

## CAPÍTULO 5

### Proyecciones de las Fuentes de Empleo

#### **5.1. Características Generales del Modelo de Regresión Lineal Múltiple de las Fuentes de Empleo a Nivel Nacional.**

Para la proyección de las fuentes de empleo se usará el método de regresión lineal múltiple. Donde la variable dependiente  $Y$ , representará el número de fuentes de empleo dignas. Es decir, solo hemos tomado aquellas fuentes de empleo que ofrecen a la población económicamente activa un ingreso mayor al de un salario mínimo, excluyendo aquellas de menos de un salario mínimo y a los trabajadores si sueldo. Esta información, fue tomada de las Encuestas Nacionales de Empleo, estas encuestas específicas por empleo, presentan ventaja sobre los censos, ya que incorporan diferentes preguntas que permiten identificar mejor a las personas económicamente activas.

Para explicar el comportamiento de las fuentes de empleo se usarán tres variables explicativas, las cuales serán el PIB, la PEA y la Población Total de 12 años y más.

Los datos pertenecientes al PIB se obtuvieron del Sistema de Cuentas Nacionales de México (INEGI), para el periodo 1991-2003. Dichos datos se encuentran en precios constantes de 1993. La PEA y la población total de 12 años y más, fueron tomadas de la Encuesta Nacional de Empleo en el periodo 1991-2003, con excepción de los años 1992 y

1994, años en los cuales la encuesta no fue levantada, para estos años se usó el promedio aritmético entre los años contiguos<sup>1</sup>.

De esta manera el modelo de regresión lineal múltiple, quedaría de la siguiente manera:

$$Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + e_i$$

$$i = 1, \dots, N$$

Donde:

- Y<sub>i</sub>: observaciones muestrales de la variable endógena **fuentes de empleo digno**, cuyo comportamiento pretende explicar el modelo.
- X<sub>1i</sub>: observaciones muestrales de la variable exógena **PIB**, variable elegida como explicativa e influyente sobre la variable Y.
- X<sub>2i</sub>: observaciones muestrales de la variable exógena **PEA**, variable elegida como explicativa e influyente sobre la variable Y.
- X<sub>3i</sub>: observaciones muestrales de la variable exógena **Población de 12 años y más** variable elegida como explicativa e influyente sobre la variable Y.

---

<sup>1</sup> Ver anexo: Modelos de Regresión.

Después de la obtención de datos, estos se computaron en el programa StatView, de esta manera se pudo observar las características siguientes. Una  $R^2 = .988$ , esto quiere decir que las variables elegidas explican en un 98.8% el comportamiento de las fuentes de empleo, así se muestra que el modelo es bueno.

Por otra parte en el análisis de varianza (ANOVA) para la hipótesis nula de inexistencia de modelo o de coeficientes iguales a cero, se obtuvo un F-value de 252.829 con un P-value  $< .0001$ , con lo que se rechaza la hipótesis nula y se concluye que al menos una variable es explicativa y al menos un coeficiente es diferente a cero. De esta manera el valor para los coeficientes quedan de la siguiente manera<sup>2</sup>:

Fuentes de Empleo =  $-10720220.046 + .012 \text{ PIB} - .635 \text{ PEA} + .607 \text{ Pob. 12 años y más}$

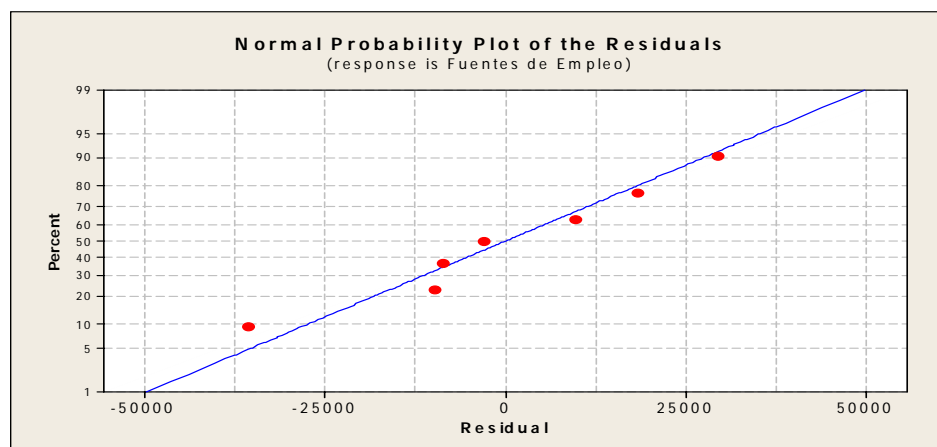
Ahora es importante saber si el modelo cumple con las hipótesis básicas expuestas en el apartado 3.4. Para esto se aplicó el contraste Durbin-Watson, que nos ayuda a descubrir la posible correlación de los residuos. El valor del estadístico de contraste  $d$ , fue igual a 2.15, por su parte, de acuerdo al número de observaciones y regresores a un nivel de significación  $\alpha = .05$ , los valores  $d_L$  y  $d_U$  son de .715 y 1.816 respectivamente, lo que nos lleva a aceptar la hipótesis nula, por lo tanto no se tiene autocorrelación en los residuos.

