CAPITULO 3
METODOLOGÍA

3.1 Métodos Matemáticos

El crecimiento de la población se define como el cambio que experimenta una población en cierto lapso de tiempo. En general, existen dos formas de medir dicho cambio: de manera directa que consiste en restar a la población existente al final del período (al 31 de Diciembre en caso de ser período de un año) la población que había al inicio (al 1 Enero del mismo año); o por medio de la ecuación compensadora.

Además de las formas anteriores, es posible hacer la proyección sobre la base de un supuesto matemático que suponga que el crecimiento de la misma se ajusta a una función determinada como el crecimiento geométrico o exponencial. Cabe mencionar, que en estos casos también es necesario de algunos datos adicionales como: la población inicial, el período de tiempo entre el inicio y el final de la proyección así como la tasa de crecimiento que es obtenida mediante el uso de la función respectiva. Además, este tipo de metodología no es utilizada en proyecciones de poblaciones grandes, a menos de que se trate de períodos cortos de tiempo ó bien, cuando no se dispongan de datos confiables que permitan emplear otros procedimientos de proyección. Asimismo, es conveniente mencionar que cuando proyectar la población según sexo y grupos de edad resulta favorable considerar a cada grupo de edad y sexo, por separado y aplicarles sus correspondientes tasas de crecimiento que deben ser estimadas previamente.[1] A continuación, son analizadas brevemente las funciones de tipo lineal, exponencial y geométrico, así como la curva logística y la ecuación compensadora para estos fines.

3.1.1 Crecimiento Lineal

Es el más sencillo de los supuestos de ritmo de crecimiento ya que considera un crecimiento absoluto constante en el número de individuos en una población año con año o bien período tras período. Por ejemplo, si del 1° de Enero1965 al 1° de Enero de 1972 hubo un cambio en la población de 120,000 a 190,000 habitantes, el crecimiento lineal supondría que dicho incremento de 70,000 habitantes en 7 años se debió a un incremento constante de 10,000 personas por año. La ecuación empleada pata el cálculo de dicha tasa es la siguiente:

$$P_f = P_i * (1 + r * t)$$

De donde se obtiene que:

$$r = [(P_f/P_i) - 1)] * 1/t$$

donde la P_f significa la población al final del período, P_i la población al inicio del período, r es la tasa anual de crecimiento y finalmente t representa el intervalo de tiempo en años y fracciones que dura el período.

[1] Welti, Carlos. "Demografía II" p.74.

3.1.2 Crecimiento Geométrico

Este método de proyección crece más rápidamente que el lineal ya que en este caso se podría decir que la población existente o inicial se va a estar reinvirtiendo cada intervalo de tiempo. Es decir, la población en el tiempo (t + 1), va a estar dada por la población en el tiempo t multiplicada por (1 + r_t), la población en el tiempo (t + 2) va a estar dada por la población inicial en el tiempo t y multiplicada dos veces por (1 + r_t) y así sucesivamente. Lo anterior se puede representar de la siguiente manera:

$$P_{t+1} = P_t * (1 + r_t)$$

$$P_{t+2} = P_t * (1 + r_t) * (1 + r_t) = P_t * (1 + r_t)^2$$

De este modo, la ecuación generalizadora para este tipo de crecimiento estaría dada por:

$$P_f = P_i * (1 + r_i)^t$$

De donde se desprende que:

$$r = [(P_f/P_i)^{(1/t)}] - 1$$

ya que la r inicial se considera constante para todos los años posteriores y donde la notación empleada es la misma que la usada para el método lineal.

3.1.3 Crecimiento Exponencial

El modelo de crecimiento exponencial representa un crecimiento muy rápido y continuo de la población. Este tipo de crecimiento resulta ser más razonable que los anteriores aunque es necesario tener cuidado y hacer su aplicación a corto plazo pues se corre el riesgo de que en el futuro muy lejano la población se dispare, lo que podría arrojar un resultado ilógico o fuera de lo normal. En este caso, y de manera similar a los supuestos de crecimiento anteriores, la ecuación que representa dicho crecimiento es la siguiente:

$$P_f = P_i * \exp(rt)$$

De donde se obtiene por medio de cálculos algebraicos que:

$$r = [(\ln(P_f/P_i) - 1)] * 1/t$$

Sin embargo, ninguno de estos tipos de crecimiento tiene un límite superior de tamaño para la población en un período dado. Lo anterior los hace un tanto incongruentes con la realidad ya que todas las poblaciones crecen hasta llegar a un punto en el cual ya no es posible sobrevivir para todos. Por tal motivo, surge la idea de emplear la curva logística como una forma de representar el crecimiento de poblaciones tomando en cuenta un límite máximo de crecimiento.

