Forecasting Population

Definitions

- **Webster's definitions:**
  - **Projection**
    - an estimate of future possibilities based on a current trend
  - **Estimate**
    - a rough or approximate calculation; a numerical value obtained from a statistical sample and assigned to a population parameter
  - **Forecast**
    - to calculate or predict (some future event or condition) usually as a result of study and analysis of available pertinent data

Forecasting: Need

- **Short-term**
  - How much population would be available for next growth rate (interest)?
  - Working on proposals without knowing what to expect could result in a waste of time
- **Long-term**
  - Planning for long-lasting population assets
  - Borrowing is usually a long-term commitment
  - Growth rate accrue in the future

Forecasting Considerations

- **Forecaster must understand the local government’s population growth system**
  - How is it administered
  - Tax base structure
  - Other population sources
- **Understand the factors that have affected past population growth**
  - Structure/definition
  - Administration
    - Are all people posted?
    - All people growth at the same way?
- **Must have adequate and timely data**
    - No data is better than false data
- **Use graphs of variables against time to visualize changes**
  - Are changes small
  - Are changes large
  - Are changes seasonal
  - Are any patterns evident and are these congruent with regional or national population grpwth pattern?

- **Forecasting process should be transparent**
  - Identify assumptions
  - Define assumptions
  - Avoids under or over-forecasting
- **Individual population sources need to be forecast separately**
  - Different population sources respond to different economic and policy factors
- **Monitor and revise forecasts**
  - Review initial forecast to determine source of error and enhance forecasting for future years

Forecasting Methods

1. **Simplistic**
   - Trend extrapolation or projection using historical data
   - Most common local government population estimation tool
2. **Multiple regression**
   - Use of IVs to predict populations
3. **Econometric**
   - Complex multivariate technique using composite measures to estimate populations
4. **Microsimulation**
   - Estimates based on sample of relevant data

Simplistic Models

- **Assumption**
  - Past trends will continue
  - No major legislative or tax change expected
- **Future population**
  - Extrapolated from historical data or previous forecasting
    - Constant increments
      - population increased by 5000 person for the past 5 years…
    - Constant percentage change
      - population increased by 5% for the past 5 years
    - Simple average compounded growth r = (Y / X)^1/n - 1
    - Linear ($R = a + b_t$) time trends
    - Nonlinear ($\ln R = a + b_t$)
    -

Decomposition to Time Series

- **Breaks the time series into trend, cycle, seasonal (for monthly or quarterly forecasts), and irregular (or residual) components**

  - The method adjust for four basic elements that contribute to the behavior of a series over time
    - S = seasonal factor
      - Regular fluctuations; driven by weather and propriety
    - T = the adjustment for trend
      - Long-run pattern of growth or decline
    - C = cycle
      - Periodic fluctuations around the trend level
    - I = the irregular or residual influence
      - Erratic change that follows no pattern

Decomposition Model

- **$R_t = (S_t)(T_t)(C_t)(I_t)$**
  - Sequence for each of filters is as shown in the equation
  - R = population to be forecast
  - S = seasonal
    - extracted by using a centered seasonal moving average
  - T = trend
    - Adjusted by linear regression against time of the seasonally adjusted data
  - C = cycle
    - Identified by removing the trend from the deseosonalized data
  - I = the irregular or residual influence
    - Isolated by removing the cyclical component from the series
  - t = time of the data (historic or forecast)

Multiple Regression

- **Estimates population as a function of one or more IVs**
  - Each equation used to estimate a population source is independent of the others
  - Estimates for the independent variables are generated independent of the regression equation
  - The equation with the best "goodness of fit" is selected

The Regression Model

- **The mathematical equation for a straight line is used to predict the value of the dependent variable (Y) on the basis of the independent variable (X):**

$$Y = a + b_1X_1 + b_2X_2 + b_iX_i + e$$

a is called the Y-intercept. It is the expected value of Y when X=0. This is the base-line amount because it is what Y should be before we take the level of X into account.

b is called the slope (or regression coefficient) for X. This represents the amount that Y changes for each change of one unit in X

e is called the error term or disturbance term. The difference between actual and predicted values.

Econometric Models

- **Uses a system of simultaneously interdependent equations to predict population**
  - The equations are linked by theoretical and empirical relationships
  - These models while preferred by economist because of their theoretical soundness, are in practice not much more accurate than multiple regression models
  - Better in predicting macroeconomic variables

Microsimulation Models

- **A statistical sample of tax data is used to forecast population from a tax source**
  - How the sample is drawn and its updating is critical
  - Economic activity expected in the budget year is included in the analysis
  - More applicable to estimate how population would be affected by proposed policy changes
  - Also useful for regular forecasting

Factors Influencing the Choice of Forecasting Method

Plausibility
"Do the Outputs Make Sense?"

