growth functions change continuously and are differentiable with respect to time. The second assumption is that the population's growth rate stays the same, regardless of the population size. Such a model is applicable if the change in population rate is proportional to the population. If P=P(t), then:

$$\frac{dP}{dt} \propto P \qquad or \qquad \frac{dP}{dt} = rP$$

Where P is the number of population at a given time and r is constant of proportionality, the difference between growth rate and death rate. If more people died in the year than were born, then the constant of proportionality has a negative sign (exponential decay), and if more are born then it is positive (exponential growth). This is a simple differential equation and I will try to find the solution. If P=P(t) then suppose that at some initial time t = 0, the sample contains the initial value of the number of people P(0). After a certain period of time t, the number of people changed to a certain value of P(t). In order to integrate I made the change of variables so that the arguments being integrated are at their differentials. After solving the integral, an elementary function is obtained in which the boundary values must be included:

$$\int_{P_0}^{P} \frac{dP}{P} = \int_0^t r dt \quad \rightarrow \quad \ln\left(\frac{P}{P_0}\right) = rt$$

Using the laws of logarithms, this equation takes a more elegant form:

$$P(t) = P_0 e^{rt}$$

Where P(t) is a population at a given time. The Malthusian growth law is not precise for a population over a long period. It does not take into account immigration and emigration. The function also predicts that for the positive r, $P(t)\ tends\ to\ \infty$ as time passes (grows), which is physically impossible. To see that this is true, I worked on a second derivation that shows that for every real r, the elementary function takes positive value in the real number domain.

$$\frac{d^2P}{dt^2} = r^2 P_0 e^{rt}$$

Also, instead of t= 0, I can take any starting time (t_0) from where the measurement started, where the only change would be in the argument of the function which would be the difference between those two measured times.

3.2. Logistic population model

To improve the previous model, I found that Belgian mathematician Verhulst developed a more precise mathematical model. For example, with the increase of population, there will be an increase in disease, increased competition for the limited resources and limited available space which all serve to slow down the growth rate. Mathematically it means that birth and death rate are functions of time. The assumption we must make here is to ignore the effects of immigration and emigration, meaning we have a closed environment. In a closed environment, the population can reach the maximum number of people, the so-called maximum carrying capacity (Crauder et al., 2008, p. 399). If P=P(t) then, the new model of a rate of change function is written as:

$$\frac{dP}{dt} = rP(1 - \frac{P}{K})$$

Where K is constant, the maximum carrying capacity of the population. This result to me is intuitively pleasing. As the population increases, the ratio $\frac{P}{K}$ is coming closer to 1, hence $\lim_{t\to\infty}(1 - \frac{P}{K}) = 0$. Therefore, I expect asymptotical behaviour of the function $as\ t \to \infty$. To solve this differential equation, I will separate the variables:

$$\frac{dP}{P(1 - \frac{P}{K})} = rdt$$

The left side of equation cannot be integrated directly, so I will decompose it into partial fractions:

$$\frac{1}{P\left(1 - \frac{P}{K}\right)} = \frac{1}{P} + \frac{1}{K - P}$$

Now, I can integrate the both sides of the differential equation:

$$\int_{P_0}^{P} \left(\frac{dP}{P} + \frac{dP}{K - P}\right) = \int_{0}^{t} rdt$$

Evaluating the integrand of both sides yields:

$$[\ln P - \ln(K - P)]_{P_0}^{P} = [rt]_{P_0}^{P}$$

Using the laws of logarithms, we get the following form:

$$\ln\left[\frac{\frac{P}{P_0}}{\frac{K - P}{K - P_0}}\right] = rt \quad \rightarrow \quad \frac{P(K - P_0)}{P_0(K - P)} = e^{rt}$$

Now, the purpose is to find P(t), I will do that by multiplying and rearranging the fraction:

$$P(K - P_0) = e^{rt}P_0(K - P) \quad \rightarrow \quad PK - PP_0 + PP_0e^{rt} = KP_0e^{rt}$$

From the last obtained expression, I will extract P:

$$P(K - P_0 + P_0 e^{rt}) = P_0 K e^{rt}$$

So, P expressed as a function of time is:

$$P(t) = \frac{K P_0 e^{rt}}{K\left(1 - \frac{P_0}{K} + \frac{P_0}{K} e^{rt}\right)}$$

By arranging the fraction, we get a more elegant form:

$$P(t) = \frac{P_0 e^{rt}}{1 + \frac{P_0 (e^{rt} - 1)}{K}}$$

This is a logistic function.

