

INSTITUTO POLITÉCNICO NACIONAL

ESCUELA SUPERIOR DE FÍSICA Y MATEMÁTICAS



Análisis estadístico de indicadores demográficos y suscriptores de televisión satelital.

Gabriela Teresa Pacheco Cabello

Métodos estadísticos univariados y multivariados
aplicados al análisis de datos.

México, D.F. 2006

INDICE

INTRODUCCION	4
1 PLANTEAMIENTO DEL PROBLEMA	5
2 JUSTIFICACION	6
3 OBJETIVO	7
4 RESUMEN	8
5 MARCO TEORICO	9
5.1 Regresión lineal	9
5.1.1 Prueba de Hipótesis	10
5.1.2 Análisis residual	12
5.2 Autocorrelación. Estadístico de Durbin – Watson	13
5.3 Prueba de falta de ajuste	14
5.4 Coeficiente de correlación.	15
5.5 Intervalos de confianza por la aproximación de Ruben.	16
5.6 Desventajas de la regresión.	17
5.6.1 Estrategia para evitar los problemas en la regresión.	17
5.7 Modelos de regresión múltiple.	18
5.7.1 Análisis residual para el modelo de regresión múltiple.	19
5.7.2 Prueba de hipótesis en la regresión múltiple.	19
5.7.3 Prueba de porciones del modelo de regresión múltiple.	20
5.8 Colinealidad.	21
5.9 Construcción de modelos.	22
5.9.1 Enfoque de mejores subconjuntos para la construcción de modelos.	22
5.9.2 Pasos para la construcción del modelo.	22
6 MARCO CONCEPTUAL.	24
6.1 Desarrollo del sector de televisión por satélite en México y Canadá.	24
6.2 Los distintos sistemas de televisión.	25
6.3 Historia de la televisión satelital en México.	27
6.4 Indicadores demográficos.	33

7	METODOLOGIA.	37
8	ANALISIS DE DATOS.	38
8.1	MODELO PILOTO (Área metropolitana).	38
8.1.1	Correlación.	38
8.1.2	Mejor subconjunto de regresión.	40
8.1.3	Modelo de regresión.	41
8.1.4	Análisis de residuos.	43
8.1.5	Puntos extremos.	45
8.2	MODELO COMPLETO.	46
8.2.1	Correlaciones incluyendo todos los datos	46
8.2.3	Modelo de regresión.	49
8.3	MODELO DE REGRESION LINEAL PARA PROVINCIA.	57
8.3.1	Correlación.	57
8.3.2	Mejor subconjunto.	58
8.3.3	Modelo de regresión.	59
9	CONCLUSIONES.	65
10	REFERENCIAS.	67

INTRODUCCION

El promedio de suscriptores mensuales a un sistema de televisión SKY varía con respecto a la delegación o municipio donde se instalen, por lo que se desea realizar un modelo que nos permita predecir el número de suscriptores mensuales en promedio que existirán en determinado municipio de la República Mexicana, con base en ciertos indicadores demográficos que se encuentren relacionados.

Al estudiar la relación que existe entre los diferentes indicadores y el promedio mensual de suscriptores se buscara el mejor modelo de regresión múltiple que relacione la mayor parte de los datos, con el fin de poder predecir que tanto afecta cada uno de los indicadores que se encuentren en el mejor modelo y el promedio total de suscriptores. Si se logra obtener un buen modelo, se considerará la opción de pronosticar el promedio de instalaciones mensuales, para los siguientes meses, usando los pronósticos de los indicadores demográficos.

1 PLANTEAMIENTO DEL PROBLEMA

Estudiar la relación entre algunos indicadores demográficos y el promedio mensual de suscriptores de televisión satelital SKY

2 JUSTIFICACION

La televisión juega un papel importante en nuestra sociedad, por lo que debe tomarse en cuenta el alcance que tiene en cada vivienda.

El poder predecir el número de suscriptores mensuales nos permitiría considerar con anticipación diversos factores que intervienen en las suscripciones, por ejemplo:

1. El inventario mínimo que deben existir para dar un buen servicio.
2. El total de ingresos que recibiría la empresa por este concepto.
3. Las supervisiones necesarias ha realizar, la cantidad de personal capacitado para dichas supervisiones, los honorarios a pagar por esta razón, y el tiempo invertido en ese trabajo.

3 OBJETIVO

Encontrar la relación existente entre ciertos indicadores demográficos y el promedio de suscriptores mensuales por municipio de un sistema de televisión satelital. De encontrarse un modelo de regresión múltiple adecuado se considerará la opción de pronosticar con base en estas variables.

4 RESUMEN

Existen muchos factores que intervienen para que una familia mexicana pueda suscribirse a un sistema de televisión por satélite, de hecho existe una enorme variación en el número de suscriptores mensuales por municipio, por lo cual se realizó un estudio sobre diversos indicadores demográficos con el fin de observar que relación guardan con el número de suscriptores.

Se creó un modelo de regresión múltiple con los indicadores demográficos que se encontraran más relacionados con el número de suscriptores, obteniendo un coeficiente de determinación alto, lo cual nos permite predecir como se comporta el número de suscriptores con un nivel de confianza del 90%.