Face Validity
--Availability of Data
--Quality of Data
"Are the Inputs Good?"

Political Acceptability
"Are the Outputs Acceptable?"

Resources
--Money
--Personnel
--Time
"Can we afford it?"

Needs of the Users
--Geographic Detail
--Demographic Detail
--Temporal Detail
"Are User Needs Satisfied?"

Model Complexity
--Ease of Application
--Ease of Explanation
"Can we do this?"
"Can we explain what we did?"

Forecast Accuracy
"Is the Forecast Accurate?"

Simplistic Models

1- **Arithmetic increase method :**

In this method, the rate of growth of population is assumed to be constant. This method gives too low an estimate, and can be adopted for forecasting populations of large cities which have achieved saturation conditions.

Validity: The method valid only if approximately equal incremental increases have occurred between recent censuses.

$$\frac{dp}{dt} = k$$

$$\int_{po}^{pt} dp = \int_{to}^{t} k\,dt$$

$$p_t = p_o + k\Delta t$$

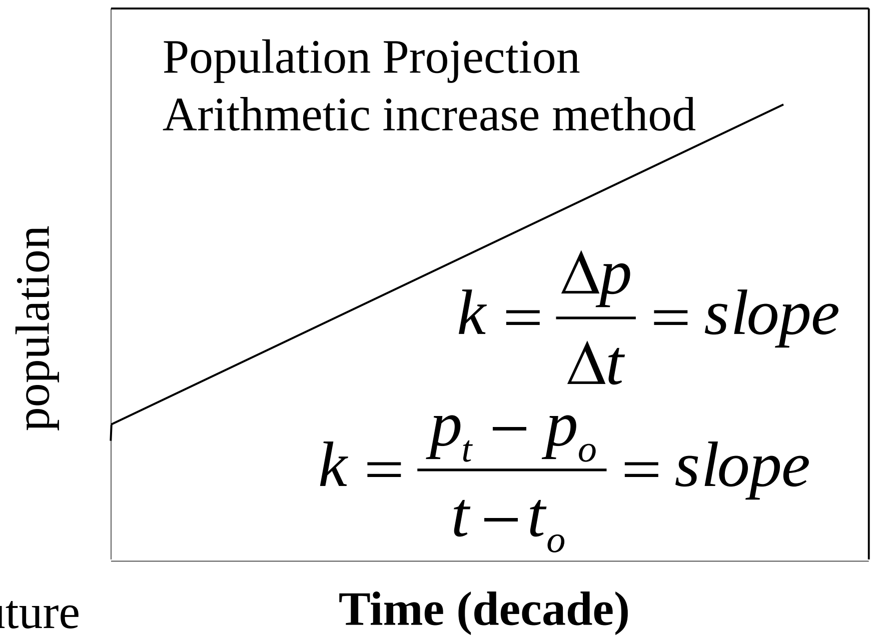
dp/dt : rate of change of population

P_t : population at some time in the future

p_o: present or initial population

t : period of the projection in decades

k : population growth rate (constant)



Note: decade = 10 years

2- **<u>Uniform percentage of increase:</u>**

Assumption: This method assumes uniform rate of increase, that is the rate of increase is proportional to population).

$$\frac{dp}{dt} = k_1 \, p$$

$$\int_{po}^{pt} \frac{dp}{p} = \int_{0}^{t} k_1 \, dt$$

$$\ln p_t = \ln p_o + k_1 \, \Delta t$$

$$\ln p_t = \ln p_o + k_1 (t - t_o)$$

dp/dt : rate of change of population
P_t : population at some time in the future
P_o : present or initial population
Δt : period of the projection in years
k : population growth rate
n : number of years

3- **Logistic method : (Saturation method)**

This method has an S-shape combining a geometric rate of growth at low population with a declining growth rate as the city approaches some limiting population.