Where,

P(t) -number of population at a time t
P_0 - initial number of population
r- growth rate
K- carrying capacity= maximum amount of people a population can achieve

The numerator of the logistic function contains the same expression as for exponential growth. What is then the difference between those two? The logistic function contains the behaviour of the exponential growth, but, at some point, that growth slows down. To check this, we need to perform the second derivation and find its zero points, in order to detect the inflection point.

$$\frac{dP}{dt} = rP\left(1 - \frac{P}{K}\right) = P'$$

In this step, I marked the first derivative with P'. Using implicit derivation and product rule I get the following:

$$\frac{d^2P}{dt^2} = rP'\left(1 - \frac{P}{K}\right) - r\frac{P}{K}P'$$

Substituting P'into the previous equation and solving it for $\frac{d^2P}{dt^2} = 0$, gives zero points:

$$\frac{d^2P}{dt^2} = r^2 P\left(1 - \frac{P}{K}\right)\left(1 - \frac{2P}{K}\right) = 0$$

$$-\infty \qquad P = 0 \qquad P = \frac{K}{2} \qquad P = K \qquad +\infty$$

Logistic function	$\frac{d^2P}{dt^2}$	+	+	−	−
Exponential function	$\frac{d^2P}{dt^2}$	+	+	+	+

For logistic function, P=0 and P=K are no inflection points. They represent constant solutions of logistic function which have no variation and therefore, no inflection points. Point of inflection occurs when $\frac{d^2P}{dt^2} = \frac{K}{2}$. For logistic function from $P \in \left]0,\ \frac{K}{2}\right[$ function has concave up shape (positive second derivative) and from $P \in \left]\frac{K}{2},\ K\right[$ function has concave down shape (negative second derivative). The exponential function has concave up shape for all $P \in]-\infty, +\infty[$, because its second derivative is always positive. Therefore, the exponential growth model of population can be only applied from $P \in \left]0,\ \frac{K}{2}\right[$ because only then both logistic and exponential functions have concave up shape.

4. Application of models on actual data

To test these two models, I will use actual data of the Indian population from the Office of the Registrar General & Census Commissioner of India from 1901 till 2011 (ORG & CCI, 2020).

4.1. Analysing the data

Table 1

In table 1 below we can see the corresponding number of India's population recorded every 10 years.

<u>Table 1:</u> *Shows Indian population data from 1901-2011*

Year	1901	1911	1921	1931	1941	1951	1961	1971	1981	1991	2001	2011
Population (in millions)	238.396	252.093	251.321	278.977	318.661	361.088	439.235	548.160	683.329	846.421	1028.737	1210.193

Source: (ORG & CCI, 2020 p. 41)

4.2. Fitting data using rates of change

Using literature (Crauder et al., 2008, p. 406) I learned how to find the growth rate and maximum carrying capacity. I will model the logistic function using the slope of the curve of graph 1. I will do this by transferring the factor P in the derivative formula of the logistic function. In that way, I will have the linear formula with P as the independent variable:

$$\frac{\frac{dP}{dt}}{P} = r\left(1 - \frac{P}{K}\right)$$

I will calculate the slope (derivative) for a specific time t intervals using the formula for slope:

$$slope\ = \frac{P(t) - P(t-10)}{10}.$$

Table 2

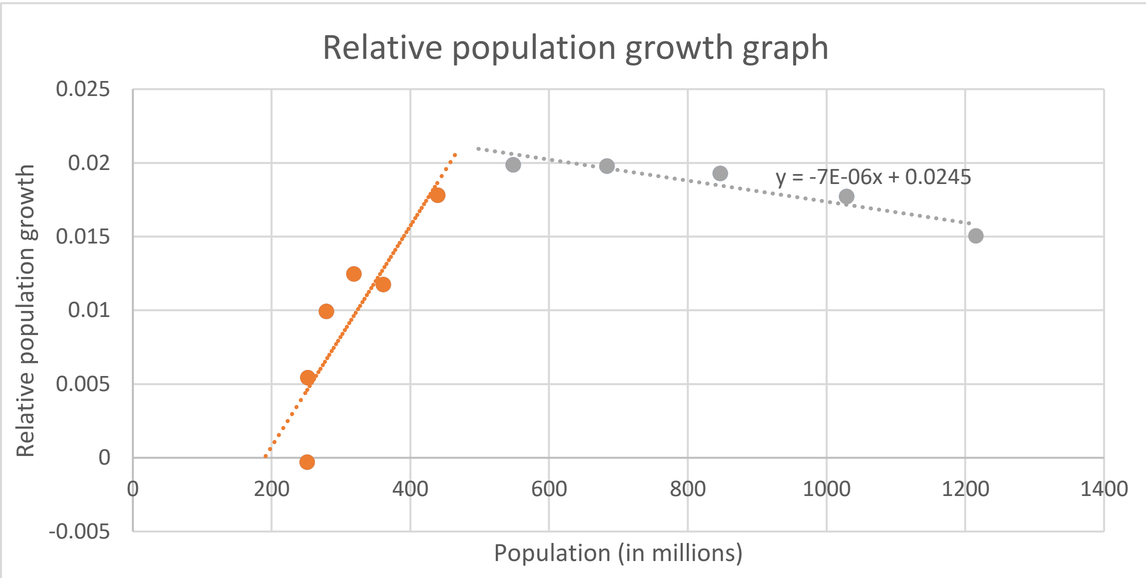
In table 2 below I have calculated the values for slope at given time intervals. All slopes are rounded to three significant digits except the last two because they differ in decimal points and therefore I rounded them to four.

Time interval	1901-1911	1911-1921	1921-1931	1931-1941	1941-1951	1951-1961	1961-1971	1971-1981	1981-1991	1991-2001	2001-2011
Slope $\frac{dP}{dt} \cdot 10^6$	1.40	-0.777	2.77	3.97	4.24	0.781	10.9	13.5	16.3	18.23	18.21

Table 2: *Shows the values of slopes*

Since $\frac{\frac{dP}{dt}}{P}$ is relative growth of population that I will replace it with y because it represents dependent variable. Also, I will multiply r with the members within the parentheses to get a linear equation: $y = r - \frac{rP}{K}$. Let m be the slope $\frac{r}{K}$, then $y = -mP + r$.

Graph 2:



Graph 2: *Shows the relationship between relative growth and population*

In graph 2 above, on the x-axis is the population (in millions) and on the y-axis is the relative population growth. For me, this graph presents quite unexpected results. Since theory showed me that it should be a linear graph with a negative slope, I expected it to be such through the whole domain. Now I need to see what to do with the data from 1911 to 1961 that can be approximated by a linear function with a positive slope.

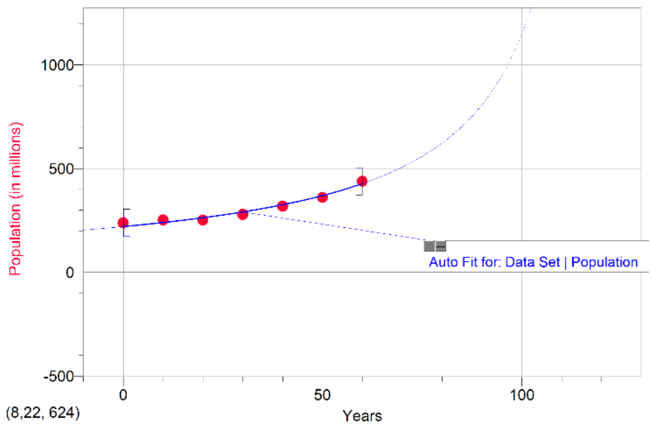
On one hand, a logistic curve can be applied to data from 1971 to 2001 because the slope of the linear graph corresponds to the theory. On the other hand, the data from 1911 to 1961 cannot. Also, an exponential function cannot be applied in this period since the graph of relative growth doesn't show a constant with a value of growth rate r, since $\frac{\dot{P}}{P} = r$ for exponential growth. Although it is not a logistic or exponential function, there is a possibility that it is maybe a rational function.