Es importante hacer un análisis para conocer el comportamiento de los residuos, ya que las hipótesis de un modelo de regresión indican que estos se deben distribuir de forma normal,

---

<sup>2</sup> Ver Anexo: Modelos de Regresión.

para saberlo, los datos fueron introducidos en el programa MINITAB, el cual arrojó la siguiente gráfica:



La cual nos permite concluir que la hipótesis se cumple, es decir, los residuos se distribuyen de forma normal.

### 5.2.1. Construcción del Modelo de Fuentes de Empleo a Nivel Nacional

Una vez que se ha obtenido el modelo de regresión se aplicó un estudio para saber si todas las variables eran representativas dentro de la regresión. Según el método “Todas las regresiones”, el mejor modelo es el que integra las tres variables propuestas ya que posee la mayor  $R^2$  y  $R^2$  ajustada, además su varianza es la menor, de los siete modelos construidos.

Para el método retrospectivo se procedió a calcular las F-parciales para cada  $X_i$  observándose que la F-parcial para  $X_1$  en un modelo con  $X_2$  y  $X_3$  es 14.868736, la F-parcial para  $X_2$  en un modelo con  $X_1$  y  $X_3$  es 9.048064 y por último la F-parcial para  $X_3$

en un modelo con  $X_2$  y  $X_1$  es 13.024881. Posteriormente la menor de ellas se comparó con una  $F_{(0.1)} = 3.23$  y como es mayor que esta última,  $X_2$  si explica las fuentes de empleo. Concluyendo que el mejor modelo es aquel que incluye las tres variables.

En el análisis del método prospectivo se observó que la variable explicativa con mayor correlación era el PIB, con un valor de .986, lo que nos permite concluir que el mejor modelo es aquel que solo incluye al PIB. Con el cálculo de las F-parciales obtenemos que para  $X_2$  en un modelo con  $X_1$  es de .015876 y para  $X_3$  en un modelo con  $X_1$  es de 2.2201. El mayor valor corresponde a  $X_3$  y este es menor a una  $F_{(0.1)} = 3,23$ , por lo que no es significativa para el modelo.

Al igual que en el método prospectivo, el análisis de Stepwise arrojó que la única variable representativa dentro del modelo era el PIB, lo cual se observó al hacer el análisis mediante StatView.

Como se puede observar, dos de los métodos sugieren el uso de un modelo con los tres regresores propuestos en un principio y por otra parte los otros dos sugieren un modelo lineal simple que solo contemple al PIB.

Debido a los resultados arrojados por los métodos de construcción de modelos, procedemos a la creación de un modelo de regresión lineal simple, donde la única variable explicativa va a ser el PIB. Al hacerlo obtenemos el siguiente modelo:

$$Y_i = \beta_0 + \beta_1 x_i + e_i$$

$$i = 1, \dots, N$$

Donde:

$Y_i$ : observaciones muestrales de la variable endógena **fuentes de empleo digno**, cuyo comportamiento pretende explicar el modelo.

$X_i$ : observaciones muestrales de la variable exógena **PIB** variable elegida como explicativa e influyente sobre la variable  $Y$ .

