CAPÍTULO 3

METODOLOGÍA

En el siguiente capítulo se presentan, definiciones de algunos conceptos demográficos que se utilizan para la elaboración de una tabla de mortalidad al igual que la metodología que se va a seguir para la proyección de la con el objetivo de que el lector se familiarice con ellos.

3.1 Población Estacionaria

La tabla de mortalidad o tabla de vida, es un instrumento que permite medir la probabilidad de vida y de muerte de acuerdo a la edad de la persona. Este esquema describe de una forma muy completa como se comporta la mortalidad en una población, además de ser la base del modelo de población estacionaria.

La tabla de mortalidad cuenta con ciertas características principales, las cuales se detallan a continuación:

i) Describe un comportamiento de la mortalidad de la población por edades.

ii) Esta se considera como una de las más importantes ya que nos permite obtener varias probabilidades y algunas medidas de mortalidad (como la tasa de mortalidad), que nos permite calcular el número de sobrevivientes existentes en una población.

iii) Permite establecer un comportamiento de la mortalidad por edades, además de utilizarse como una medida que resume las características de la población. Un ejemplo es la esperanza de vida al nacer, que se considera como el mejor de los indicadores de la mortalidad de una población.

iv) Puede considerarse como un modelo teórico de población al que llamaremos población estacionaria.

v) Tiene varias aplicaciones, como la de estimar la mortalidad, evaluar programas de salud, medir la fecundidad, migración entre otras.

La Población Estacionaria se considera como un modelo teórico de población, el cual considera una mortalidad y un número de nacimientos constante. Como resultado de estos supuestos la población total y la forma de la distribución por edades no tiene ningún tipo de variación. También considera que la tasa de natalidad y de mortalidad es equivalente y como característica final es que no existe ningún tipo de crecimiento en la población. Este modelo sirve para proyectar la población por edades además de permitir estudios sobre la población.

De a cuerdo a Ortega, la tabla de mortalidad cuenta con la siguiente notación:

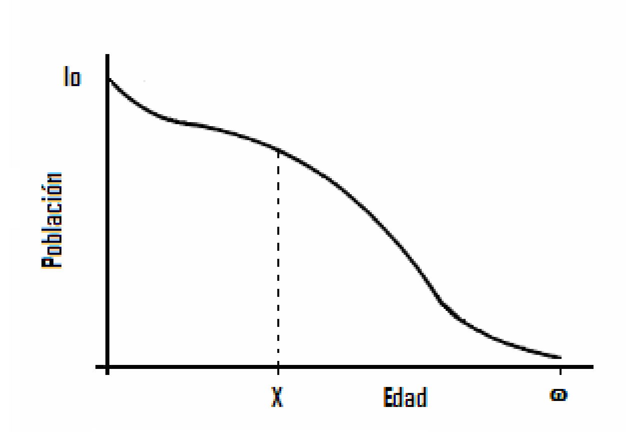
l_x: Número de personas que alcanzan con vida la edad exacta x, de una generación inicial de l_o nacimientos.

l_o: Número de personas con los que inicia la tabla.

ω (Omega): Edad en la cual el número de sobrevivientes es cero.

En la gráfica 1, observemos que la función de sobrevivencia es una función decreciente y positiva.

Gráfica 1



Fuente: Elaboración propia

d_x: Número de muertes ocurridas, a una generación inicial de l_o nacimientos, entre las edades exactas “x” y “x+n”.

Esta expresión puede también la podemos expresar como:

(1) $$d_x = l_x - l_{x+1}$$

De acuerdo a lo anterior podemos generalizar este termino a

(2) $${}_n d_x = l_x - l_{x+n}$$

que simboliza, el número de defunciones entre las edades “x” y “x+n”.

q_x: Representa la probabilidad de muerte o la probabilidad que tiene una persona de edad exacta “x” de fallecer dentro del año que sigue al momento en que alcanza dicha edad.