Es importante indicar que el comportamiento se estudio a nivel municipal, sin embargo debemos tener en cuenta que aun dentro de cada municipio existen enormes diferencias en el numero de suscriptores, ya que nuestro país es rico en diferencias sociales y económicas.

5 MARCO TEORICO

5.1 Regresión lineal

En muchas situaciones una relación lineal puede ser útil para describir la dependencia de una variable Y sobre otra variable X, por ejemplo podemos escribir el modelo lineal de primer orden.

$$Y = \beta_0 + \beta_1 X + \varepsilon ,$$

en donde $\beta_0, \beta_1, \varepsilon$ son desconocidos y de hecho ε es difícil de encontrar pues cambia de una observación a otra. ε es un error aleatorio con media cero y varianza σ^2 .

Los términos β_0 y β_1 son fijos y podemos usar información conocida a través de una serie de observaciones para estimarlos, así que podemos escribir

$$\hat{Y} = b_0 + b_1 X ,$$

en donde $\hat{Y}$ denota el valor predicho de Y para un valor de X dado.

Si disponemos de un conjunto de n observaciones $(X_1, Y_1), \dots, (X_n, Y_n)$ podemos usar el procedimiento de mínimos cuadrados para estimar los valores de β_0 y β_1 mediante

$$b_1 = \frac{\sum_{i=1}^n X_i Y_i - \frac{1}{n} \left[\left(\sum_{i=1}^n X_i \right) \left(\sum_{i=1}^n Y_i \right) \right]}{\sum_{i=1}^n X_i^2 - \frac{1}{n} \left(\sum_{i=1}^n X_i \right)^2} = \frac{S_{XY}}{S_{XX}}$$
$$b_0 = \bar{Y} - b_1 \bar{X}$$

$S_{XY} = \sum_{i=1}^n X_i Y_i - \frac{1}{n} \left[\left(\sum_{i=1}^n X_i \right) \left(\sum_{i=1}^n Y_i \right) \right]$, se llama suma corregida de cuadrados de productos cruzados de X y Y.

$S_{XX} = \sum_{i=1}^n X_i^2 - \frac{1}{n} \left[\left(\sum_{i=1}^n X_i \right)^2 \right]$, se llama suma corregida de cuadrados de X.

Para probar la validez del modelo suele ser necesario obtener una estimación de la varianza σ^2 . Esto se logra calculando el error cuadrático medio, que es una estimación insesgada de σ^2 .

5.1.1 Prueba de Hipótesis

Para probar la suficiencia del modelo es necesario probar hipótesis en torno a los parámetros del modelo y la construcción de ciertos intervalos de confianza.

Para probar hipótesis respecto a la pendiente y la ordenada al origen del modelo de regresión, debemos hacer la suposición adicional de que la componente del error se distribuye normalmente.

Supongamos que deseamos probar la hipótesis

$$H_0 : \beta_0 = \beta_{0,0} \text{ vs}$$

$$H_1 : \beta_0 \neq \beta_{0,0}.$$

En este caso la estadística de prueba es

$$t_0 = \frac{b_0 - \beta_{0,0}}{\sqrt{MS_E \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{XX}} \right)}}$$

$$\text{donde } MSE = \frac{\sum_{i=1}^n (y_i - \hat{y})^2}{n-2}, \text{ es el cuadrado medio del error}$$

y se rechaza $H_0 : \beta_0 = \beta_{0,0}$ al nivel α si $|t_0| > t_{\frac{\alpha}{2}, n-2}$

Si se desea probar

$$H_0 : \beta_1 = \beta_{1,0} \text{ vs.}$$

$$H_1 : \beta_1 \neq \beta_{1,0}.$$

Entonces la estadística de prueba es

$$t_0 = \frac{b_1 - \beta_{1,0}}{\sqrt{\frac{MS_E}{S_{XX}}}}$$

y se rechaza $H_0 : \beta_1 = \beta_{1,0}$ al nivel α si $|t_0| > t_{\frac{\alpha}{2}, n-2}$

Un caso particular importante ocurre cuando se desea probar la significación de la regresión

$H_0 : \beta_I = 0$ vs.

$H_1 : \beta_I \neq 0$.

Pues el hecho de no rechazar $H_0 : \beta_I = 0$ es equivalente a concluir que no hay regresión lineal entre X y Y.

El procedimiento de prueba para $H_0 : \beta_I = 0$ suele arreglarse en una tabla de análisis de varianza como se muestra a continuación

FUENTE DE VARIACION	SUMA DE CUADRADOS	GRADOS DE LIBERTAD	MEDIA CUADRATICA	F_0
REGRESION	$SS_R = b_1 S_{XY}$	1	$MS_R = SS_R/1$	MS_R/MS_E
ERROR RESIDUAL	$SS_E = SS_T - SS_R$	n-2	$MS_E = SS_E/(n-2)$	
TOTAL	$SS_T = S_{YY}$	n-1		

Donde $S_{yy} = \sum_{i=1}^n (y_i - \bar{y})^2$

Entonces se rechaza $H_0 : \beta_I = 0$ al nivel α si $F_0 > F_{\alpha, 1, n-2}$

El ajuste de un modelo de regresión requiere de varias suposiciones para la estimación de parámetros, así como para las pruebas de hipótesis y la estimación de intervalos de confianza.