Las características de este modelo se muestran a continuación. Una  $R^2 = .971$ , esto quiere decir que las variables elegidas explican en un 97.1% el comportamiento de las fuentes de empleo, así se muestra que el modelo es bueno. Un F-value de 371.712 con un P-value  $< .0001$ , lo que nos permite rechazar la hipótesis nula y se concluye que la variable es explicativa y al su coeficiente es diferente a cero. De esta manera el valor para los coeficientes quedan de la siguiente manera<sup>3</sup>:

$$\text{Fuentes de Empleo} = - 6078792.878 + .021 \text{ PIB}$$

Ahora es importante saber si el modelo cumple con las hipótesis básicas expuestas en el apartado 3.4. Para esto se aplicó el contraste Durbin-Watson. El valor del estadístico de contraste  $d$ , fue igual a 1.389, por su parte, de acuerdo al número de observaciones y regresores a un nivel de significación  $\alpha = .05$ , los valores  $d_L$  y  $d_U$  son de 1.01 y 1.34

---

<sup>3</sup> Ver Anexo: Modelos de regresión.

respectivamente, lo que nos lleva a aceptar la hipótesis nula, por lo tanto no se tiene autocorrelación en los residuos.

Por su parte, al analizar el comportamiento de los residuos por medio del programa MINITAB, se observó que estos se comportaban de forma normal.

Solo falta verificar la existencia de homocedasticidad, para esto se realizó el contraste de Golfeld-Quandt, el cual se basa en que la distribución de las perturbaciones aleatorias es normal. Para este contraste fue necesario el cálculo del estadístico  $F_{GD}$ , el valor obtenido para el mismo fue de 3.0086, el cual se contrastó con una  $F_{(2,2)}$  a un nivel de significación de .05, siendo su valor de 19. Lo que nos permite aceptar la hipótesis nula, que es existencia de homocedasticidad.

Como se obtuvo que el modelo cumple con las hipótesis básicas. Ahora solo resta aplicar los resultados obtenidos en el capítulo anterior a los modelos de regresión múltiple y simple obtenidos, para conocer el número de fuentes de empleo dignas en la República Mexicana, en ambos modelos.

Para el modelo de regresión múltiple, al sustituir los valores proyectados del PIB, PEA y Población de 12 años y más se obtiene un pronóstico de las fuentes de empleo de 33,895,176.

Después se procedió a calcular un intervalo de confianza, para esta estimación, lo cual se hizo tomando una distribución t-student con los grados de libertad correspondientes, es

decir, 9 (13-(3+1)), con un nivel de confianza de .05. Cuyo valor en tablas es de 1.833. El intervalo viene dado por:

$$\hat{Y}_p = 1.83 \cdot 864157.1538$$

Quedándonos el siguiente intervalo de confianza (32,313,768 , 35,476,583).

Si deseáramos usar el modelo simple de fuentes de empleo contra PIB, el valor que alcanzarían las fuentes de empleo serían igual a 39,673,647.

## **5.2. Características Generales del Modelo de Regresión Lineal Múltiple de las Fuentes de Empleo a Nivel Estatal.**

De manera análoga a la creación del modelo descrito en el apartado 5.1 se desarrolló un modelo para la proyección de las fuentes de empleo a nivel estatal. En esta ocasión, la información, fue tomada de las Encuestas Nacionales de Empleo por Entidad Federativa, para el periodo 1996-2002.

Para explicar el comportamiento de las fuentes de empleo en el Estado de Puebla se usarán dos variables explicativas, las cuales serán el PIB, y la Población Total de 12 años y más.

Los datos pertenecientes al PIB se obtuvieron del Sistema de Cuentas Nacionales de México (INEGI), para el periodo 1996-2002. En este caso el PIB se encuentra expresado



como porcentaje, el cual, corresponde a la participación del Estado en el PIB. La población total de 12 años y más, fue tomada de la Encuesta Nacional de Empleo por Entidad Federativa<sup>4</sup>.

De esta manera el modelo de regresión lineal múltiple, quedaría de la siguiente manera:

$$Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + e_i$$
$$i = 1, \dots, N$$

Donde:

Y<sub>i</sub>: observaciones muestrales de la variable endógena **fuentes de empleo digno**, cuyo comportamiento pretende explicar el modelo.

X<sub>1i</sub>: observaciones muestrales de la variable exógena **PIB**, variable elegida como explicativa e influyente sobre la variable Y.

X<sub>2i</sub>: observaciones muestrales de la variable exógena **Población de 12 años y más** variable elegida como explicativa e influyente sobre la variable Y.