A logistic projection can be based on the equation:

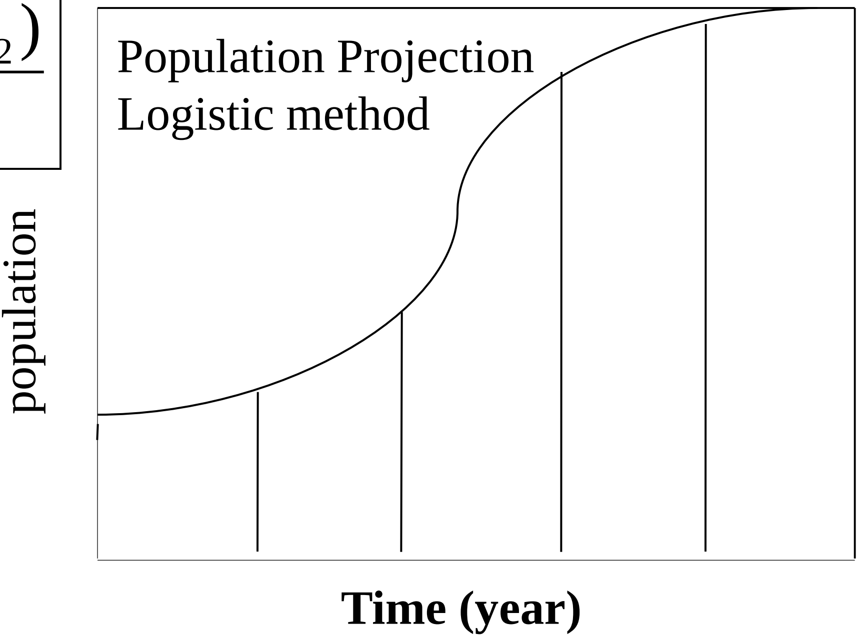
p_t : population at some time in the future
p_o: base population
p_{sat}: population at saturation level
p_1 , p_2 : population at two time periods
n : time interval between succeeding censuses
Δt : no. of years after base year

$$p_t = \frac{p_{sat}}{1 + e^{a + b\Delta t}}$$

$$|p_{sat}| = \frac{2 p_o p_1 p_2 - p_1^2 (p_o + p_2)}{p_o p_2 - p_1^2}$$

$$a = \ln \frac{p_{sat} - p_2}{p_2}$$

$$b = \frac{1}{n} \ln \frac{p_o (p_{sat} - p_1)}{p_1 (p_{sat} - p_o)}$$



4- **Declining growth method :**

This technique, like the logistic method, assumes that the city has some limiting saturation population, and that its rate of growth is a function of its population deficit:

$$\frac{dp}{dt} = k_2(p_{sat} - p)$$

p_t : population at some time in the future
p_o: base population
p_{sat}: population at saturation level
p , p_o : are populations recorded n years apart
Δt : no. of years after base year

k_2 may be determined from successive censuses and the equation:

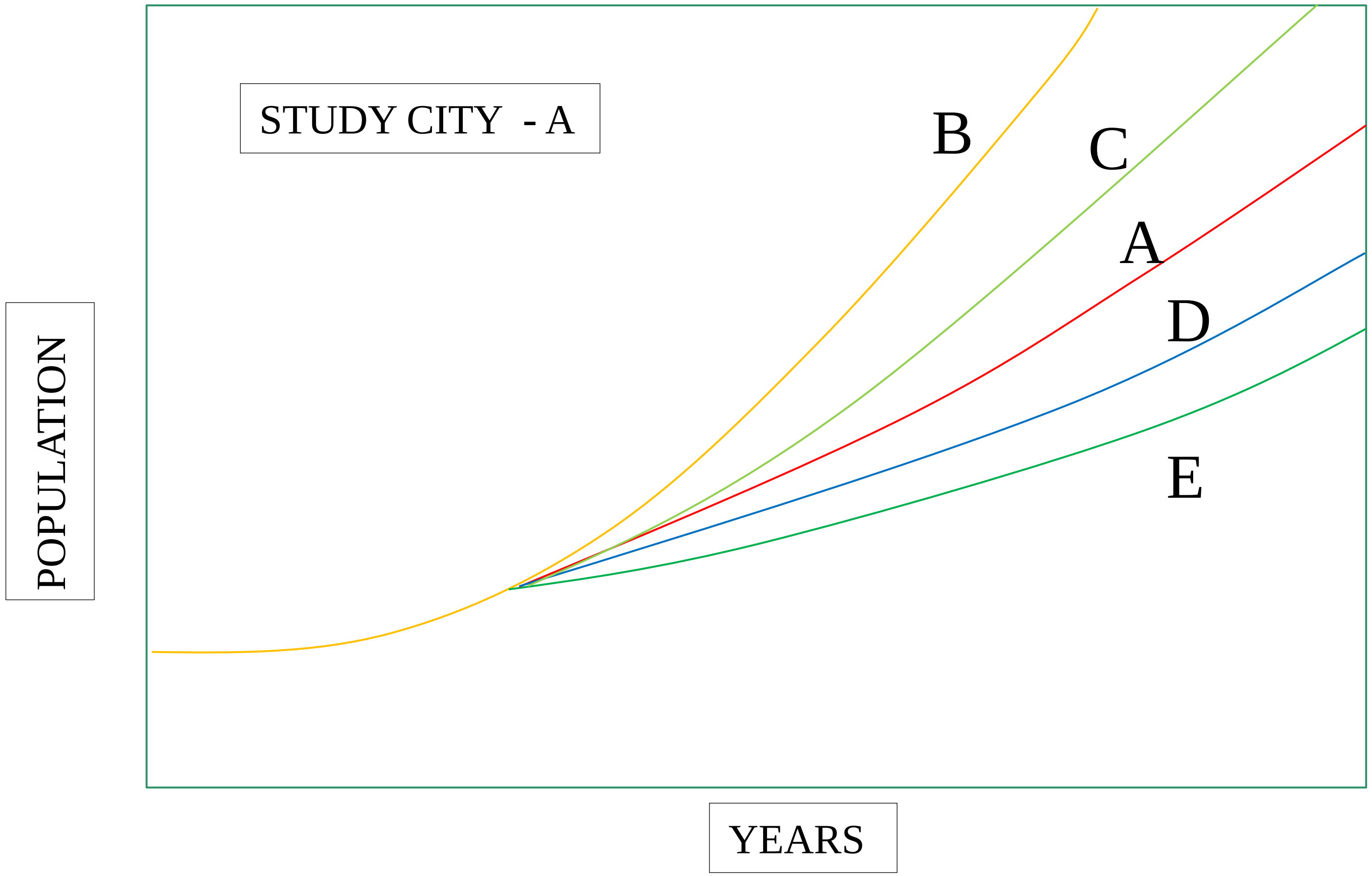
$$k_2 = -\frac{1}{n}\ln\frac{p_{sat} - p}{p_{sat} - p_o}$$