Debemos considerar revisar la validez de estas suposiciones y conducir los análisis para examinar la adecuación del modelo que se ha considerado en forma tentativa.

5.1.2 Análisis residual

Definimos los residuos del modelo de regresión lineal como

$$e_i = y_i - \hat{y}_i, \quad i = 1, 2, \dots, n.$$

El análisis de residuos es útil para confirmar la suposición de normalidad de los errores y en la determinación de si los términos adicionales en el modelo serían de utilidad.

Como una verificación aproximada de la normalidad, es posible estandarizar los residuos calculando

$$d_i = \frac{e_i}{\sqrt{MS_E}}, \quad i = 1, 2, \dots, n.$$

Si los errores son normales, entonces de acuerdo a la regla empírica, aproximadamente

68% de los residuos estandarizados deben caer en el intervalo $[-1, 1]$

95% de los residuos estandarizados deben caer en el intervalo $[-2, 2]$

99.7% de los residuos estandarizados deben caer en el intervalo $[-3, 3]$

A menudo resulta útil graficar los residuos

- 1) En secuencia de tiempo (si se conoce).
- 2) Contra los valores predichos por la ecuación de regresión.
- 3) Contra la variable independiente X.

En una situación normal los residuos seguirán un patrón aleatorio y estarán contenidos en una franja alrededor del 0.

Las situaciones en donde la varianza no es constante se identifican por cambios en la dispersión de los residuos a través del tiempo, o al cambiar los valores de las variables.

Cuando los residuos muestran patrones no aleatorios no lineales hay una indicación de insuficiencia del modelo, en este caso deben agregarse términos de orden superior al modelo, como

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \varepsilon.$$

5.2 Autocorrelación. Estadístico de Durbin – Watson

Una de las suposiciones básicas del modelo de regresión es que los errores son variables aleatorias independientes. Esta suposición frecuentemente se viola cuando los datos son recopilados en periodos sucesivos y adyacentes de tiempo. De esta forma es muy probable que residuos positivos sigan a otros residuos positivos y que residuos negativos sigan a otros negativos. Un patrón de residuos de este estilo se conoce como autocorrelación. En este caso, la validez de un modelo ajustado puede quedar en duda.

La forma más fácil de detectar la autocorrelación en un conjunto de datos es graficar los residuos en orden cronológico. Si existe un efecto de auto correlación positiva, se verán grupos de residuos con el mismo signo y se detectará un patrón evidente.

Además de las gráficas residuales, la autocorrelación puede medirse con el estadístico de Durbin_Watson, que mide la correlación entre cada residuo y el residuo inmediato anterior al periodo de interés. El estadístico de Durbin-Watson se define como

$$D = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=2}^n e_i^2}.$$

Cuando los residuos sucesivos tienen una autocorrelación positiva, el valor de D es cercano a cero. Si no existe correlación, el valor de D será cercano a 2. Si existe una correlación negativa, D será mayor que 2 y puede llegar incluso a su valor máximo de 4.

5.3 Prueba de falta de ajuste

Al ajustar un modelo de regresión, nos gustaría saber si el orden del modelo asumido en forma tentativa es correcto.

Una prueba de bondad de ajuste del modelo de regresión consiste en probar las hipótesis

H_0 : El modelo ajusta adecuadamente los datos vs.

H_1 : El modelo no ajusta los datos

Para realizar esta prueba es necesario tener observaciones repetidas al menos en un nivel de X.

La prueba implica dividir la suma de cuadrados del error residual en dos componentes:

La suma de cuadrados debida a la falta de ajuste del modelo SS_{FA} y la suma de cuadrados del error puro SS_{EP} .

Este procedimiento puede agregarse fácilmente en la tabla de análisis de varianza usada para probar la significación de la regresión en la siguiente forma.

FUENTE DE VARIACION	SUMA DE CUADRADOS	GRADOS DE LIBERTAD	MEDIA CUADRATICA	F_0
REGRESION	$SS_R = b_1 S_{XY}$	1	$MS_R = SS_R/1$	MS_R/MS_E
ERROR RESIDUAL	$SS_E = SS_T - SS_R$	$n-2$	$MS_E = SS_E/(n-2)$	
FALTA DE AJUSTE	SS_{FA}	$m-2$	$MS_{FA} = SS_{FA}/(m-2)$	MS_{FA}/MS_{EP}
ERROR PURO	SS_{EP}	$m-n$	$MS_{EP} = SS_{EP}/(n-m)$	
TOTAL	$SS_T = S_{YY}$	$n-1$		

La estadística de prueba para la falta de ajuste es

$$F_0 = \frac{MS_{FA}}{MS_{EP}}$$

y se rechaza H_0 : El modelo ajusta adecuadamente los datos, al nivel α si $F_0 > F_{\alpha, m-2, n-m}$.

5.4 Coeficiente de correlación.

Hasta ahora hemos supuesto que X es una variable matemática y que Y es una variable aleatoria.

Muchas situaciones de análisis de regresión implican casos en los que tanto X como Y son variables aleatorias.