Después de la obtención de datos, estos se computaron en el programa StatView, de esta manera se pudo observar, de acuerdo con el análisis realizado por el mismo programa, que

---

<sup>4</sup> Ver anexo: Modelos de Regresión.

se tiene una  $R^2 = .995$ , esto quiere decir que las variables elegidas explican en un 99.5% el comportamiento de las fuentes de empleo, lo cual muestra que el modelo es bueno.

Por otra parte en el análisis de varianza (ANOVA) para la hipótesis nula de inexistencia de modelo o de coeficientes iguales a cero, se obtuvo un F-value de 207.951 con un P-value  $< .0001$ , con lo que se rechaza la hipótesis nula y se concluye que al menos una variable es explicativa y al menos un coeficiente es diferente a cero. De esta manera el valor para los coeficientes quedan de la siguiente manera<sup>5</sup>:

$$\text{Fuentes de Empleo} = -941024.538 - 420849.917 \text{ PIB} + 1.82 \text{ Pob. 12 años y más}$$

Ahora es importante saber si el modelo cumple con las hipótesis básicas expuestas en el apartado 3.4. Para esto se aplicó el contraste Durbin-Watson, que nos ayuda a descubrir la posible correlación de los residuos. El valor del estadístico de contraste  $d$ , fue igual a 1.751, por su parte, de acuerdo al número de observaciones y regresores a un nivel de significación  $\alpha = .05$ , los valores  $d_L$  y  $d_U$  son de .467 y 1.896 respectivamente. De esta manera al realizar el contraste notamos que el valor del estadístico cae en la zona de indiferencia, lo que nos lleva a aceptar la hipótesis nula de inexistencia de autocorrelación.

Por su parte, al analizar el comportamiento de los residuos por medio del programa MINITAB, se observó que estos se comportaban de forma normal.

---

<sup>5</sup> Ver Anexo: Modelos de regresión.

### **5.2.1. Construcción del Modelo de Fuentes de Empleo a Nivel Estatal**

Una vez que se ha obtenido el modelo de regresión se aplicó un estudio para saber si todas las variables eran representativas dentro de la regresión, para esto se aplicaron los métodos descritos en el apartado 3.4.1.

Para el método retrospectivo se procedió a calcular las F-parciales siendo la menor de 6.335289, la cual corresponde al PIB. Esta se compara con una F con parámetros (1,4), que tiene un valor de 7.709. Como  $6.335289 < 7.709$  se elimina la variable PIB y solo se conserva la Población de 12 años y más.

El método prospectivo mostró el mismo resultado ya que la variable que tiene mayor correlación con el empleo es la población de 12 años y más. Al calcular las F-parciales de la otra variable (PIB) con la población, se obtuvo que es de 6.335289 y al compararla con una F con parámetros (1,4), se muestra que solo la Población de 12 años y más es representativa y por lo tanto se elimina el PIB del modelo.

Al igual que en el método prospectivo, el análisis de Stepwise arrojó que la única variable representativa dentro del modelo era el PIB, lo cual se observó al hacer el análisis mediante StatView.

De esta manera se procede a crear un modelo de regresión lineal simple con las observaciones correspondientes a las fuentes de empleo y a la población de 12 años y más. Al hacerlo obtenemos el siguiente modelo:

$$Y_i = \beta_0 + \beta_1 x_i + e_i$$

$$i = 1, \dots, N$$

Donde:

Y<sub>i</sub>: observaciones muestrales de la variable endógena **fuentes de empleo digno**, cuyo comportamiento pretende explicar el modelo.

X<sub>i</sub>: observaciones muestrales de la variable exógena **Población de 12 años y más** variable elegida como explicativa e influyente sobre la variable Y.

Las características de este modelo se muestran a continuación. Una  $R^2 = .975$ , esto quiere decir que las variables elegidas explican en un 97.5% el comportamiento de las fuentes de empleo, así se muestra que el modelo es bueno. Un F-value de 198.174 con un P-value < .0001, lo que nos permite rechazar la hipótesis nula y se concluye que la variable es explicativa y al su coeficiente es diferente a cero. De esta manera el valor para los coeficientes quedan de la siguiente manera<sup>6</sup>:

$$\text{Fuentes de Empleo} = -926767.380 + 0.552 \text{ Pob. 12 años y más}$$

Revisando el contraste Durbin-Watson, obtenemos que el valor del estadístico de contraste  $d$ , fue igual a 1.58, por su parte, de acuerdo al número de observaciones y regresores a un

---

<sup>6</sup> Ver Anexo: Modelos de regresión.

nivel de significación  $\alpha = .05$ , los valores  $d_L$  y  $d_U$  son de .7 y 1.356 respectivamente. De esta manera al realizar el contraste notamos que el valor del estadístico  $d$  es mayor que el valor  $d_U$ , lo que nos lleva a aceptar la hipótesis nula, es decir, no existe autocorrelación.