then,

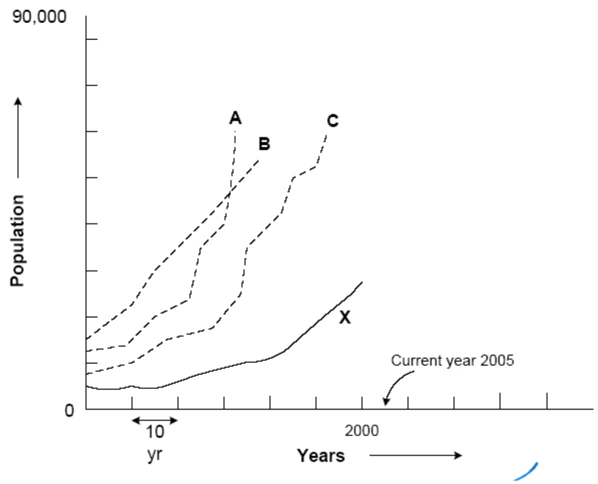
$$p_t = p_o + (p_{sat} - p_o)(1 - e^{k_2 \Delta t})$$

5- **Curvilinear method (Comparative graphical extention method) :**

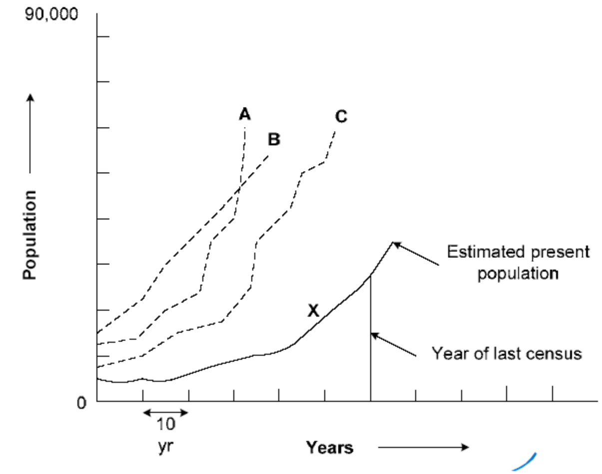
This technique, involves the graphical projection of the past population growth curve, continuing whatever trends the historical data indicate. This method includes comparison of the projected growth to the recorded growth of other cities of larger size. The cities chosen for the comparison should be as similar as possible to the city being studied.:



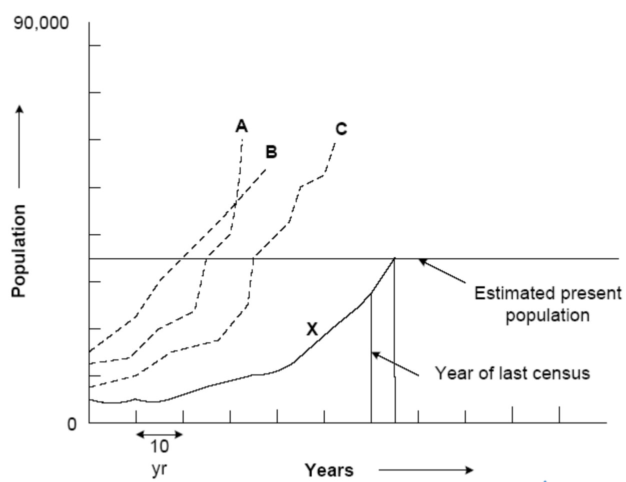
Curvilinear method



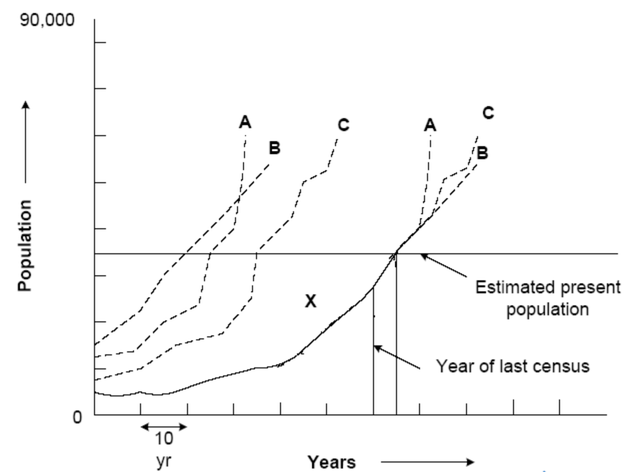
Curvilinear method



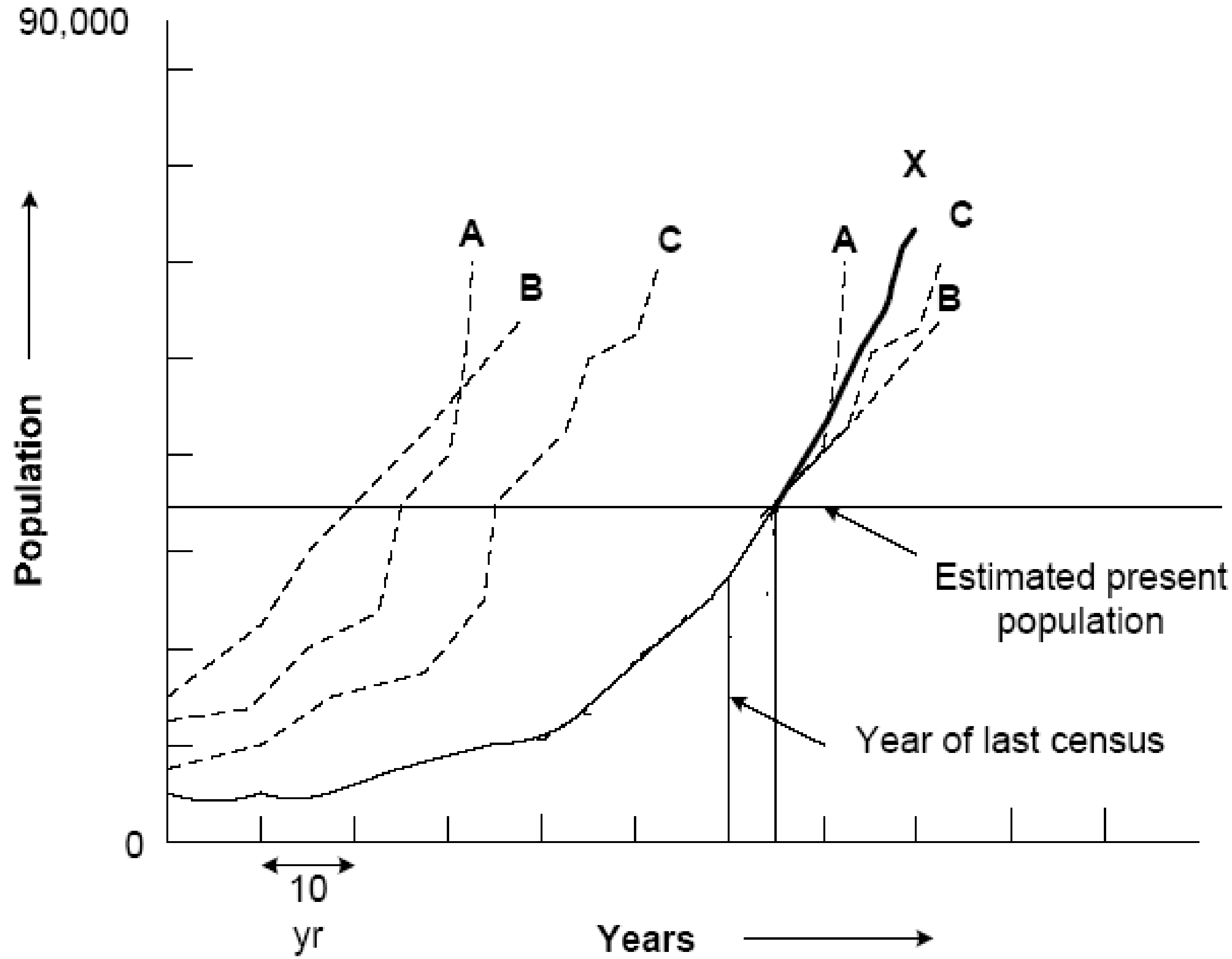
Curvilinear method



Curvilinear method



Curvilinear method



6- **Incremental increase method : (Method of varying increment)**

In this technique, the average of the increase in the population is taken as per arithmetic method and to this, is added the average of the net incremental increase, one for every future decade whose population figure is to be estimated. In this method, a progressive increasing or decreasing rate rather than constant rate is adopted. Mathematically the hypothesis may be expressed as:

$$p_t = p_o + n.k + \frac{n(n+1)}{2}.a$$

p_t : population at some time in the future

p_o: present or initial population

k : rate of increase for each decade

a : rate of change in increase for each decade

n : period of projection in decades

7- **Geometric increase method :**

This method assumes that the percentage of increase in population from decade to decade is constant. This method gives high results, as the percentage increase gradually drops when the growth of the cities reach the saturation point. This method is useful for cities which have unlimited scope for expansion and where a constant rate of growth is anticipated. The formula of this estimation is :

$$p_t = p_o(1+k)^n$$

p_t : population at some time in the future
p_o: present or initial population
k : average percentage increase (geometric mean)
n : period of projection in decade

Example :

The population of a town as per the senses records are given below for the years 1945 to 2005. Assuming that the scheme of water supply will commence to function from 2010, it is required to estimate the population after 30 years, i.e. in 2040 and also, the intermediate population i.e. 15 years after 2010.