3.1.4 Curva Logística

Durante mucho tiempo la curva logística sirvió de modelo para representar el crecimiento de diferentes poblaciones. "Sin embargo, como la conducta de los seres humanos varía constantemente a consecuencia de actos deliberados, no debe esperarse de ordinario que el crecimiento de una población se ajuste a una fórmula matemática precisa. Aunque ya no se acepta como ley universal, la curva logística todavía se sigue empleando en algunas proyecciones demográficas, pues suele representar una tendencia que por lo menos es razonable"[2].

Cabe mencionar que la principal limitante que surge al emplear la curva logística es el hecho de definir una población máxima ya que resulta ser además de difícil, demasiado riesgoso.

La ecuación[3] que describe dicha curva es la siguiente:

$$P_t = \frac{K}{(1 + e^{a+bt})}.$$

[2] Naciones Unidas. "Métodos para preparar proyecciones de la población por sexo y edad" p.2

[3] Shryock, Henry & siegel, Jacob. "The Methods and Materials of Demography". p.443

donde:

K es la constante que representa el límite máximo que podría alcanzar la población en el futuro también denominada asíntota superior.

P_t	es la población estimada al tiempo t.
e=exp	es la base de logaritmo natural.
t	representa el período de tiempo de la proyección.
b	es una constante que representa la proporción de cambio para períodos sucesivos de t.
a	constante.

El valor de K resulta ser demasiado arbitrario al mismo tiempo que difícil de establecer considerando únicamente la variable de población por lo que muchas de las veces se recurre a usar algún indicador de densidad demográfica máxima y de allí se procede a deducir una población máxima para un valor de K. Por su parte, los valores de a y b en la segunda ecuación pueden ser estimados mediante mínimos cuadrados.

3.1.5 Ecuación Compensadora

El crecimiento de la población que puede ser positivo en caso de incrementarse y negativo en caso de verse disminuida la población inicial, puede llamarse crecimiento absoluto y viene dado por la siguiente ecuación que es la llamada ecuación compensadora:

$$C^z = B^z - D^z + I^z - E^z$$

Donde:

C^z representa el creciemiento total de la población en el año z.

B^z son los nacimientos en el año z.

D^z son las defunciones en el año z.

I^z son los inmigrantes en el año z.

E^z son los emigrantes en el año z.

Esta ecuación constituye una de las maneras más comunes y sencillas de calcular el crecimiento que ha experimentado una población en un año o período determinado ya que una vez calculada la magnitud del crecimiento ésta es sumada a la población inicial y de esta forma se obtiene la población que habrá en el futuro. Sin embargo, esta cifra no es del todo representativa pues no permite calcular la intensidad del crecimiento. De este modo, surgió la necesidad de desarrollar supuestos que en casos específicos resultan ser de gran utilidad para estimar cuál será el número de individuos que habrá en una población dentro de un tiempo específico por medio del cálculo de una tasa de crecimiento para lo cual se puede hacer uso de diferentes supuestos.

3.2 El método de los componentes

Desde hace muchos años, el "método de los componentes" se ha conocido como el modelo básico para elaborar estimaciones demográficas. Su principio básico consiste en desagregar el crecimiento de la población en sus componentes demográficos fundamentales por medio de la ecuación compensadora.

$N_{x+k} = N_x + B_{x+k} - D_{x+k} + SM_{x+k}$

donde: N_{x+k} es la población estimada k años después del año base x

N_x es la población en el año base x

B_{x+k} es el número de nacimientos ocurridos entre el año x y el año x + k

D_{x+k} es el número de defunciones ocurridas entre el año x y el año x + k

SM_{x+k} es el saldo migratorio que tuvo lugar entre el año x y el año x + k

Un segundo aspecto importante dentro de este método es que puede ser aplicado a sub-poblaciones o cohortes de personas con una o más características: sexo, edad, grupos étnicos, etc. Cuando se utiliza el método de esta forma, la población total viene dada por:

$$N_{x+k} = \sum_{i=1}^{n} Nc_{i,x+k}$$

donde: $Nc_{i,\, x+k}$ corresponde a la población de la cohorte i en tiempo x + k y

$$Nc_{i,\, x+k} = Nc_{i,x} + Bc_{i,x+k} - Dc_{i,x+k} + SMc_{i,x+k}$$

Los términos están definidos igual que antes, salvo por la especificación de la cohorte dada c_i. Una adecuada aplicación del método de los componentes requiere precisión de los datos correspondientes a sus componentes demográficos. Si alguno de ellos fuera desconocido se hace necesario llevar a cabo un diagnóstico de la dinámica de dicho componente en la región en consideración, y con base en esto postular hipótesis respecto al cambio esperado para el periodo postcensal.

El nombre de este método se debe a que su estrategia metodológica consiste en proyectar los tres componentes del cambio poblacional: mortalidad, natalidad y migración[4]. Para nuestro caso, se hará un análisis de la participación de la población en la economía, ya que como se mencionó anteriormente, lo que se quiere conocer es la Población Económicamente Activa, en el año 2015.