También puede ser representada como:

(3) $$q_x = \frac{l_x - l_{x+1}}{l_x} = \frac{d_x}{l_x}$$

Para el cálculo de la probabilidad dentro de un intervalo en edades de “n” años, la fórmula queda de la siguiente manera:

(4) $${}_{n}q_{x} = \frac{l_x - l_{x+n}}{l_x} = \frac{{}_{n}d_x}{l_x}$$

Si conocemos el valor de $\mathbf{l_o}$ y el de $\mathbf{q_x}$ para cada edad o grupos de edades, la fórmula queda de la siguiente manera:

(5) $$l_x \cdot {}_{n}q_x = {}_{n}d_x$$

$\mathbf{p_x}$**:** Representa la probabilidad de sobrevivencia o la probabilidad que tiene una persona de edad exacta “x” de llegar con vida a la edad “x+1.

(6) $$p_x = \frac{l_{x+1}}{l_x}$$

Si calculamos la probabilidad, de que una persona de edad “x”, sobreviva hasta el siguiente periodo o de que muera en el siguiente periodo, la fórmula, queda:

(7) $$p_x + q_x = 1$$

Al despejar la fórmula anterior p_x queda de la siguiente manera:

(8) $$p_x = 1 - q_x$$

Al generalizar la fórmula **(6)** para un periodo de n años y teniendo el número de sobreviviente a edad “x” y “x+n”, la fórmula queda:

(9) $${}_{n}p_x = \frac{l_{x+n}}{l_x}$$

Si calculamos la probabilidad de que una persona de edad x llegue con vida a edad “x+1” y cuando tenga edad “x+1” calculamos la probabilidad de que llegue con vida a edad “x+2” y así sucesivamente hasta “x+n” la fórmula queda de la siguiente manera:

$$ {}_{n}p_{x} = p_{x} \cdot p_{x+1} \cdot p_{x+2} \cdot \ldots \cdot p_{x+n} \quad \textbf{(10)} $$

En general podemos decir que:

$$ {}_{n}p_{x} \cdot {}_{m}p_{x+n} = {}_{n+m}p_{x} \quad \textbf{(11)} $$

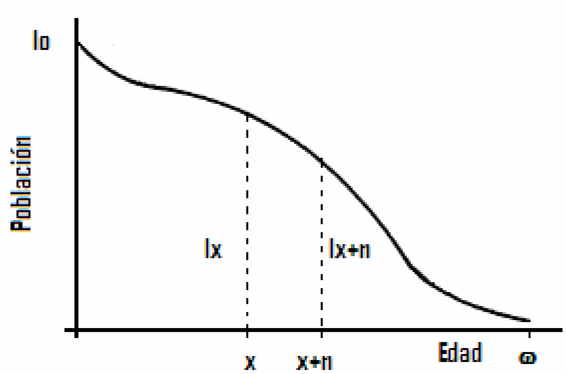
Cabe mencionar que la notación es internacional y usada por los actuarios y que mas adelante se utilizara la notación utilizada por los demógrafos un ejemplo de esto es la siguiente fórmula, donde $_{\mathbf{x}}\mathbf{p_o}$ es notación ocupada por actuarios y **p(x)** por demógrafos:

$$ p(x) = {}_{x}p_{0} = \frac{l_x}{l_0} \quad \textbf{(12)} $$

$\mathbf{{}_{n}L_{x}}$**:** Simboliza el tiempo vivido entre las edades “x” y “x+n”.

Matemáticamente lo podemos ver como la integral entre dos puntos, en este caso “x” y “x+1”, como se muestra en la gráfica 2.

Gráfica 2



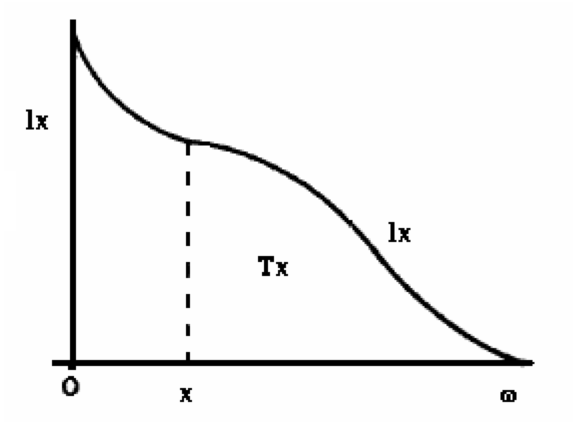
Fuente: Elaboración propia

La notación con la que se representa es:

$$ {}_{n}L_{x} = \int_{x}^{x+n} l_a \, da \quad \textbf{(13)} $$

T_x: Representa el tiempo vivido entre las edades “x” y omega.