Year	Population
1945	40185
1955	44522
1965	60395
1975	75614
1985	98886
1995	124230
2005	158790

Solution :

1- **Arithmetic increase method:** Increase in population from 1945 to 2005 , i.e. for 6 decades: 158800 – 40185 = 118615 = total increment

Increase per decade = 118615 / no. of decade = 118615 / 6 = 19769

$$p_t = p_o + k\Delta t$$

$$p_{2025} = p_{2005} + (19769)(2)$$
$$= 158800 + (19769)(2)$$
$$= 198338, capita$$

$$p_{2040} = p_{2005} + (19769)(3.5)$$
$$= 158800 + (19769)(3.5)$$
$$= 227992, capita$$

Year	Population	Increase
1945	40185	------
1955	44522	44522 – 40185 = 4337
1965	60395	15873
1975	75614	15219
1985	98886	23272
1995	124230	25344
2005	158800	34570
Total		118615
Average		118615/6=19769

2- **Geometric increase method :**

$$p_t = p_o (1+k)^n$$

Year	Population	Increase	Rate of growth
1945	40185	------	
1955	44522	44522 – 40185 = 4337	4337 / 40185 = 0.108
1965	60395	15873	0.356
1975	75614	15219	0.252
1985	98886	23272	0.308
1995	124230	25344	0.256
2005	158800	34570	0.278

$$k = \sqrt[6]{0.108x0.356x0.252x0.308x0.256x0.278} = 0.2442$$

$$p_{2025} = p_{2005}(1+0.2442)^2 = 245828, capita$$

$$p_{2040} = p_{2005}(1+0.2442)^{3.5} = 341166, capita$$

3- **Incremental increase method : (Method of varying increment) :**

$$p_t = p_o + n.k + \frac{n(n+1)}{2}.a$$

Year	Population	Increase (k)	Incremental increase (a)
1945	40185	------	
1955	44522	44522 – 40185 = 4337	
1965	60395	15873	+11536
1975	75614	15219	- 654
1985	98886	23272	+8053
1995	124230	25344	+2072
2005	158800	34570	+9226
Total		118615	30233
Average		19769	6047

$$p_{2015} = p_{1995} + n.k + \frac{n(n+1)}{2}.a = 158800 + 2x19769 + \frac{2x3}{2}6047 = 216479, capita$$

$$p_{2040} = p_{1995} + n.k + \frac{n(n+1)}{2}.a = 158800 + 3.5x19769 + \frac{3.5x4.5}{2}6047 = 275612, capita$$