El primer paso en la aplicación de la técnica requiere hacer un análisis detallado del comportamiento pasado de cada uno de los componentes, con la finalidad de tratar de establecer tendencias sobre el comportamiento esperado para el futuro. En segundo lugar se debe establecer una población base distribuida por sexo y grupos de edad. Generalmente se toma el último censo de población como referencia y se realizan las correcciones de los diferentes errores que pudiera contener el censo: cobertura, declaraciones de edad y lugar de residencia, etc., que pudieran haber presentado este censo. También se hace necesario, trasladar la población del censo y su distribución por sexo y edad al 30 de junio del año que se quiere utilizar como referencia. Si se desean realizar proyecciones quinquenales entonces es conveniente efectuar este traslado hasta un año terminado en cero o en cinco.

Como última etapa, a partir de la población base se proyecta cada grupo de edad hacia el nuevo período (generalmente 5 años) de acuerdo con las tendencias esperadas en los patrones de fecundidad, mortalidad y migración. El proceso se repite tantas veces como sea necesario hasta alcanzar el año hasta el cual se pretende proyectar esta población.

[4] Estimaciones y Proyecciones de Población, Sesión 12.
http://ccp.ucr.ac.cr/cursos/demografia_03/materia/12_estimacion.htm

Por último, ya con la población proyectada en el año deseado, se aplican las tasas de participación específicas por grupos de edad, que fueron proyectadas anteriormente, de esta manera se obtiene la PEA.

3.2.1 Métodos para la Elaboración de Poblaciones Bases Confiables

Para poder trabajar con el método de componentes en necesario tener una población base confiable, sobre la cual haremos todos los cálculos antes mencionados. Generalmente los datos empleados pertenecen al último censo, desgraciadamente, estos datos no se encuentran listos para usarlos, ya que es necesario actualizarlos, o bien pueden contener errores de cobertura, sobretodo en las estadísticas vitales y en las declaraciones inexactas de edad. En esa etapa, no solo se buscará corregir errores, también se pretenderá poner la población en fecha 30 de junio del año a tomar como base.

3.2.1.1 Técnicas para Actualizar la Información Demográfica

Pocas veces se dispone de información actualizada para iniciar directamente la proyección de la población, o algunas veces puede ser reciente pero diferida a un mes diferente al que se necesita para establecerla como población base. En este caso se puede hacer uso del prorrateo, de la interpolación respecto al tiempo y de los modelos matemáticos anteriormente mencionados.

Prorrateo

Este método consiste en actualizar la información de la población por sexo y edad del último censo disponible. Por ejemplo, si el censo fue levantado en el año de 1998, la población base estaría fijada en el año 2000, por lo que sería necesario actualizar la información dos años. Es decir, si en el censo de 1998, fueron empadronadas 218000 personas y en el año 2000 se calculó que existió una población de 243000 habitantes, se obtiene una tasa simple de cambio entre el total de la población en el año del censo y la existente en el año 2000. Entonces quedaría de la siguiente manera: 243000/218000=1.1468, y esta relación se usa para estimar el total de habitantes de cada grupo de edad y sexo de los datos provenientes del censo.

Interpolación con respecto al tiempo

Este método es similar al prorrateo. Ya que de igual manera, actualiza la información del último censo a la fecha base que se desea. Pero este método permite actualizar los datos en periodos que no son necesariamente múltiplos de cinco. La hipótesis básica es que las condiciones demográficas van variando periódica y gradualmente a lo largo del tiempo y los errores originados por este método en el quinquenio son poco importantes. La desventaja que ofrece, es que al ser interpolación requiere de datos anteriores y posteriores a la fecha a la que se desea interpolar. De esta manera, si deseamos la población a mediados del año 1999, necesitaríamos contar con la información del Conteo de Población y Vivienda del 1995 y del XII Censo de Población y Vivienda.

Una vez que se cuenta con esta información, se calcula la proporción entre la fecha deseada y los mismos, por ejemplo, si el conteo fue realizado a principio del 2000, la actualización a mediados de 1999 significa que es 4 años después del Conteo de Población y Vivienda y medio año antes del XII Censo de Población, por lo que la proporción sería 4 a 1/2, es decir, 8 a 4[5]. Ya que se cuenta con las proporciones, se calcula el promedio ponderado de la población en los dos periodos. De esta manera, la población a mediados del año 1999 viene dada por:

$$\frac{8 \cdot Pob_{1995} + 4 \cdot Pob_{2000}}{12}$$

Este procedimiento se lleva a cabo para cada edad sin importar de cual se trate, ya que el método supone uniformidad de la población dentro de cada grupo.