El coeficiente de correlación de la muestra

$$r = \frac{S_{XY}}{\sqrt{S_{XX}S_{YY}}}$$

mide el grado de asociación lineal entre X y Y.

En el caso de una variable matemática X, r no tiene significado porque la magnitud de r depende del espaciamiento para X.

La cantidad

$$R^2 = \frac{SS_R}{S_{yy}} = r^2$$

se denomina coeficiente de determinación y se emplea a menudo para juzgar la suficiencia del modelo de regresión.

Con frecuencia nos referimos a R^2 como el grado de variabilidad en los datos explicado por el modelo de regresión.

La estadística R^2 debe usarse con precaución puesto que siempre es posible hacer a R^2 unitaria agregando suficientes términos al modelo.

5.5 Intervalos de confianza por la aproximación de Ruben.

Sea μ el punto crítico superior $\alpha/2$ de la distribución normal estándar.

Sean

$$r^* = \frac{r}{(1-r^2)^{1/2}}$$

$$a = 2N-3 \cdot \mu^2$$

$$b = r^*[(2N-3)(2N-5)]^{1/2}$$

$$c = (2N-5 \cdot \mu^2)(r^*)^2 - 2\mu^2$$

donde $\mu = Z_{\alpha/2} = 1.96$ para el 95% de confiabilidad y r el coeficiente de correlación.

Suponga ahora que y_1 & y_2 son las raíces de la ecuación $ay^2 - 2by + c = 0$, entonces los límites de confianza superior e inferior del $(1-\alpha)$ 100% para ρ son

$$\left[\frac{y_1}{(1+y_1^2)^{1/2}}, \frac{y_2}{(1+y_2^2)^{1/2}} \right],$$

Por lo tanto si el cero se encuentra en este intervalo, podemos concluir que la correlación no es significativa.

5.6 Desventajas de la regresión.

El análisis de regresión es de amplio uso y se presta a malas interpretaciones cuando se aplica en los negocios y en la economía.

Algunas dificultades involucradas en el análisis de regresión pueden ser:

- 1) Olvidar las suposiciones de la regresión por mínimos cuadrados.
- 2) No saber cómo evaluar las suposiciones de la regresión por mínimos cuadrados.
- 3) No conocer las alternativas de la regresión por mínimos cuadrados si se viola alguna suposición.
- 4) Usar un modelo de regresión sin conocimiento del tema.

5.6.1 Estrategia para evitar los problemas en la regresión.

- 1) Comenzar con un diagrama de dispersión para observar la posible relación entre X y Y.
- 2) Verificar que se cumplen las suposiciones del modelo de regresión antes de usar los resultados con fines de predicción.
- 3) Graficar los residuos en función de la variable independiente para verificar la adecuación del modelo y la suposición de homoscedasticidad (varianza constante).
- 4) Verificar la suposición de normalidad de los residuos.
- 5) Si los datos se obtuvieron en secuencia de tiempo graficar los residuos contra el orden y calcular el estadístico de Durbin-Watson para evaluar la auto correlación.
- 6) Si existe alguna violación a las suposiciones en los puntos 3) a 5) buscar algún modelo alternativo según indique la evaluación.
- 7) Si no existe violación a las suposiciones, se puede realizar inferencia con el modelo desarrollado así como predicción.

5.7 Modelos de regresión múltiple.

En un modelo de regresión múltiple, en general la variable dependiente o respuesta Y puede relacionarse con p variables independientes, en la forma

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon.$$

Los parámetros β_j representan el cambio esperado en la respuesta Y debido a un cambio unitario en la variable X_j , suponiendo que el resto de las variables independientes se mantienen constantes.

Se supone que el término de error en el modelo ε , satisface $E(\varepsilon) = 0$, $V(\varepsilon) = \sigma^2$ y que las ε_i son variables aleatorias no correlacionadas.

El método de mínimos cuadrados puede usarse para estimar los coeficientes del modelo de regresión. Para ello es necesario tener $n > p$ observaciones $(X_{11}, X_{21}, \dots, X_{p1}, Y_1), \dots, (X_{1n}, X_{2n}, \dots, X_{pn}, Y_n)$.

El modelo ajustado puede usarse entonces con propósitos de predicción

$$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2 + \dots + b_p X_p.$$

En la regresión múltiple, debido a que hay al menos dos variables explicativas, el coeficiente de determinación múltiple r^2 representa la proporción de la variación en el Y explicada por el conjunto de variables independientes seleccionadas.

Cuando se trata de regresión múltiple algunos autores sugieren calcular un valor r^2 ajustado que refleje tanto el número de variables explicativas como el tamaño de la muestra. Esto resulta necesario y útil cuando se comparan dos o más modelos de regresión múltiple que predicen la misma variable dependiente en términos de diferente número de variables explicativas.

5.7.1 Análisis residual para el modelo de regresión múltiple.

Recordemos que los residuos del modelo de regresión se definen como

$$e_i = y_i - \hat{y}_i, \quad i = 1, 2, \dots, n.$$

Al analizar un modelo de regresión múltiple, las gráficas residuales son útiles para evaluar la validez del modelo. Podemos pensar en que al menos debe graficarse

- 1) Los residuos vs. los valores predichos por el modelo.
- 2) Los residuos en función de cada variable predictora.
- 3) Los residuos vs. el tiempo (si se conoce la secuencia).