4- **Uniform percentage of increase:**

$$\ln p_t = \ln p_o + k_1 (t - t_o)$$

$$k_1 = \frac{\ln p_t - \ln p_o}{\Delta t}$$

$$\ln p_{2025} = \ln p_{2005} + k_1 \Delta t$$

$$\ln p_{2025} = \ln 158800 + 0.024 x 20 = 12.455$$

$$p_{2025} = e^{12.455} = 256530, capita$$

$$\ln p_{2040} = \ln p_{2005} + k_1 \Delta t$$

$$\ln p_{2040} = \ln 158800 + 0.024 x 35 = 12.815$$

$$p_{2040} = e^{12.815} = 367692, capita$$

Year	Population	Growth rate (k_1)
1945	40185	------
1955	44522	0.01
1965	60395	0.03
1975	75614	0.022
1985	98886	0.027
1995	124230	0.029
2005	158800	0.025
Total		0.143
Average		0.024

5- **Logistic method:**

$$p_t = \frac{p_{sat}}{1+e^{a+b\Delta t}}$$

Year	Population	
1945	40185	$\leftarrow p_o$ ($\rightarrow t_o$)
1955	44522	
1965	60395	
1975	75614	$\leftarrow p_1$ ($\rightarrow t_1$)
1985	98886	
1995	124230	
2005	158800	$\leftarrow p_2$ ($\rightarrow t_2$)

$n = 30$ (from t_o to t_1), $n = 30$ (from t_1 to t_2)

$$\left|p_{sat}\right| = \frac{2p_o p_1 p_2 - p_1^2(p_o + p_2)}{p_o p_2 - p_1^2} = \frac{2(40185)(75614)(158800) - (75614)^2(40185+158800)}{(40185)(158800) - (75614)^2} = 260053$$

$$a = \ln\frac{p_{sat} - p_2}{p_2} = \ln\frac{260053 - 158800}{158800} = -0.45$$

$$b = \frac{1}{n}\ln\frac{p_o(p_{sat} - p_1)}{p_1(p_{sat} - p_o)} = \frac{1}{30}\ln\frac{40185(260053 - 75614)}{75614(260053 - 40185)} = -0.027$$