4.3. Making a model of a rational function

I will get rational function by the appropriate fit for $t \in [1901, 1961]$, in Logger Pro program. Afterwards the program will estimate the most precise values of coefficients for my function. The rational function of the best fit given is:

$$y = \frac{9559}{-0.350t + 43.38}$$

Graph 3



Autofit means that the Logger Pro made an automatic approximation of the best fit curve. Since the denominator cannot be equal to zero, $-0.350t + 43.38 \neq 0$, the biggest period India can endure from population growth is 124 years or till year 2025, which is also the value of vertical asymptote.

4.4. Making a model of a logistic function

The equation that approximates my relative growth data from t$\in [1971, 2011]$ given by excel is:

$$y = -7 \cdot 10^{-6}x + 0.0245$$

This equation is equivalent with the: $y = r - \frac{rP}{K}$. By comparing two equations it can be seen that r is y intercept (it isn't multiplied with the argument P) so it equals to 0.0245. Since P is the argument, independent variable x, the ratio $\frac{r}{K}$ is equal to $7 \cdot 10^{-6}$.

Table 3

Growth rate r= Y intercept	Slope=m= r/K	K=-r/m	$\frac{K}{2}$
0.0245	$-7 \cdot 10^{-6}$	3500	1750

In my calculations, it is estimated that the maximum population, maximum carrying capacity India can achieve is about 3,5 billion people. The half-carrying capacity of India's population is 1.75 billion (India's population point of inflection). From the census, the last data of India's population is 1210.860. The ratio of last recorded population and carrying capacity K is $\frac{P}{K} = \frac{1210.193}{3500} = 0.346$

Given that the ratio of numbers is less than 0.5, this means that the function is just entering the inflection region.

4.5. Combining two models

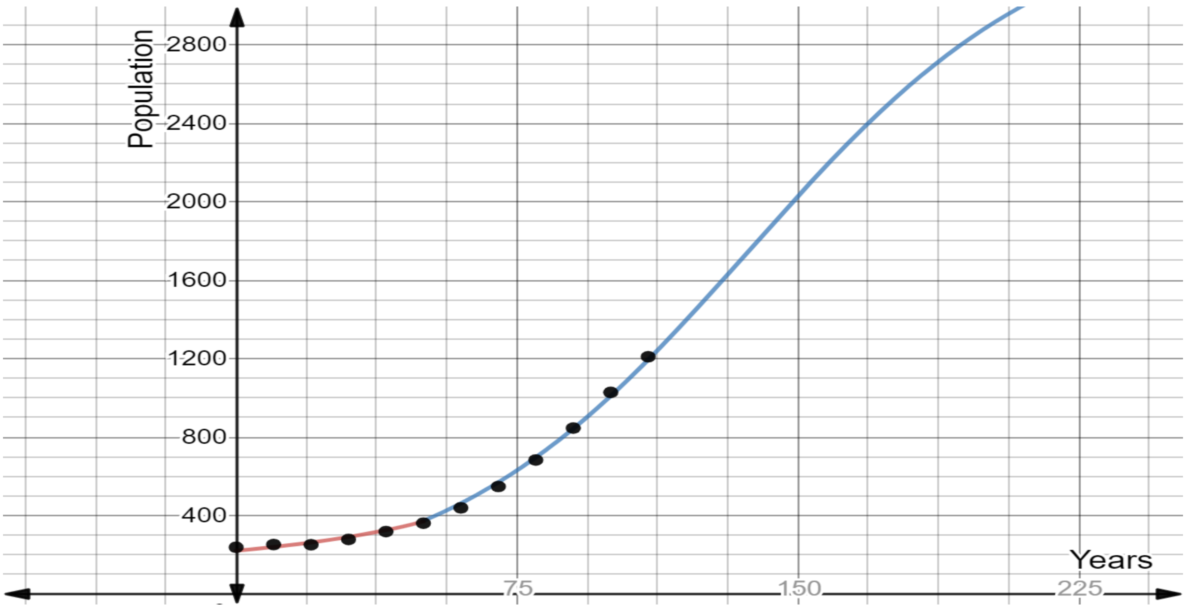
For models to be applicable India must be modelled using two piecewise functions (piecewise functions are used when we cannot describe something by a single equation, instead we describe it using a combination of functions which are defined for the given interval). The functions that approximate the Indian population in a given time interval are:

$$P(t) = \begin{cases} \frac{9559}{-0.350t + 43.38}, & 1901 \leq t \leq 1961 \\ \frac{238.396e^{0.0245t}}{1 + 0.0681(e^{0.0245t} - 1)}, & 1961 < t \leq 2011 \end{cases}$$

I rounded up to three significant figures because it has an accuracy of up to five thousandths, so when dealing with large numbers, it doesn't make a difference.