3.2.1.2 Métodos para el Ajuste de Información Defectuosa.

Cuando se levantan los censos es muy común encontrar errores derivados de declaraciones inexactas de edad, como pueden ser los redondeos de edades pares o empadronamiento defectuoso, en lugares que quedan sin censar. Otro error común es el subregistro de

[5] Naciones Unidas. "Métodos para proyecciones de población por sexo y edad" p.11

nacimientos o de defunciones por grupos de edad específicos, este tipo de aberración es más común en el grupo de edad de 0 a 4 años de edad, ya que es el que más comúnmente queda sin registrar las defunciones o se tiene un registro tardío de los nacimientos. El no contemplar este tipo de errores nos conduciría a obtener resultados subestimados o sobreestimados, por eso la mejor opción es eliminar en lo posible este tipo de irregularidades. Para corregirlos, existen diversos métodos a emplear, que dependen del grupo de edad y de sus características, de esta manera, se tiene lo siguiente:

- **0 a 4 años de edad.-** Este grupo resulta quizá el más importante, ya que como se ha mencionado, en este grupo las defunciones son las más deficientes. Por lo tanto, el número de personas entre estas edades va a ser sustituido por los valores obtenidos para el número de nacimientos esperados, por medio de la proyección de las tasa de fecundidad.

- **5 a 9 años de edad.-** En general, este grupo es el más confiable, sin embargo, existen ocasiones en las que las estadísticas son incompletas lo que lleva a tener errores. Para corregirlos, se deben sumar las mitad de las cifras registradas a edad 5 y 10, más las cifras pertenecientes a las edades de 6,7,8,9 años y de esta manera obtendremos la población para ese grupo de edad[6].

- **10 a 74 años.-** En este grupo de edad, la fórmula de ajuste emplea la cifra correspondiente al grupo de edad que se esta ajustando, así como las cifra de de los

[6] Ibid, p.15

dos grupos anteriores y posteriores. Cada uno de los grupos se puede ajustar por medio de esta fórmula si se cuenta con la información en grupos quinquenales y hasta los 85 años de edad.

$$NAP_x = \frac{-N_{-2} + 4N_{-1} + 10N + 4N_1 - N_2}{16}$$

donde:

NAP_x se refiere al número ajustado de personas correspondientes al grupo quinquenal con edad inicial *x* que se desea calcular

N corresponde al número de personas observadas en el grupo quinquenal

N_{-1} y N_{-2} son las personas registradas en cada uno de los dos grupos quinquenales anteriores al que se desea ajustar

N_1 y N_2 representa el número de individuos que conforman los dos grupos quinquenales siguientes

Este ajuste se debe llevar a cabo para ambos sexos, aunque representa mayor importancia para el caso de las mujeres, las cuales tienden a disminuirse la edad. Es importante tener en cuenta que antes de aplicar estos métodos, es necesario tener en cuenta que no hayan ocurrido en el pasado eventos como guerras u otros que pudieran ocasionar que la tasa de natalidad se haya visto disminuida.

- Mayores de 75.- es sabido que las personas en edades avanzadas tiene a exagerar su edad, lo que ocasiona que las cifras obtenidas en el censo sean excesivas en dichos grupos de edad. Sin embargo, el número de personas en edades avanzadas es tan pequeño debido a la mortalidad que resulta de poca importancia y generalmente no es necesario hacerle ajustes.

Al hacer estos ajuste resulta lógico que la suma de la población ajustada no corresponda con la del censo, para corregir este perturbación basta con prorratear la información.

3.2.2. Proyección de la Mortalidad

Primeramente se estiman las tasas específicas de mortalidad[7] (${}_nm_x$), por sexo y grupos de edad (grupos quinquenales), para los años anteriores. Estas tasas pueden ser determinadas por medio de las estadísticas vitales o por medio de métodos indirectos para aquellos casos en que no se cuenta con información vital de calidad. De acuerdo con su evolución pasada y analizando las perspectivas futuras en mortalidad, se proyecta el patrón de comportamiento futuro de estas tasas, considerando que pueden llegar a estabilizarse en determinado momento. Con esta información, se estima el comportamiento de las tablas de vida para el futuro y específicamente se determinan las relaciones de supervivencia (${}_nP_{x,x+n-1}$), por sexo y grupos de edad, para cada uno de los períodos en que se van a realizar las proyecciones. Son precisamente estas relaciones de supervivencias las que permiten proyectar cada grupo n años hacia el futuro, suponiendo que no existe migración.

[7] Ver Anexo: Conceptos y Medidas Poblacionales.

A continuación se tiene que seleccionar el nivel al que corresponde la esperanza de vida actual y futura para ambos sexos y aplicar dicha tabla a la población base previamente calculada. Al aplicar estas tasas de sobrevivencia a un grupo de edad específico, se obtendrá la población que habrá en el grupo de edad siguiente en un periodo de cinco años.