Por su parte, al analizar el comportamiento de los residuos por medio del programa MINITAB, se observó que estos se comportaban de forma normal.

Solo falta verificar la existencia de homocedasticidad, para esto se realizó el contraste de Golfeld-Quandt, el cual se basa en que la distribución de las perturbaciones aleatorias es normal. Para este contraste fue necesario el cálculo del estadístico  $F_{GD}$ , el valor obtenido para el mismo fue de 0, lo que nos permite aceptar la hipótesis nula, que es existencia de homocedasticidad.

Al observar que el modelo cumple con las hipótesis básicas, solo resta aplicar los resultados obtenidos en el capítulo anterior para conocer el número de fuentes de empleo dignos en el Estado de Puebla. De esta manera y de acuerdo a los resultados obtenidos por los métodos demográficos anteriormente explicados para el año 2015 la población de 12 años para el Estado de Puebla será de 5,659,095. Al sustituir este valor en nuestro modelo de regresión lineal simple es 2,197, 053.

### **5.3 Estimación del PIB**

Ya que conocemos el valor de cada una de las variables que interviene en nuestros modelos, solo nos resta hacer el pronóstico para la variable PIB, para esto se empleará una regresión lineal simple con respecto al tiempo. Es importante mencionar, que para la proyección de esta variable se intentó hacer uso de métodos de series de tiempo, pero los resultados observados no fueron exitosos. De esta manera se procedió a graficar el comportamiento que ha presentado a través de los años, observándose un comportamiento lineal, por eso se ha procedido a proyectarlo por medio de esta herramienta.

El modelo es el siguiente:

$$Y_i = \beta_0 + \beta_1 x_i + e_i$$

$$i = 1, \dots, N$$

Donde:

$Y_i$ : observaciones muestrales de la variable endógena **PIB**, cuyo comportamiento pretende explicar el modelo.

$X_i$ : observaciones muestrales de la variable exógena **Periodos de tiempo** (años) variable elegida como explicativa e influyente sobre la variable Y.

Al introducir los datos correspondientes a la República Mexicana en el programa StatView observamos los siguientes resultados. Una  $R^2 = .934$ , es decir, los periodos de tiempo

explican en un 93.4% el comportamiento del PIB, y se observa que el modelo es bueno. En el análisis de varianza se obtuvo un F-value de 156.707 con un P-value < .0001, lo que nos permite rechazar la hipótesis nula y se concluye que la variable es explicativa y su coeficiente es diferente a cero. De esta manera el valor para los coeficientes quedan de la siguiente manera:

$$\text{PIB} = 1112256464.077 + 42657247 \text{ Periodos de tiempo}$$

Para conocer el valor del PIB en el año 2015, basta con sustituir en la variable Periodo de tiempo el valor 25, que es el valor que tomará la variable Periodo de tiempo en el año 2015.

El valor esperado para el PIB en el año 2015 a precios constantes de 1993 es 2,178,687,639.

Ahora hay que determinar el porcentaje del PIB que ha de obtener Puebla para el año 2015. Se procede de la misma manera que en el caso de la República Mexicana. Obteniendo los siguientes índices. Una  $R^2 = .975$ , es decir, los periodos de tiempo explican en un 97.5% el comportamiento del PIB, y se observa que el modelo es bueno. En el análisis de varianza se obtuvo un F-value de 200.818 con un P-value < .0001, lo que nos permite rechazar la hipótesis nula y se concluye que la variable es explicativa y su coeficiente es diferente a cero. De esta manera el valor para los coeficientes quedan de la siguiente manera:

$$\text{PIB} = 4.614 + .168 \text{ Periodos de tiempo}$$

Para conocer el porcentaje de participación que tendrá el Estado de Puebla en el PIB para el año 2015, basta con sustituir en la variable Periodo de tiempo el valor 25, que es valor que tomará la variable Periodo de tiempo en el año 2015. De esta manera, la participación esperada para el año en interés es 8.814.