Graph 4:

In graph 4 below, the x-axis represents the time in years where 1901 is taken as time origin t=0 and y-axis represent the population in millions. The graph was plot in Desmos online program. Black dots represent census data. Red line (rational function) is function defined for $1901 + t,\ t \in [0,60]$ and the blue one (logistic function) for $1901 + t, t \in]60, 110]$.



Graph 4: *Shows Indian population and applied piecewise functions*

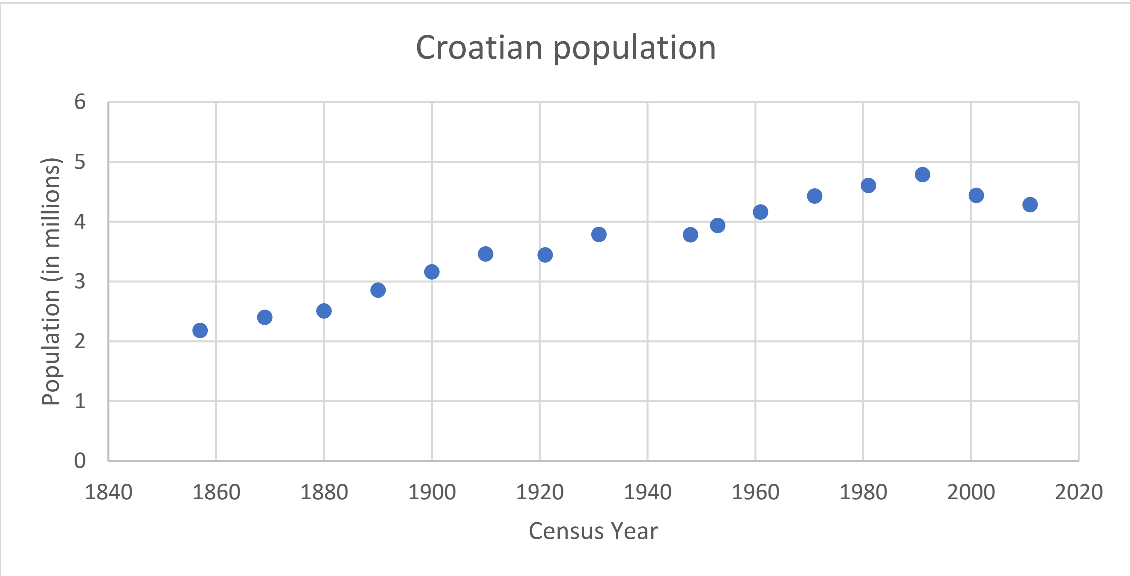
Table 4 (see appendix) shows the results of an exponential and logistic model population number in millions. In the last two columns, I calculated the relative error, which is the difference between the results of the model and the real data divided by the value of real data. Even though functions were created at defined intervals, I made interpolation to see how well they can predict data (the process of approximating data within the interval at which the function is obtained).

In my judgment, in India before 1941 as an English colony, there was an oppression system and If we try to explain it socio-economically, then the colonial influence is noticeable until 1961, and then the population dynamics change in line with the new socio-economic situation.

5. Investigating the Croatian data

Graph 5:

In graph 5 below, the x-axis represents census years and y-axis corresponding population in millions.