3.2.3 Proyección de la Fecundidad

Al igual que se realizó con la mortalidad, se requieren estimar el comportamiento pasado de las tasas específicas de fecundidad y de la tasa global de fecundidad[8]. Nuevamente estas estimaciones pueden llevarse a cabo por medio de las estadísticas vitales o de métodos indirectos. Con esto se pretende proyectar estas tasas hacia el futuro, específicamente para el período de interés. Del mismo modo que se realizó con la mortalidad, se supone que la tasa global de fecundidad debe llegar a estabilizarse en cierto momento. En este caso se espera que alcance el nivel de reemplazo (TGF) para cierto año específico, y las tasas específicas se proyectan hacia ese valor. En este sentido generalmente se consideran tres supuestos básicos:

Hipótesis baja: La fecundidad alcanza el nivel de reemplazo dentro de un período de tiempo relativamente corto. En este caso se espera un decrecimiento de la fecundidad más fuerte de lo esperado.

[8] Ver Anexo: Conceptos y Medidas Poblacionales.

Hipótesis media o esperada: Supone que la fecundidad alcanza el nivel de reemplazo siguiendo el patrón de descenso que se ha venido experimentando en los años anteriores.

Hipótesis alta: Supone que la fecundidad alcanza el nivel de reemplazo en un término mayor del esperado, lo cual indica que el ritmo de descenso en la fecundidad disminuye para el período en que se pretende proyectar la población.

Esta tasas proyectadas permiten estimar los nacimientos que se esperaría ocurrieran en cada período.

Para el caso de la fecundidad, se usará el Método de la Fecundidad por Periodo[9] , este consiste en calcular las tasas de fecundidad específicas por grupo de edad, desde los 15 hasta los 49 años de edad. Así, para la obtención de las tasas en los años posteriores se pueden extrapolar las tendencias de los años anteriores o hacer un análisis de series de tiempo.

Es importante tener en cuenta que el supuesto ha utilizar depende de cada situación, es decir, de la información con la que se cuente y las necesidades.

Ya que se tienen calculadas las tasas mencionadas anteriormente, estas serán aplicadas a la población femenina, para obtener el número de nacimientos esperados. Así se usarán las tasas de fecundidad específicas por grupo de edad f_x^z para cada año de la proyección, en nuestro caso 2000, 2005, 2010 y 2015, para lo cual es necesario que sean previamente proyectadas. Una vez obtenidas, son aplicadas a la población femenina para obtener los nacimientos esperados en cada año.

[9] Shryock Henry & Siegel, Jacob. "The Methods and Materials of Demography". P.455

La población femenina a emplear es aquella que ha sido calculada previamente por medio de las probabilidades de supervivencia. Es decir, el grupo de mujeres del grupo 15-19 del año 2005, estará dado por el grupo de 10-14 del año 2000, con su respectiva probabilidad de sobrevivencia.

Ahora se debe determinar el número de nacimientos esperados para le periodo de 5 años, entonces se obtiene un promedio de los nacimientos obtenidos al inicio y final del quinquenio y después es multiplicado por 5, para así tener el total de nacimientos durante el quinquenio. Posteriormente debemos saber cuales de estos nacimientos son hombres y cuales son mujeres. Para tal efecto se consideró que por cada 100 hombres nacen 105 mujeres. Por ejemplo, si se quiere saber el número de total de nacimientos varones basta multiplicar por 100/205 por el total de nacimientos. Lo anterior, puede ser expresado de la siguiente manera:

$$B^{z...z+5}(Masc) = \frac{(B^{z} + B^{z+5})}{2} * 5 * \frac{100}{205}$$

$$B^{z...z+5}(Fem) = \frac{(B^{z} + B^{z+5})}{2} * 5 * \frac{105}{205}$$

3.2.4 Proyección de la Migración

Indudablemente el componente migratorio es el más complejo para predecir su comportamiento futuro. No basta con efectuar un análisis retrospectivo para poder analizar lo que podría pasar en el futuro. Se requiere además, de todo un estudio sobre diferentes elementos tales como: la economía, la situación política, situación laboral, climatológica, etc., pero no solo de la localidad para la cual se está proyectando la población, sino para las localidades vecinas. Aun con estudios de este tipo resulta sumamente difícil poder aproximar lo que se espera de este componente a mediano o largo plazo. Por esta razón, muchas veces se proyecta la población suponiendo únicamente una migración constante o decreciente en su magnitud, utilizando como referencia lo que se ha venido presentando en el pasado. También es posible, suponer que no se presentará migración en el futuro.

Dentro de lo métodos empleados en el cálculo de la migración se encuentran el de estadísticas vitales y el de coeficiente de supervivencia, ambos métodos representan una buena opción en el caso de que no se cuente con ningún tipo de registro sobre el número de migrantes. El primero de ellos estima únicamente la población migrante para años anteriores por medio de las diferencias de población existentes entre dos censos sucesivos y restándoles a estas últimas el crecimiento natural (número de nacimientos menos el número de defunciones) de la población durante el periodo intercensal. El segundo, estima el número de migrantes restando de la población censal existente, la población del censo anterior aplicándoles las probabilidades de vida. Sin embargo, si se cuenta con información, se pueden calcular las tasas de migración[10] correspondientes y de esta manera analizar el comportamiento de la migración a través de los años. Esta última opción es la mejor, puesto que es más confiable.