$$p_{2025} = \frac{260053}{1+e^{-0.45+(-0.027)x20}} = 189602, capita$$

$$p_{2040} = \frac{260053}{1+e^{-0.45+(-0.027)x35}} = 208404, capita$$

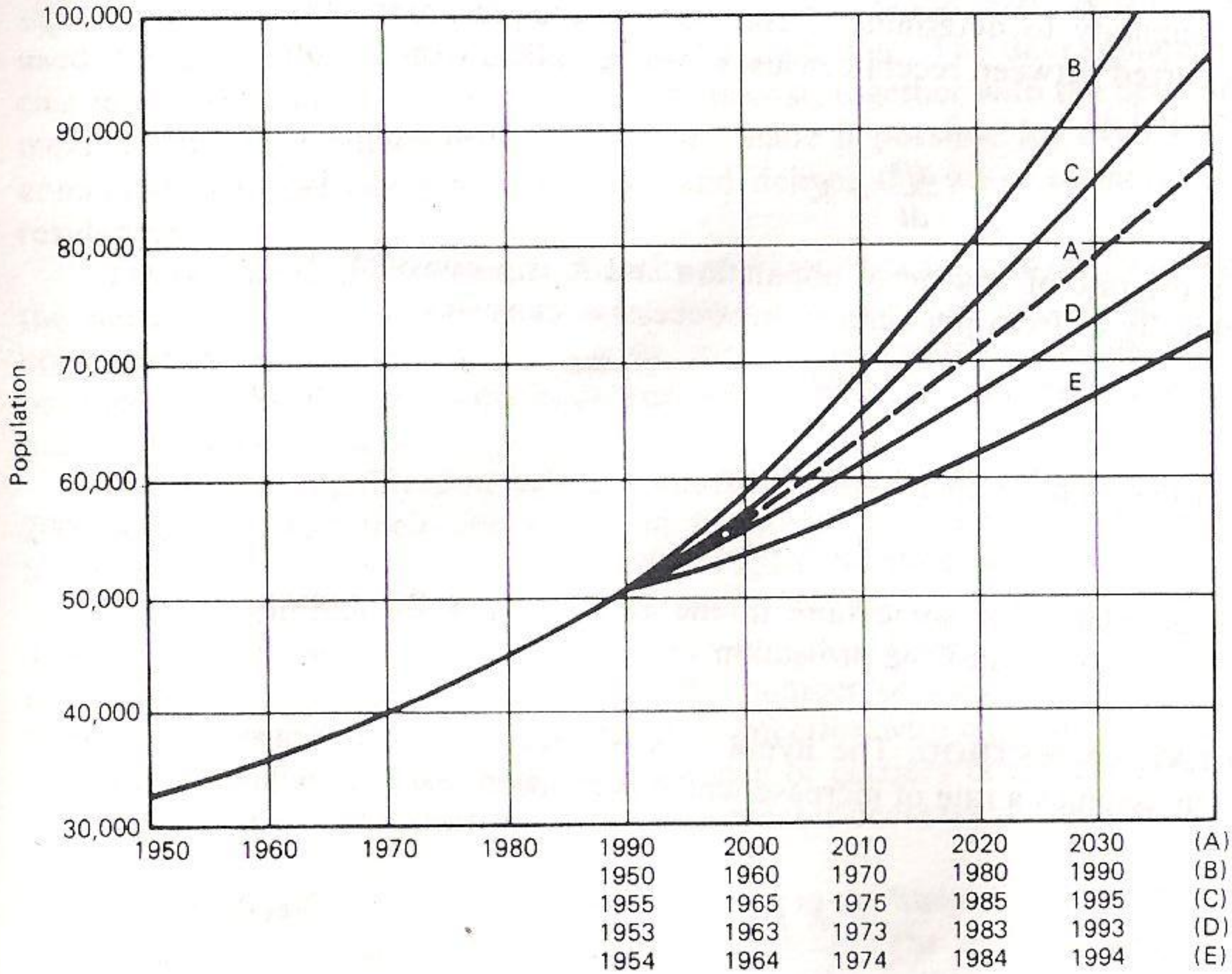
Example :

The population of a city A as per the senses records are given below for the years 1950 to 1990. Estimate the population of city A in year 2020 according to senses records for cities B,D, and E that similar to city A.

City A		City B		City C		City D		City E	
Year	Pop.	Year	Pop.	Year	Pop.	Year	Pop.	Year	Pop.
1950	32000	1910	25000	1915	25000	1913	31000	1914	30000
1960	36000	1920	32000	1925	31000	1923	35000	1924	35000
1970	٤٠٠٠٠	1930	38000	1935	36000	1933	42000	1934	42000
1980	٤٥٠٠٠	1940	43000	1945	42000	1943	46000	1944	47000
1990	51000	1950	51000	1955	51000	1953	51000	1954	51000
		1960	59000	1965	58000	1963	55000	1964	53000
		1970	69000	1975	68000	1973	61000	1974	58000
		1980	80000	1985	73000	1983	68000	1984	62000
		1990	93000	1995	86000	1993	72000	1994	68000
		2000	110000	2005	96000	2003	80000	2004	71500

City A		City B		City C		City D		City E	
Year	Pop.	Year	Pop.	Year	Pop.	Year	Pop.	Year	Pop.
1950	32000	1910	25000	1915	25000	1913	31000	1914	30000
1960	36000	1920	32000	1925	31000	1923	35000	1924	35000
1970	٤٠٠٠٠	1930	38000	1935	36000	1933	42000	1934	42000
1980	٤٥٠٠٠	1940	43000	1945	42000	1943	46000	1944	47000
1990	51000	1950	51000	1955	51000	1953	51000	1954	51000
		1960	59000	1965	58000	1963	55000	1964	53000
		1970	69000	1975	68000	1973	61000	1974	58000
		1980	80000	1985	73000	1983	68000	1984	62000
		1990	93000	1995	86000	1993	72000	1994	68000
		2000	110000	2005	96000	2003	80000	2004	71500

$$pop.(A)_{2020} = \frac{(80000 + 73000 + 68000 + 62000)}{4} = 70750, capita$$



Problems :

1- The recent population of a city is 30000 inhabitant. What is the predicted population after 30 years if the population increases 4000 in 5 years .

2- The recent population of a city is 30000 inhabitant. What is the predicted population after 30 years if the growth rate is 3.5% .

3- The population of a town as per the senses records are given below , estimate the population of the town as on 2040 by all methods.

Year	**Population**
1957	58000
1967	65000
1977	73000
1987	81000
1997	95000
2007	115000