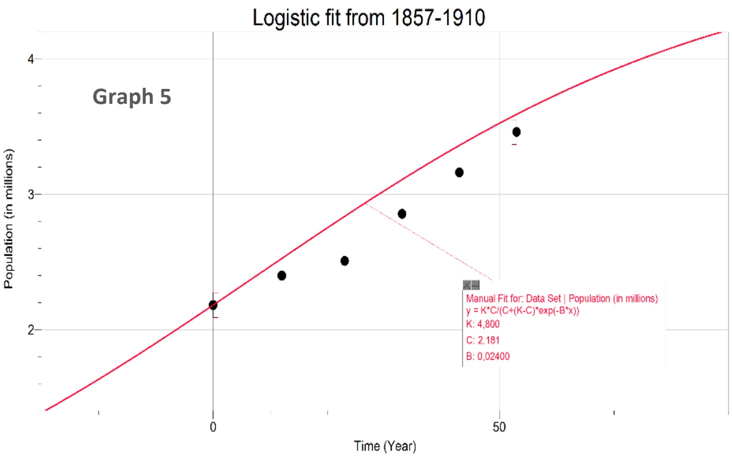


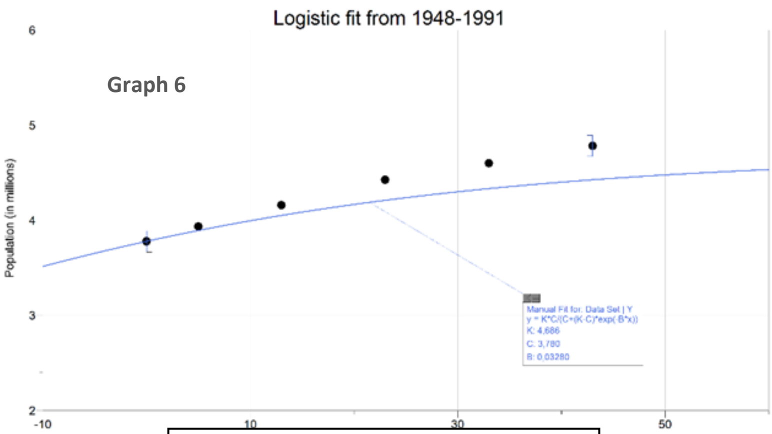
Graph 4: *Shows Croatian population data from 1857-2011*

Let us now examine the graph. The first thing I noticed was that the curve is not continuously increasing. There are periods with stagnation in the population number and even decreasing. This can't be explained by exponential growth nor with logistic growth. The curve is interrupted in three war periods. From the first World war beginning in 1914, the Second world war beginning in 1939 and from 1991 when the Homeland War began. These three intervals can't be explained by my equations. The interruption can be explained by the war and the decrease in population due to war conditions. On the other hand, exponential and logistic equations are based on the assumptions of natural population increase and natural dying resulting in a stable growth rate. In Croatia, there was an unnatural death and population decline as a result of a violent death due to war. As social conditions have changed since World War II, Croatia has achieved population growth, but with a different initial population and with a different growth rate. The logistic model can be applied but only as a piecewise function.

For modelling, I will only use points where the population grows continuously in order to achieve an appropriate logistic model. Points with a decrease or stagnation in population will not be included in developing appropriate piecewise functions. I will try to explain them by functions discontinuity (the difference between the predicted value of functions which are lacking in relation to natural growth) got by extrapolation. The logistic model and approach I learned from the Indian population do not make equally good predictions for the Croatian population as seen from graphs 6 and 7.