[10] Ver Anexo: Conceptos y Medidas Poblacionales.

3.2.5 Proyección de la Participación Económica

Para la proyección de la participación económica, es necesario el estudio de la tendencia que han tenido en los últimos años las tasas de participación específicas por grupos de edad[11], es importante mencionar que el primer grupo de edad va de los 12 a los 14 años, los demás se comportan de la misma para los estudios anteriores, siendo el último grupo el de 65 años en adelante. La evolución de las tasas de participación ha mostrado una tendencia constante en los últimos once años para el caso de los hombres, por su lado, en el caso de las mujeres se ha observado un incremento en sus tasas de participación. Para el estudio de esta variable lo más recomendable hacer un análisis de series de tiempo. Una vez con, las tasas de participación proyectadas, bastará aplicarlas a nuestra población proyectada en el año deseado para obtener la PEA.

Dentro de los estudios de series de tiempo se encuentran: modelo multiplicativo, promedios móviles, suavizamiento exponencial.

3.2.5.1 Modelo multiplicativo

$$Y_t = S_t * T_t * I_t$$

Donde:

S_t Se obtiene mediante el cálculo de promedios móviles centrados

[11] Ver Anexo: Conceptos y Medidas Poblacionales.

T_t Tendencia proyectada mediante la ecuación de regresión

$T_t = b_0 + b_1 t$

I_t Componente Aleatorio

3.2.5.2 Promedios Móviles

$$\text{Pr}\,omediomóvil = \frac{Suma\_de\_los\_n\_valores\_más\_recientes}{n}$$

3.2.5.3 Suavizamiento exponencial

$$F_{t+1} = F_t + \alpha(Y_t - F_t) \qquad\qquad F_2 = Y_1$$

donde:

F_{t+1} Pronóstico para el periodo t+1

F_t Pronóstico para el periodo t

α Constante de suavizamiento (entre 0 y 1)

Y_t Valor real de la serie en el tiempo t

3.3 Modelo Poblacional

Para el cálculo y análisis de las tasas de mortalidad, natalidad y migración, es necesario conocer la población en años consecutivos, para nuestro caso, necesitaremos la población en los años 1990, 1995, 1996, 1997, 1998, 1999 y 2000, desgraciadamente, solo conocemos la población en los años 1990, 1995 y 2000, esta información proviene de los Censos Nacionales y del Conteo de Población y Vivienda. De esta manera se hace necesario usar un modelo poblacional para estimar la población en los años mencionados. Para esto emplearemos la ecuación conocida como ley logística del crecimiento de una población[12].

Denótese por *p(t)* la población en el tiempo *t* y por *r(t,p)* la diferencia de sus tasas de natalidad y mortalidad. Si se supone que la población es cerrada, se debe demostrar que el *dp/dt*, el cambio de la población es igual a *r*. tomando un modelo con *r* constante, es decir una *r* que no dependa del tiempo no de la población. La ecuación diferencial que representa el crecimiento del tiempo esta dada por:

$$\frac{dp(t)}{dt} = a \cdot p(t) \qquad a = cte.$$ [13]

Si además la población el tiempo t_0 está dada por $p(t_0)$ entonces la solución al problema inicial viene dado por:

[12] Braun, M. "Ecuaciones Diferenciales y sus Aplicaciones" p. 28
[13] Ecuación conocida como La Ley de Maltus para el crecimiento de una población.

$$p(t) = p_0 \cdot \exp[a \cdot (t - t_0)]$$

Sin embargo, es necesario considerar que cuando la población es grande, este modelo no puede ser exacto ya que no refleja la competencia de los individuos por comida, espacio, etc. Por lo anterior, se debe introducir un término de competición, en este caso introduciremos $(-b \cdot p^2)$ donde b es constante, ya que el promedio estadístico del número de encuentro por unidad de tiempo es proporcional a p^2. Esta ecuación fue propuesta por el biólogo Verhulst y se expresa de la siguiente manera:

$$\frac{dp}{dt} = ap - bp^2 \qquad\qquad p(t_0) = p_0$$

Donde la constante *b* es muy pequeña en comparación con *a,* de tal modo que si *p* no es demasiado grande, entonces el término $-bp^2$ es insignificante comparado con *ap,* por lo que la población crece exponencialmente. Sin embargo si *p* es grande el término $-bp^2$ debe tomarse en cuenta ya que disminuye la tasa de crecimiento de la población. Así mientas más industrializado sea un país, tanto más espacio disponible posee y cuanto más alimento tiene, entonces más pequeño será el coeficiente *b*.