Gráfica 3



Fuente: Elaboración propia

Se representa mediante la siguiente fórmula:

$$T_x = \int_x^w l_a da \tag{14}$$

La fórmula general se puede generalizar, como:

$$T_x = \int_x^{x+1} 1_a da + \int_{x+1}^{x+2} 1_a da + \cdots \tag{15}$$

$$= L_x + L_{x+1} + \cdots$$

e^o_x: Representa la esperanza de vida a la edad x.

Si dividimos el tiempo vivido **T_x** por el número de personas a edad “x” **l_x**, obtenemos la esperanza de vida a edad “x”.

$$e_x^o = \frac{T_x}{l_x} = \frac{\int_x^w l_a da}{l_x} \tag{16}$$

Anteriormente vimos que una tabla de mortalidad o de vida, consiste en seguir una generación de personas a través del tiempo a las que se les aplican condiciones de mortalidad y se le establece un número de sobrevivientes a cada edad.

Ahora denotaremos la notación que es utilizada por los demógrafos en la población estacionaria.

N(t): Población total en el tiempo t

B(t): Nacimientos que ocurren en el año t

N(x,t): Representa el número de personas que tienen edad x en el tiempo t. esto lo podemos ver como:

(17) $$B(t-x)*p(x)= N(x,t)$$

Para determinar la población total en el tiempo t, se obtiene dando a los “x” valores de cero a omega, entonces

(18) $$N(t) = \int_{0}^{w} B(t-x)p(x)dx$$

Al realizar cambios en los limites de integración y dejarlos de “x” a “x+n”, podemos obtener la cantidad de personas que existen en un grupo de edades, quedando la fórmula de las siguiente manera.

(19) $$N^{t}(x, x+n) = \int_{x}^{x+n} B(t-a)p(a)da$$

De forma similar a lo anterior, podemos obtener las defunciones totales y las defunciones en un grupo por edades entre “x” y “x+n”, solamente se necesita multiplicar el numero de sobrevivientes a edad “x” por la fuerza de mortalidad o tasa instantánea de mortalidad que se representa por **μ(x)** que están representadas por la siguiente formulas:

Para las defunciones totales, es:

(20) $$D(t) = \int_0^w B(t-x)\,p(x)\mu(x)dx$$

Para defunciones por grupo de edades entre “x” y “x+n”, es:

(21) $$D^t(x, x+n) = \int_x^{x+n} B(t-a)\,p(a)\mu(a)da$$

Anteriormente se mencionaron cinco características de la población estacionaria, ahora se representaran en seis características de acuerdo a la notación que se ha visto hasta el momento.

i) Si hacemos **B(t-x)= l_o** en **(18)**, obtendremos la población total

(22) $$N(t) = \int_0^w l_0 * \frac{l_x}{l_0} dx = \int_0^w l_x dx = T_0 = cte$$

Esto nos indica que en una población estacionaria el número de personas en el tiempo se mantiene constante y esto es representado por **T_0** de la tabla de mortalidad.

ii) Si sustituimos **B(t-x)=l$_o$** obtendremos la población estacionaria por edades y esto será igual a **$_nL_x$** que representa la tabla de mortalidad.

(23) $$N(x, x+n) = \int_{x}^{x+n} l_a da =_n L_x = cte$$

Para las defunciones totales, tenemos:

(24) $$D(t) = -\int_{0}^{w} l_0 * \frac{l_x}{l_0} * \frac{1}{l_x} * \frac{dl_x}{d_x} d_x = -\int_{0}^{w} dl_x = -(l_w - l_0) = l_0$$

Significa que las muertes en el año son constantes en el tiempo.

iii) Para defunciones por edades la obtendremos remplazando l$_o$ en **(21)**. Otra forma de obtener la fórmula es integrar de "x" hasta "x+n".