5.7.2 Prueba de hipótesis en la regresión múltiple.

Al igual que en regresión lineal simple, es deseable establecer la significancia del modelo de regresión múltiple. Dado que en un modelo de regresión múltiple se tiene más de una variable explicativa las hipótesis adecuadas son

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0 \quad \text{vs.}$$

$$H_1 : \beta_j \neq 0 \text{ para algún } j.$$

El rechazo de H_0 implica que al menos una de las variables independientes $X_1, X_2, \dots, X_p$ contribuye significativamente al modelo.

El procedimiento de prueba es una generalización del utilizado en la regresión lineal simple, de acuerdo a la siguiente tabla de análisis de varianza

FUENTE DE VARIACION	SUMA DE CUADRADOS	GRADOS DE LIBERTAD	MEDIA CUADRÁTICA	F_0
REGRESION	SS_R	P	$MS_R = SS_R/p$	MS_R/MS_E
ERROR RESIDUAL	SS_E	$n-p-1$	$MS_E = SS_E/(n-p-1)$	
TOTAL	SS_T	$n-1$		

Se rechaza $H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0$ al nivel α si $F_0 > F_{\alpha, p, n-p-1}$.

Con frecuencia será conveniente realizar pruebas de hipótesis respecto a los coeficientes individuales de la regresión.

Las hipótesis para probar la significancia de cualquier coeficiente de regresión individual digamos β_j , son

$$H_0 : \beta_j = 0 \quad \text{vs.}$$

$$H_1 : \beta_j \neq 0.$$

Si no se rechaza $H_0 : \beta_j = 0$, entonces la variable X_j puede ser eliminada del modelo. La estadística de prueba adecuada para esta hipótesis es

$$t_0 = \frac{b_j}{S_{b_j}}.$$

Se rechaza $H_0 : \beta_j = 0$, si $|t_0| > t_{\frac{\alpha}{2}, n-p-1}$.

5.7.3 Prueba de porciones del modelo de regresión múltiple.

Al desarrollar un modelo de regresión múltiple, el objetivo es usar sólo aquellas variables explicativas que son útiles para predecir el valor de la variable dependiente. Si una variable no es útil para el pronóstico, debe eliminarse del modelo y entonces se puede usar un modelo con menos variables explicativas.

Un método para determinar la contribución de una variable explicativa, se denomina prueba F parcial. Este criterio determina la contribución a la suma de cuadrados de la regresión que hace cada variable explicativa una vez que se han incluido al modelo todas las demás variables explicativas en el modelo. La nueva variable explicativa se incluye en el modelo sólo si lo mejora de una manera considerable.

Para determinar la contribución de la variable j al modelo, suponiendo que el resto de las variables estén incluidas se tiene

$$SS_R(X_j | \text{todas las variables excepto } X_j) =$$

$$SS_R(\text{todas las variables incluso } X_j) - SS_R(\text{todas las variables excepto } X_j).$$

La prueba de hipótesis adecuada para determinar la contribución de X_j al modelo es

H_0 : La variable X_j no mejora significativamente el modelo (incluidas el resto de las variables en el modelo) vs

H_1 : La variable mejora significativamente el modelo (incluidas el resto de las variables en el modelo)

La estadística de prueba adecuada es entonces

$$F_0 = \frac{SS_R(X_j | \text{todas las variables excepto } X_j)}{MS_E}$$

Se rechaza H_0 : La variable X_j no mejora significativamente el modelo, al nivel α si $F_0 > F_{\alpha, 1, n-p-1}$.

5.8 Colinealidad.

Un problema importante en la regresión múltiple es la posible colinealidad de las variables explicativas, debida a que algunas variables tienen una alta correlación entre sí.

Un método para medir la colinealidad utiliza el factor inflacionario de varianza VIF definido como

$$VIF_j = \frac{1}{1 - R_j^2}$$

donde R_j^2 es el coeficiente de determinación múltiple de la variable explicativa X_j con todas las demás variables. Es decir es el coeficiente de determinación del modelo:

$$X_j = a_0 + a_1 X_1 + a_2 X_2 + \dots + a_p X_p$$

$$X_1, \dots, X_p \neq X_j$$

Si un conjunto de variables explicativas es no correlacionado, entonces VIF_j será igual a 1. Si el conjunto tiene una correlación alta VIF_j puede exceder a 5.

5.9 Construcción de modelos.

5.9.1 Enfoque de mejores subconjuntos para la construcción de modelos.

Este enfoque evalúa ya sea todos los modelos de regresión posibles para un conjunto de variables independientes o los mejores subconjuntos de modelos para un número dado de variables independientes, de acuerdo a dos criterios: la r^2 ajustada y el estadístico C_p .

El estadístico C_p mide las diferencias de un modelo de regresión ajustado con respecto al “modelo verdadero”, junto con un error aleatorio. El estadístico se define como

$$C_p = \frac{(1 - R_p^2)(n - T)}{1 - R_T^2} - (n - 2(p + 1)).$$

p - número de variables independientes incluidas en el modelo de regresión.

T - número total de parámetros (incluso la ordenada al origen) que se estimarán en el modelo de regresión completo.