Resolviendo la ecuación se tiene:

$$\int_{p_0}^{p} \frac{dr}{ar - br^2} = \int_{t_0}^{t} ds = t - t_0$$

Pero para poder resolver la primera integral es necesario emplear fracciones parciales con lo que se obtiene lo siguiente:

$$\frac{1}{ar - br^2} \equiv \frac{1}{r \cdot (a - br)} = \frac{A}{r} + \frac{B}{a - br}$$

Para encontrar *A y B* se observa que:

$$\frac{A}{r} + \frac{B}{a + br} = \frac{A \cdot (a - br) + Br}{r \cdot (a - br)} = \frac{Aa + (B - bA) \cdot r}{r \cdot (a - br)}$$

Por tanto, *Aa+ (B-bA)*r=1* y como esta ecuación es cierta para $r \in \Re$ se llega a que:

$$Aa = 1 \qquad y \qquad B - bA = 0$$

$$A = \frac{1}{a} \qquad y \qquad B = \frac{b}{a}$$

Una vez obtenidos los valores para A y B se puede resolver la integral anterior

$$\int_{p_0}^{p} \frac{dr}{r \cdot (a - br)} = \frac{1}{a} \cdot \int_{p_0}^{p} \left(\frac{1}{r} + \frac{b}{a - br} \right) dr$$

si se hace $u = a - br$, entonces $du = -b$ y :

$$\int_{p_0}^{p} \frac{dr}{r \cdot (a - br)} = \frac{1}{a} \cdot \left[\left(\ln p - \ln p_0 \right) - \int_{a-bp_0}^{a-bp} \frac{du}{u} \right]$$

$$\int_{p_0}^{p} \frac{dr}{r \cdot (a - br)} = \frac{1}{a} \cdot \left[\left(\ln \frac{p}{p_0} \right) + \left(-\ln\left|a - bp\right| + \ln\left|a - bp_0\right| \right) \right]$$

$$\int_{p_0}^{p} \frac{dr}{r \cdot (a - br)} = \frac{1}{a} \cdot \left[\left(\ln \frac{p}{p_0} \right) + \left(\ln \left| \frac{a - bp_0}{a - bp} \right| \right) \right]$$

$$\int_{p_0}^{p} \frac{dr}{r \cdot (a - br)} = \frac{1}{a} \cdot \ln \frac{p}{p_0} \cdot \left| \frac{a - bp_0}{a - bp} \right|$$

Por tanto,

$$a \cdot (t - t_0) = \ln \frac{p}{p_0} \cdot \frac{a - bp_0}{a - bp}$$

$$\exp\left[a \cdot \left(t - t_0\right)\right] = \frac{p}{p_0} \cdot \frac{a - bp_0}{a - bp}$$

Si se factoriza y se despeja p se llega a la siguiente ecuación:

$$p(t) = \frac{ap_0}{bp_0 + (a - bp_0) \cdot \exp\left[-a \cdot (t - t_0)\right]}$$ [14]

[14] Ecuación a emplear para proyectar la población en los años deseados.

Obsérvese que cuando $t \rightarrow \infty$:

$$p(t) \rightarrow \frac{a}{b}$$

Lo que quiere decir que independientemente del valor inicial, la población tenderá al valor límite a/b. Nótese que *p(t)* es una función monótona creciente en t para $0 \langle p_0 \langle a/b$. Además, *dp/dt* es creciente si *p(t)<a/2b* y decreciente si *p(t)>a/2b*. De esta manera, la gráfica para *p(t)* tiene la forma de *"S"* denominada logística.

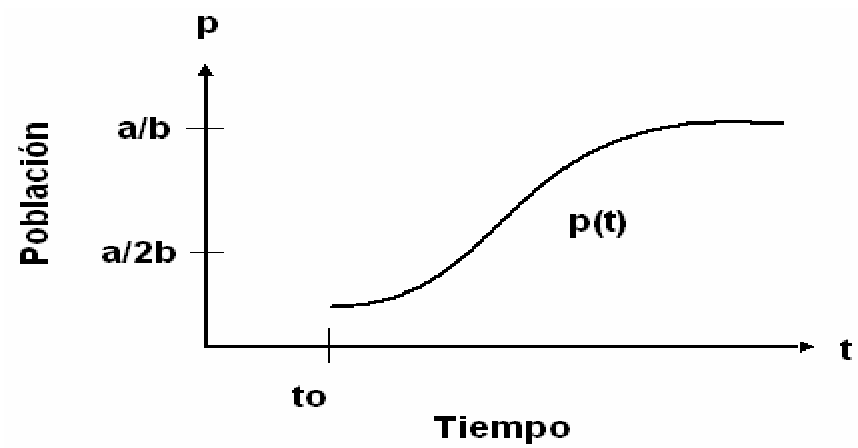

Figura 3.1 Gráfica de p(t)

Fuente: Braun, M. "Ecuaciones Diferenciales y sus Aplicaciones" p. 30

De la gráfica se concluye que el tiempo antes de que la población alcance la mitad de su valor límite es un periodo de crecimiento acelerado. A partir de este punto, la tasa de crecimiento disminuye hasta llegar a cero.