(25) $$D(x, x+n) = -\int_{x}^{x+n} dl_a = -(l_{x+n} - l_x) =_n d_x$$

Con la información anterior podemos sacar nuevas tasas, como la de mortalidad y la de mortalidad.

Para la tasa bruta de natalidad la representamos como, el cociente entre nacimientos anuales y la población total.

(26) $$b = \frac{B(t)}{N(t)} = \frac{l_0}{T_0} = \frac{1}{e_0^0}$$

De forma similar para la tasa de mortalidad, lo representaremos como el cociente entre muertes anuales y la población total.

$$d = \frac{D(t)}{N(t)} = \frac{l_0}{T_0} = \frac{1}{e_0^0} \qquad \textbf{(27)}$$

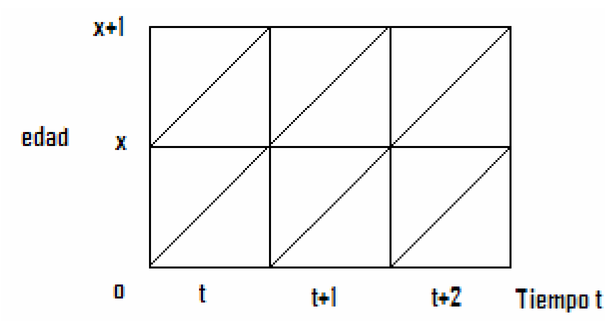
iv) Para terminar con las características de la población estacionaria podemos determinar la tasa de crecimiento natural, como la diferencia entre la tasa bruta de mortalidad y la de natalidad que deberá de ser igual a cero.

$$r = b - d = 0 \qquad \textbf{(28)}$$

3.2 Diagrama de Lexis

Para el enfoque que le daremos al diagrama de Lexis, el eje de las "x" representa el tiempo "t" mientras que el eje de las "y" representa la edad de la persona. En la gráfica 4 se muestra la forma del diagrama.

Gráfica 4

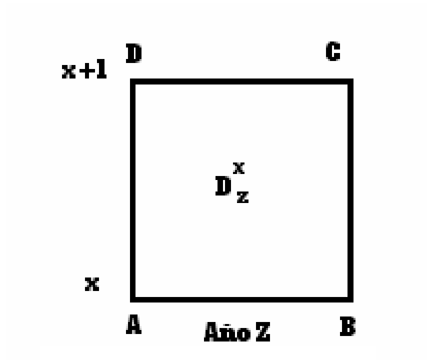


Fuente: Elaboración propia

Las líneas diagonales a 45 grados representan la línea de vida de cada individuo de una población que parte de la edad "x" igual a cero y en el momento en el que nace el tiempo t.

Si deseamos calcular la tasa de mortalidad de las personas de edad “x” para un año “z”, lo podemos representar con el diagrama de Lexis de la siguiente forma:

Gráfica 5



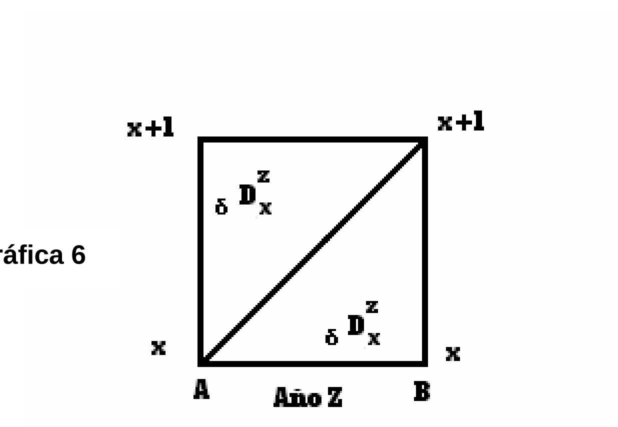
Fuente: Elaboración propia

Cuando los datos contienen errores, se necesita separar las defunciones que ocurren en cada cohorte y se utilizan los factores de separación (f_x).