R_p^2 – coeficiente de determinación múltiple para un modelo de regresión con p variables independientes.

R_T^2 – coeficiente de determinación múltiple para un modelo de regresión completo que contiene los T parámetros estimados.

Cuando un modelo de regresión con p variables independientes contiene sólo diferencias aleatorias con un “modelo verdadero”, el valor promedio de C_p es $p+1$, el número de parámetros. Entonces al evaluar muchos modelos de regresión alternativos, la meta es encontrar modelos cuyo C_p sea cercano o menor que $p+1$.

5.9.2 Pasos para la construcción del modelo.

1. Elegir un conjunto de variables independientes para considerar su inclusión en el modelo de regresión.
2. Ajustar el modelo de regresión completo que incluya todas las variables independientes que se van a analizar de manera que se pueda determinar VIF para cada variable.
3. Determinar si alguna variable independiente tiene $VIF > 5$.
4. Existen 3 resultados posibles:
 - a) Ninguna variable independiente tiene $VIF > 5$. Proceder al paso 5.

- b) Una variable independiente tiene $VIF > 5$. Eliminar dicha variable y proceder al paso 5.
 - c) Más de una variable tiene $VIF > 5$. Eliminar la variable independiente con el VIF más alto y regresar al paso 2.
-
5. Realizar una regresión de los mejores subconjuntos o de todos los subconjuntos con el resto de las variables independientes.
 6. Enumerar todos los modelos con $C_p \leq p+1$.
 7. Entre los modelos enumerados en el paso 6, elegir el mejor modelo.
 8. Realizar un análisis completo del modelo elegido que incluya el análisis residual
 9. De acuerdo con los resultados del análisis residual agregar términos curvilíneos si es necesario y analizar de nuevo los datos.
 10. Usar el modelo seleccionado para el pronóstico.

6 MARCO CONCEPTUAL.

La televisión –como su propio nombre indica– es “ver de lejos”, es decir, llevar ante los ojos de un público de espectadores cosas que puedan ver en cualquier sitio, desde cualquier lugar y distancia.

Los medios en general se usan con un fin específico, la audiencia selecciona el medio y los contenidos según sus necesidades, hay quienes prefieren informarse y otros ven series porque les gusta identificarse con la realidad en ellas representada. Así, los hombres de clase social alta suelen ver programas de actualidad, informativos, deportes y concursos; mientras que las mujeres con un nivel de educación bajo, de una clase social inferior y que no trabajan, ven más televisión que los anteriores y ven series, películas y programas infantiles. La tercera edad ve todo tipo de programas, y junto con las personas de menor nivel intelectual, son los que más tiempo ven la televisión, porque es su forma de entretenimiento y es su medio preferido de ocio. Gran parte de los adultos utilizan la televisión a falta de compañía para no sentirse tan solos.¹

6.1 Desarrollo del sector de televisión por satélite en México y Canadá.

Los primeros pasos de DTH en México comenzaron en marzo de 1994 cuando Televisa anunció su proyecto de iniciar el servicio de televisión vía satélite DTH. Sin embargo, su iniciación se postergó debido a que el satélite PAS III no pudo ser colocado en órbita en los tiempos proyectados. En 1995, Televisa, el grupo estadounidense News Corporation, la empresa brasileña O’Globo y la compañía Tele Communications Internacional Incorporated (TCI) firmaron un convenio para prestar de manera conjunta el servicio de DTH en Latinoamérica. El servicio proporcionado fue a nivel latinoamericano. El nombre que se le asignó a este proyecto fue de Sky Entertainment Services, sin embargo para fines publicitarios se le conoce simplemente como Sky. Al igual que DirecTV, Sky comenzó a funcionar durante el segundo semestre de 1996.

En 1995, las empresas Hughes Communications de Estados Unidos, la Organización Cisneros de Venezuela, la Televisión Abril de nacionalidad brasileña y Multivisión de México (MVS), dieron a conocer la creación del consorcio Galaxy Latin America. Se trató de una alianza estratégica para lanzar al mercado DirecTV, un servicio de televisión vía satélite DTH con cobertura latinoamericana, con un servicio diseñado especialmente para el público latinoamericano con programación en español y portugués. Dicha inversión fue el resultado de la expansión hacia Latinoamérica de otra empresa filial de DTH, Hughes Communications en Estados Unidos a finales de 1993. Para comercializar el nuevo sistema, Hughes creó la filial DirecTV que mediante una inversión de 600 millones

¹ En referencia a: Lorenzo Vilches “usos de la televisión”: diversión o entretenimiento, utilidad social e información. <http://www.rppnet.com.ar/efectostv.htm>

de dólares proporciona el servicio DTH en Estados Unidos.[9] El inicio de operaciones se dio en junio de 1994. En México el comienzo de DirecTV se llevó a cabo el segundo semestre de 1996.

A pesar de las perspectivas tan positivas actualmente las dos empresas están en busca de una fusión, y esto se dificulta por el monopolio que generarían. Las inversiones en infraestructura de DTH en México han sido significativas, pero su recuperación se dificulta, entre otras cosas, por la caída en la demanda originando que se vieran obligadas a reducir tarifas. Sky esperaba tener utilidades hasta el 2004 y para DirecTV la situación es parecida. La fusión se está meditando, ya que se considera que el mercado de DTH en México no es tan amplio como para soportar dos grandes competidores y que ambos alcancen la rentabilidad.