Graph 6,7: *Show applied logistic fit from 1857-1910 and 1948-1991*





$$P(t) = \frac{2.181e^{0.024t}}{1 + \frac{2.181}{4.80}(e^{0.024t} - 1)}$$

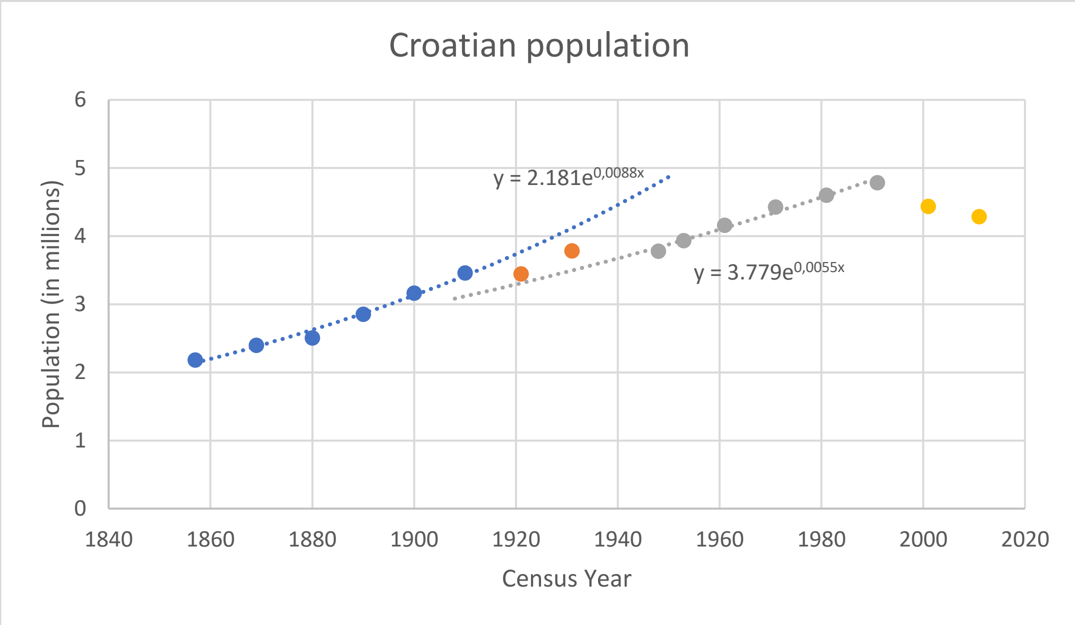
$$P(t) = \frac{3.780e^{0.032t}}{1 + \frac{3.780}{4.686}(e^{0.032t} - 1)}$$

At this point, I don't know how I would explain this, but I think the exponential model would fit better. My premise as to why the logistics model cannot be fully applied is well due to another assumption I later learned. The assumption in the logistic function predicts that with a smaller population, individuals will progress much better (Zill, 2013). This assumption is valid for the fish population (there are no larger fish that would eat smaller ones), while for the human population that may not be the case. As the war has reduced population and caused great harm to people, we cannot say with certainty that those fewer people, with their welfare overall being decreased, will contribute equally to population growth.

I will try to approximate with the exponential function:

Graph 8:

In graph below, the x-axis represents census year and y-axis population in millions.



Graph 8: *Shows Croatian population data and applied piecewise exponential functions*

The equations for piecewise function are:

$$P(t) = \begin{cases} 2.181e^{0.0088t}, & 1857 \le t \le 1910 \\ 3.779e^{0.0055t}, & 1948 \le t \le 1991 \end{cases}$$

Where function 1 is, $2.181e^{0.0088t}$ modelled for the interval from the year 1857 till 1910 and function 2, $3.779e^{0.0055t}$ modelled for the interval from the year 1948 till 1991 included.

The discontinuities of the piecewise curve can predict the order of the magnitude of victims that occurred. If I bring the curves into the realm of discontinuity, $t \in]1921, 1948[$ then I can see how many people are missing. During this period, the assumptions I made at the beginning were not valid. There is no natural death, the number of dead people, as a result of violent deaths, increases and the number of new-borns decreases. If I go into the period of discontinuity, do the interpolation and calculate the population difference predicted by the function 1 and 2, I get the following shocking results of missing people:

Year	Population difference (Function 1 – Function 2)
1921	572 957
1931	683 873

In class I learned that an exponential function can be written as the sum of a potentials

$$e^{rt} = 1 + \frac{rt}{1!} + \frac{(rt)^2}{2!} + \frac{(rt)^3}{3!} + \cdots$$

When $rt \ll 1$ (significantly smaller), the sum of the potentials become linear because the members in parentheses become negligible. Then the exponential function can be approximated by the linear function:

$$e^{rt} = 1 + rt$$

Given that the r I calculated is of the order 10^{-2}, it would be perfectly justified to apply a linear fit in given periods and still get good results. In both cases my function would have the form of:

$P_0 e^{rt} \approx y = P_0 + P_0 rt.$

6. Evaluation and conclusion

6.1. Evaluation

In the Indian population, I learned to apply the logistic model in the way I read in the literature. I encountered an unexpected result in a relative growth rate graph that was not predicted by theory. I didn't solve that problem by deduction but by guessing. Luckily the rational model I found was proven to estimate precise results. In the vicinity of vertical asymptote of a rational function, the population tends to infinity which proves that the adopted rational function is best-used piece-wisely but can't be used for extrapolation, for physical reasons. My theory has the assumption that the growth rate must be always stable. The Croatian population data cannot be explained with a unique stable growth rate. It was necessary to apply the piecewise explanation. There are two different growth rates and the corresponding initial population in the case of Croatia. Different growth rates indicate different social circumstances in different historical periods. The social meltdown occurred in World War II. Extrapolation of the piecewise functions to the period of World War II produced a discontinuity in the population prediction. The order of magnitude of the demographic losses shown in historical data agrees with the discontinuity predicted by these functions (Žerjavić, 1995, p. 548).