Se conoce como factores de separación (f_x) a la relación:

$$f_x^z = \frac{{}_{\delta}D_x^z}{D_x^z} \qquad \textbf{(29)}$$

Gráfica 6



Fuente: Elaboración propia

3.3. Proyección de la Población Económicamente Activa Según Área Urbana y Rural Por Sexo y Edad

En la siguiente sección se describirá la metodología que utiliza el CELADE para la proyección de la población urbana y rural, por sexo y grupos quinquenales de edades.

Las proyecciones de la población económicamente activa se realizan por sexo y edades en los grupos quinquenales para áreas rurales urbanas.

Para la proyección de la PEA se debe utilizar la información que se ha obtenido de los últimos censos del país.

Se deberá establecer un año base para realizar la proyección, tanto para la población rural como para la urbana.

Después de realizar los pasos anteriores, se deberá determinar la tasa por actividad por sexo y área rural y urbana. Si no existe la información para el año base se realizará una interpolación de los últimos dos censos realizados.

Para definir la PEA que corresponda a nuestro país se tomará la edad mínima o edad legal a la que se puede trabajar, que es a partir de los 16 años. Pero para nuestra investigación se tomará el grupo de 15 a 19 años, porque se realizará una tabla quinquenal.

Ya que se tenga la población que se desea estudiar, población urbana o rural, se aplicarán los procedimientos establecidos por el CELADE.

Para proyectar las tasas de actividad de la población masculina se tomará uno de los dos modelos que previamente ha establecido el CELADE. Uno de los modelos se refiere a un desarrollo industrial que corresponde a personas menores de 20 años y mayores de 55 años. El otro modelo se refiere a un desarrollo semi-industrial el cual comprende a las personas entre 22 y 55 años, la diferencia con el primer modelo es que en este intervalo se concentra una mayor participación. Las tasas de actividad que se utilizarán se muestran en la tabla 3.3.1.

TABLA 3.3.1

Modelos Límites de Tasa de Actividad por edad de la Población Masculina urbana Adoptadas para el año 2030

Edad	Modelo I Industrializado	Modelo II Semi-industrializado
10 - 14	2.5	4.4
15 - 19	32.8	43.8
20 - 24	83.1	83.1
25 - 29	93.8	93.8
30 - 34	96.9	96.9
35 - 39	96.8	96.8
40 - 44	95.6	95.6
45 - 49	93.7	93.7
50 - 54	89.1	89.1
55 - 59	78.1	83.4
60 - 64	54.4	75.6
65 - 69	31.6	66.3
70 - 74	10.0	52.5
75 - 79	2.5	39.1
80 - 84	1.0	25.3

Fuente: Centro Latinoamericano de Demografía "Métodos para proyecciones demográficas", p.152, año 184

Para la proyección de las tasa de actividad masculino rural, se necesitan las tasas de participación de la población rural masculina rural. Al igual que en la proyección anterior se consideraron dos modelos. El primer modelo de los 15 a 20 años y de 55 años y más. Para los siguientes años el CELADE consideró una tasa constante, es por eso que se tomarán las tasas de actividad realizadas por el Conteo de Población y Vivienda 2005. Las tasas que considero se el CELADE se muestran en la tabla 3.3.2

TABLA 3.3.2

Modelos Límites de Tasas de Actividad por edad de la Población Masculina Rural menor de 25 años y mayor de 55 Años de edad Para el año 2030

Edad	Modelo I	Modelo II
15 - 19	52	65
20 - 24	83	88
55 - 59	81	92
60 - 64	68	85
65 - 69	45	75
70 - 74	27	65
75 - 79	10	52
80 - más	3	25

Fuente: Centro Latinoamericano de Demografía "Métodos para proyecciones demográficas", p.153, año 184

Para la proyección de las tasa de actividad de la población rural femenina y urbana femenina, se mantendrá constante para todos los años en que se proyecte.