6.2 Los distintos sistemas de televisión.

Los programas de televisión que gozamos provienen de muchas fuentes: Las estaciones locales o abiertas proporcionan información en noticias y acontecimientos locales; las redes nacionales proporcionan las noticias comprensivas, nacionales e internacionales; los estudios de películas de estudios como en Hollywood ofrecen al mercado películas que eventualmente son mostradas por la televisión por cable o por las estaciones locales, o redes nacionales.

Televisión abierta

La televisión que no está codificada por un productor, se considera abierta o gratuita. Las estaciones locales que transmiten en canales uniformes de un transmisor cercano pueden recibir gratuitamente la programación usando antenas de orejas de conejo o una antena al aire libre conectada a la espalda del receptor de televisión. A los creadores de estos programas, poseedores de los derechos de autor, se les pagan los derechos de transmisión por medio de los patrocinadores comerciales. A su vez, éstos obtienen ganancias cuando los consumidores compran sus productos.

El mercado de la televisión comercial abierta con cobertura nacional se compone de seis señales nacionales, repartidas entre dos empresas: Televisa y TV Azteca. Mientras la primera cuenta con cuatro canales y detenta los primeros lugares del rating a través de su canal **2**; TV Azteca maneja dos redes: la del **7** y la del **13**.

- **TV Azteca**

Se ha caracterizado por ser una empresa agresiva en todos los sentidos: buscó asociarse a NBC (acuerdo que finalmente no se concretó y que mantuvo en litigio a las dos compañías), hizo alianzas con Telemundo en materia de producciones y artistas, cambió su programación dándole un tinte amarillista y violento, produjo telenovelas con argumentos muy controvertidos aunque de gran aceptación

- **Televisa**

Sus canales abiertos ofrecen cerca de 650 horas semanales de programación a un auditorio nacional con una cobertura del 97% de la población en el caso del canal 2; del 90% con el 4; 79% con el 5 y 60% con el 9, lo que sumado da un total de 50 millones de tele hogares (Robina, en Crovi, 1996).

- **Televisión de paga**

La introducción de la televisión por cable al mercado, posibilitó que los usuarios gozaran de retransmisiones de programas de estaciones lejanas que no se podían recibir con antenas de orejas de conejo ni antenas al aire libre. En años recientes, la introducción de la tecnología de codificación o la criptografía ha hecho posible la DTH. La tecnología de codificación ha permitido que las empresas proveedoras den un servicio similar al de televisión por cable, en el cual estas empresas conectan a sus suscriptores permitiendo que la señal sea decodificada por su receptor.

- **DTH**

Se entiende por servicios de difusión Directa al Hogar aquellas “...señales unidireccionales encriptadas de radiocomunicación de video o de video/ audio que se transmiten por satélites con licencia de las partes, para recepción directa por parte de suscriptores mediante remuneración periódica” (D.O.F. 1995. Reglamento de Comunicación Vía Satélite). Las cadenas de televisión y las empresas productoras de programas, rutinariamente distribuyen su programación vía satélite. Una vez que llegan a su punto de distribución, estas son grabadas, e incluidas dentro de la agenda de las cadenas de televisión, para ser luego enviadas de vuelta hacia satélites en intervalos apropiados a los horarios de las zonas donde hay estaciones afiliadas.

El sistema de DTH permite la transmisión de señales de TV a los hogares directamente desde un satélite. Este sistema opera con pequeñas antenas semiparabólicas con diámetros de entre 45 y 90 centímetros —en contraste con las de otros servicios que requieren parabólicas de hasta 5 metros de diámetro— y además tienen la capacidad para transmitir más de 150 canales.

Alguna de esta programación es gratuita, pero existen muchos servicios de satélite mixtos, lo que significa que están configurados y sólo disponibles con suscripción pagada, como es Sky o Direct TV. Para tener acceso a estos sistemas se debe realizar la contratación con una de las dos empresas que proveen el servicio en México Direct TV o Sky. Estas empresas envían al usuario un número de serie único para que su receptor de satélite se pueda comunicar con la señal del mismo.

Esto libera o abre la señal para que pueda aparecer a través del televisor del hogar que posea la antena.

Los servicios de satélite para el hogar, como el de Sky y el de Direct-TV, tienen sus propios sistemas y receptores satelitales. Al menos uno de ellos tiene una capacidad de 50 canales digitales y simultáneos de TV².

6.3 Historia de la televisión satelital en México.

1965

PAJARO MADRUGADOR: En abril es colocado en órbita el Pájaro Madrugador (Early Bird), primer satélite comercial de comunicaciones, México aún no tiene infraestructura propia para la comunicación vía satélite, por lo que las imágenes provenientes de Europa llegan primero a Estados Unidos y luego se envían a nuestro país por microondas; asimismo, las señales que tienen su origen en México viajan primero vías microondas al país vecino en donde son "subidas" al satélite.

1966

INGRESO A INTELSAT: México ingresa a la Organización Internacional de Comunicaciones por Satélite (INTELSAT), con lo cual obtiene el derecho de utilizar los artefactos espaciales propiedad de ese consorcio.