6.2. Conclusion

Exponential, logistic and rational functions used and developed in this essay are capable of explaining the population dynamic of various countries. Functions used emerged from differential equations gotten from first principle (as a result of hypothetical assumptions) or by searching for the best possible fit to the data given. Different models are used in different historical periods to explain changes in population dynamics in accordance with changes in socio-economic conditions. Models are shown to fit the population data of India and Croatia in the 20th century.

Appendix

Table 4 – Indian population

Time	Year	Census data	Logistic model	Logistic relative error	Rational model	Rational relative error
0	1901	238,396	117,413	-0,507	220,355	-0,0757
10	1911	252,093	148,911	-0,409	239,712	-0,0491
20	1921	251,321	188,489	-0,250	262,798	0,0457
30	1931	278,977	238,000	-0,147	290,803	0,0424
40	1941	318,661	299,595	-0,0598	325,490	0,0214
50	1951	361,088	375,700	0,0405	369,573	0,0235
60	1961	439,235	468,936	0,0676	427,466	-0,0268
70	1971	548,160	581,981	0,0617	506,867	-0,0753
80	1981	683,329	717,330	0,0498	622,493	-0,0890
90	1991	846,421	876,963	0,0361	806,462	-0,0472
100	2001	1028,737	1061,933	0,0323	1144,790	0,113
110	2011	1210,193	1271,912	0,0504	1972,148	0,629

Table 5 – Croatian population

Year	Real data	Function 1	Function 2	Relative error (function 1)	Relative error (function 2)
1857	2.181	2.181	2.290	0	0,0502
1869	2.398	2.423	2.447	0,0107	0,0204
1880	2.506	2.670	2.599	0,065	0,0374
1890	2.854	2.915	2.746	0,021	-0,0377
1900	3.161	3.184	2.902	0,00718	-0,0820
1910	3.460	3.477	3.066	0,00476	-0,114
1921	3.443	3.830	3.257	0,112	-0,0540
1931	3.785	4.182	3.498	0,105	-0,0756
1948	3.779	4.857	3.779	0,285	0
1953	3.936	5.076	3.884	0,290	-0,0131
1961	4.159	5.446	4.059	0,309	-0,0242
1971	4.426	5.947	4.288	0,344	-0,0312
1981	4.601	6.494	4.531	0,411	-0,0153
1991	4.784	7.092	4.787	0,482	0,000633
2001	4.437	7.744	5.057	0,745	0,140
2011	4.284	8.531	5.343	0,991	0,247

Bibliography

1. Crauder. B., Evans, B. and Noell, A. (2008) Functions and Change: A Modeling Approach to College Algebra, 4th, Houghton Mifflin, ISBN-13:978-0-547-16567-7

2. Croatian Bureau of Statistics (CBS), (2020): Census of population, households and dwellings 1857-2001: Population by sex and age, by settlements, www.dzs.hr (15. 01. 2020.).

3. Office of the Registrar General & Census Commissioner, India (ORG & CCI) (2020), Size, growth rate and distribution of population, http://censusindia.gov.in/2011-prov-results/data_files/india/Final_PPT_2011_chapter3.pdf (10. 01. 2020.).

4. Zill, D. G. (2013) A First Course in Differential Equations with Modeling Applications, 10th ed. Brooks/Cole, Cengage Learning, ISBN-13: 978-1-111-82705-2, http://lya.fciencias.unam.mx/jele/EDOs2k15.2/Libros/A%20First%20Course%20in%20Differential%20Equations%20with%20Modeling%20Applications%2010e%202012%20Zill.pdf (15.12.2019.)

5. Žerjavić, V. (1995) Demographic and War Losses of Croatian The World War Two and in the Postwar Period, Časopis za suvremenu povijest, Vol. 27 No. 3, https://hrcak.srce.hr/213638 (30.01.2020.).