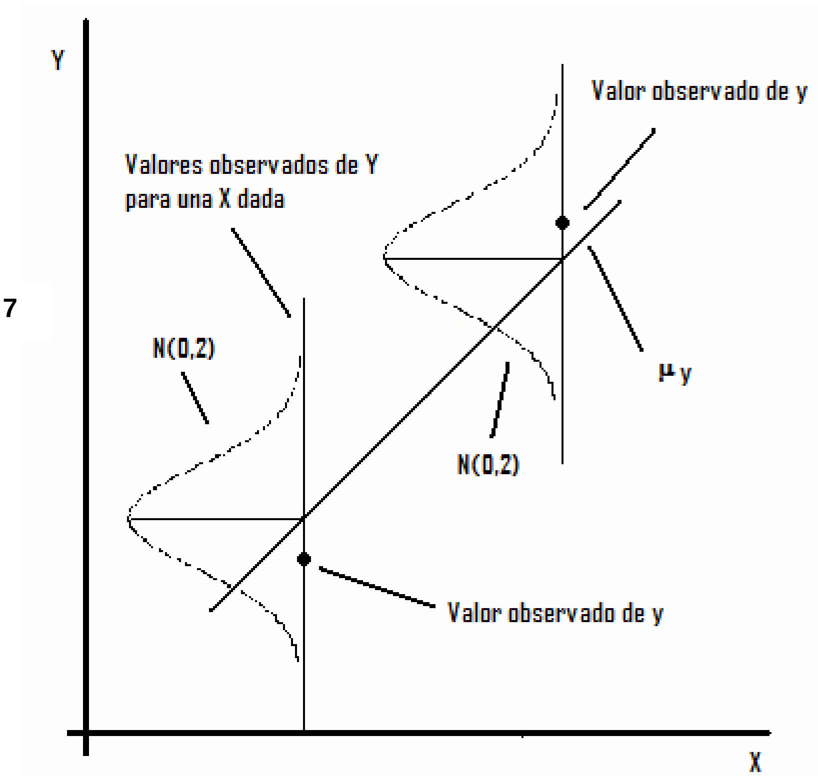
3.4 Regresión Lineal Simple

Diremos que el modelo de regresión lineal simple, es un modelo con un solo regresor "x" que tiene una relación con una respuesta "y", donde la relación es una línea recta, (Montgomery, Peck y Vining, 2002).

El objetivo que tiene la regresión lineal para nuestro estudio será el de predecir. Todo análisis de regresión simple tiene como finalidad obtener una función que nos exprese la relación entre dos variables, comúnmente se dice que “x” es nuestra variable independiente y “y” nuestra variable dependiente.

En la gráfica 7 se muestra como se generan las observaciones en la regresión lineal.

Gráfica 7



Fuente: Elaboración propia

Cuando el modelo esta compuesto por sólo una variable independiente, nuestro modelo estará representado por, lo siguiente:

$$\textbf{(3.4.1)} \qquad Y = \beta_0 + \beta_1 \cdot x + \varepsilon$$

donde,

$\beta_0 =$ Ordenada a origen

$\beta_1 =$ Pendiente

$\varepsilon =$ Error de la regresión

Los supuestos que tiene es te modelo son los siguientes:

i) $E[\varepsilon_i]=0$

ii) $E[\varepsilon_i{}^2]=\sigma^2 \quad \forall i>0$

iii) $E[\varepsilon_i\varepsilon_j]=Cov[\varepsilon_i,\varepsilon_j]=0$

iv) $E[\varepsilon_i x]=Cov[\varepsilon_i,x]=0$

De acuerdo con el i) podemos decir que $\varepsilon_i \longrightarrow N(0,\sigma^2)$, entonces

$$\textbf{(3.4.2)} \qquad E[Y]=\beta_0+\beta_1\cdot x$$

y la varianza es

$$\textbf{(3.4.3)} \qquad Var[Y]=Var[\beta_0+\beta_1\cdot x+\varepsilon]=\sigma^2$$

A los parámetros β_0 y β_1 se les llama coeficientes de regresión. La interpretación que se les da a los parámetros β_0 y β_1 es que la pendiente β_1 es el cambio que tiene la media de la distribución de "y" por cada unidad que cambia "x".

3.4.1. Estimación de los parámetros por mínimos cuadrados

Como los parámetros β_0 y β_1 son desconocidos, necesitamos estimarlos con los datos de la muestra. Si suponemos que existe n-pares de datos $(y_1,x_1),(y_2,x_2),...,(y_n,x_n)$, podemos estimar su valor, por medio del método de mínimos cuadrados.