3.4 Regresión Lineal Simple

Un modelo de regresión es muy útil para hacer una representación simplificada de la realidad, mediante el uso de expresiones simbólicas, por medio de un conjunto de relaciones matemáticas. El uso de este tipo de modelos permiten incorporar variables aleatorias y el uso de los métodos de la inferencia estadística a partir de datos observados sobre las magnitudes incluidas en el propio modelo.

El modelo de regresión lineal simple cumple con las siguientes características: lineal, uniecuacional, estático, con una variables endógena, una variable exógena, un término de error y dos parámetros. La formula teórica es:

$$Y = \beta_1 + \beta_2 X + \varepsilon$$

Donde Y representa a la variable endógena, X a la variable exógena, ε es la variable aleatoria que representa al término de error y β_1 y β_2 son los parámetros. Si consideramos los "N" elementos maestrales que se van a observar se puede expresar como:

$$Y_i = \beta_1 + \beta_2 X_i + \varepsilon_i \qquad \text{i= 1,…,N.}$$

Siendo:

Yi Observaciones muestrales de la variable Y, variable cuyo comportamiento pretende explicar el modelo.

Xi Observaciones muestrales de la variable exógena X, variable elegida como explicativa e influyente sobre la variable Y.

εi Perturbaciones aleatorias o términos de error.

β_1 y β_2 Son los parámetros estructurales del modelo que permanecen invariantes respecto a la variación muestral.

Las hipótesis básicas de cualquier modelo de regresión lineal se exponen a continuación:

- Linealidad (respecto a los parámetros)
- Las variables Xi son no estocásticas.
- Media Nula: $\forall i = 1,...,N \quad E[\varepsilon i] = 0$
- Varianza Constante: $\forall i = 1,...,N \quad V[\varepsilon_i] = \sigma^2$
- No autocorrelación: $\forall i \neq j \quad Cov[\varepsilon_i, \varepsilon_j] = E[\varepsilon_i \cdot \varepsilon_j] = 0$
- Normalidad: $\forall i = 1,...,N \quad \varepsilon_i : N(0, \sigma^2)$

Un modelo de regresión puede violar algunas de estas hipótesis, el hacerlo trae consigo que la estimación de nuestra variable exógena sea incorrecta, los problemas más frecuentes que suelen presentarse son: existencia de autocorrelación en los residuos, igualdad en las varianzas de las perturbaciones aleatorias, inexistencia de normalidad en los residuos. Para cada uno de estos problemas existen contrastes para su posible detección y formas de solucionarlos, para la obtención de un modelo confiable. Los contrastes más comunes para detectar la autocorrelación de los residuos y la igualdad en las varianzas de las perturbaciones aleatorias más comunes son Durbin-Watson y Golfeld-Quandt respectivamente.

3.4.1 Regresión Lineal Múltiple

Un modelo de regresión lineal simple guarda mucha relación con un modelo de regresión múltiple, la diferencia radica en que este último involucra más de una variable explicativa. Este tipo de modelos se presenta de la siguiente manera:

$$Yi = \beta_0 + \beta_1 x1_i + \beta_2 x2_i + \beta_3 x3_i + e_i$$

$$i = 1,\ldots,N$$

La única diferencia es que en este caso se cuenta con más de una variable explicativa, para este tipo de análisis existen diversos métodos para la construcción de un modelo, la mayor parte pretenden incorporar en el modelo solo aquellas variables que sean mas representativas. Entre los métodos para construcciones de modelos se encuentran:

- **Todas las regresiones posibles**.- Este método se basa en que de las $2^k - 1$ regresiones posibles utilizando métodos reducidos, donde K es el número de regresores, se escoge la mejor de todas, con base en la mejor R^2 y la menor $\hat{\sigma}^2$.

- **Método retrospectivo**.- Consiste en determinar la ecuación con todas las variables independientes. Calcular la F-parcial para cada *X* como si hubiera sido la última en entrar al modelo. Ver la menor, si es significativa, sí la incluimos y acabamos; si no lo es, la eliminamos y recalculamos.

- **Método Prospectivo**.- Para este método se elige la variable independiente con correlación más alta con *Y* y calculamos su ecuación. Se procede a calcular la F-parcial de todas las demás con la primera del modelo. Se observa si la F-parcial más grande es significativa. Si lo es, se incorpora y se vuelve a calcular la ecuación con ella incluida. Si no lo es, so se incluye y se da por concluida.

- **Stepwise**.- Para este método se realizan los mismos pasos que para la selección prospectiva, pero ahora al tener una nueva ecuación, calculamos las F-parciales nuevamente, incluyendo ahora la F-parcial de la primera en el nuevo modelo. De esta manera se examinará nuevamente en cada paso las variables previamente incorporadas en los pasos anteriores.

Además de usar estos métodos, se puede hacer uso de las variables denominas Dummy, las cuales nos ayudan a descubrir si hay algún tipo de diferencias entre grupos que se pudieran formar dentro de los datos.

Una vez que se cuenta con el modelo se procede a hacer la estimación de la variable endógena.