El método de mínimos cuadrados, consiste en obtener la mejor línea que minimice la suma de cuadrados de los residuales o errores.

$$\textbf{(3.4.3)} \qquad \min \sum_{i=1}^{n} \varepsilon_i^{\ 2} = \min \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 \cdot x_i)^2$$

Al realizar el método de mínimos cuadrados, nuestros valores estimados de β_0 y β_1 quedan de la siguiente manera:

$$\textbf{(3.4.4)} \qquad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \cdot \bar{x}$$

$$\textbf{(3.4.5)} \qquad \hat{\beta}_1 = \frac{\sum_{i=1}^{n} y_i \cdot x_i - \left(\sum_{i=1}^{n} y_i\right) \cdot \left(\sum_{i=1}^{n} x_i\right)}{\sum_{i=1}^{n} x_i^{\ 2} - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}$$

donde

$$\textbf{(3.4.6)} \qquad \bar{x} = \sum_{i=1}^{n} \frac{x_i}{n}$$

$$\textbf{(3.4.7)} \qquad \bar{y} = \sum_{i=1}^{n} \frac{y_i}{n}$$

Por lo tanto, nuestro modelo ajustado queda de la forma:

$$\textbf{(3.4.8)} \qquad \bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \cdot \bar{x}$$

Otra forma más cómoda de ver el modelo, es:

$$\textbf{(3.4.9)} \qquad \hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$$

donde

$$\textbf{(3.4.10)}\, S_{xy} = \sum_{i=1}^{n} y_i \cdot x_i - \left(\sum_{i=1}^{n} y_i\right) \cdot \left(\sum_{i=1}^{n} x_i\right) = \sum y_i \cdot \left(x_i - \bar{x}\right)$$

$$\textbf{(3.4.11)} \quad S_{xx} = \sum_{i=1}^{n} x_i^{\,2} - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}$$

3.4.2. Prueba de Hipótesis para β_1

$H_o : \beta_1 = 0$; existe relación lineal entre "x" y "y"

$H_a : \beta_1 \neq 0$; No existe relación lineal entre "x" y "y"

A Ho, le llamaremos hipótesis nula y a Ha, le llamaremos hipótesis alternativa.

Lo que se busca con esta hipótesis, es rechazar Ho, para poder concluir que nuestro modelo tiene una relación lineal entre "x" y "y", también podemos decir que "x" tiene un valor para poder explicar a la variable "y".

Nuestro estadístico de prueba es:

$$\textbf{(3.4.1.1)} \quad \text{Valor t=} \frac{\hat{\beta}_1}{se(\hat{\beta}_1)} \ ,$$

donde

$$se(\hat{\beta}_1) = \sqrt{\frac{MSE}{S_{xx}}} \ , \qquad \text{MSE=Error Cuadrático Medio}$$

entonces, rechazamos Ho, si

$$\textbf{(3.4.1.2)} \quad \text{Valor t>} \ \chi^2_{\frac{\alpha}{2}(n-2)}$$

Si se rechaza Ho, concluimos que existe relación lineal con la variable "x" y la variable "y".

Si, no se rechaza Ho, concluimos que no hay relación lineal entre la variable “x” y la variable “y”.

3.5 Promedios Móviles

En series de tiempo existen métodos de suavizamiento que buscan eliminar lo efectos de variaciones que existen en las observaciones. Una técnica de suavizamiento es el llamado promedio móvil que consiste en transformar los datos de la serie original en una serie de promedios móviles que permite tener menos cambios drásticos y así poder obtener ciclos o tendencias sobre el tiempo.

De forma general podemos decir que el promedio móvil lo podemos calcular como la suma de los valores más recientes y dividirlo entre n.

$$\Pr omedioMóvil = \sum_{i=1}^{n} \frac{x_i}{n},$$

Donde, x_i: son los valores de la serie

n: es el número de valores que se van a promediar

Lo anterior se realiza para varios periodos hasta encontrar el error cuadrático medio (ECM) mayor que es el que se va a utilizar para pronosticar el siguiente período.