1968

INFRAESTRUCTURA DE TELECOMUNICACIONES: Se concluyen los trabajos de la Red Nacional de Telecomunicaciones iniciados desde 1963. El sistema incluye la Red Federal de Microondas, la Estación Terrestre para Comunicaciones Espaciales de Tulancingo (para envío y recepción de señales por satélite) conectada a los satélites INTELSAT II y III, y la Torre de Telecomunicaciones en la Ciudad de México.

1976

TELEVISA INTERNACIONAL: El consorcio Televisa adquiere el 20 por ciento de las acciones de la empresa Spanish International Communication Corporation (SICC) de los Estados Unidos y funda el sistema Univisión, que incluye estaciones en Los Ángeles, Nueva York y San Antonio, gracias al cual exporta programas vía satélite y microondas hacia ese país.

1980

² http://revista.amec.com.mx/num_8_2004/Jasa_Graciela_Frias_Martha2.htm

TELEVISA POR SATELITE: En mayo Televisa contrata, con autorización de la Secretaría de Comunicaciones y Transportes, los servicios del satélite estadounidense Westar III con lo cual adquiere la posibilidad de cubrir el territorio mexicano, pues la huella del artefacto cubre nuestro país, y de transmitir directamente a Estados Unidos 19 horas diarias de programación a través de la cadena Spanish International Network (SIN) constituida por 100 estaciones afiliadas y de la que Televisa ha adquirido el 75 por ciento de acciones.

SATELITE MEXICANO EN PROYETO: El 10 de octubre la Secretaría de Comunicaciones y Transportes da a conocer que para 1985 México contará con su propio satélite que llevará el nombre de Iluicahua ("Señor del cielo" en lengua náhuatl).

1981

MEXICO ALQUILA SATELITE: Debido a que México está ampliando su infraestructura de comunicación por satélite, pero aun no cuenta con uno propio, el gobierno de la república solicita al consorcio INTELSAT que modifique la órbita de uno de sus artefactos (el INTELSAT IV AF3) para que pueda "bañar" el territorio mexicano con sus señales. El alquiler de tres transpondedores en este satélite permite a México efectuar por esa vía una parte sustancial de sus telecomunicaciones internas las cuales realizaba antes por microondas. En 1984, al concluir la vida útil de este satélite, México renta los servicios del INTELSAT V F8 para efectuar sus comunicaciones domésticas.

1982

ACTIVIDAD ESTRATEGICA: En diciembre, el artículo 28 constitucional es modificado por iniciativa del nuevo presidente, Miguel de la Madrid, con el fin de declarar a la comunicación por satélite como "una actividad estratégica a cargo exclusivo del Estado".

1983

SATELITES MORELOS: En junio la Secretaría de Comunicaciones y Transportes informa que el sistema mexicano de satélites llevará el nombre de Morelos (ya no Iluicahua, como se le pretendía llamar en el sexenio anterior) y que estará constituido por dos artefactos que serán colocados en órbita en 1985.

1990

NUEVO PROYECTO SATELITAL: Ante la proximidad de la finalización del periodo de vida útil del satélite Morelos I, prevista para 1994, y la inminente saturación del Morelos II, cuyo lapso de utilización llegará a su término en 1998, el gobierno de la república da a conocer, el 28 de julio, su proyecto de ubicar en el espacio un nuevo sistema de satélites, también integrado por dos artefactos, que llevará el nombre de Solidaridad y cuya vida útil será de 14 años. Se anuncia que el primero de los satélites será enviado al espacio en 1993.

1992

TELESECUNDARIA POR SATELITE: La Unidad de Televisión Educativa (UTE) inicia la transmisión de la telesecundaria y otros programas educativos a través del sistema de satélites Morelos. En 1993 pone en marcha su proyecto de transmitir tele conferencias a diversas instituciones de educación en la república.

1993

SOLIDARIDAD I: El 19 de noviembre es lanzado al espacio desde Guyana Francesa, a bordo de un cohete Ariane, el nuevo satélite mexicano Solidaridad I, construido, al igual que sus antecesores, los Morelos, por la empresa estadounidense Hughes.

1994

DTH: En marzo, Televisa anuncia su proyecto de iniciar el servicio de televisión vía satélite directa al hogar, conocido como Direct to Home o DTH. Se trata de un nuevo sistema que permite la transmisión de señales de TV a los hogares directamente desde un satélite. La diferencia entre el sistema DTH y otros servicios de TV directa al hogar vía satélite es que permite transmitir audio y video digitales, es decir, el sonido tiene la calidad de un compact disk, y la imagen una definición mucho mayor que la TV normal. Además, el nuevo sistema posee la ventaja de operar con pequeñas antenas semiparabólicas con diámetros de entre 45 y 90 centímetros --en contraste con las de otros servicios que requieren parabólicas de hasta 5 metros de diámetro-- y la de tener capacidad para transmitir más de 150 canales.

SOLIDARIDAD II: El 7 de octubre de 1994, el satélite Solidaridad II es colocado en órbita. Construido también por la empresa Hughes Communications es enviado al espacio, desde Guyana Francesa, a bordo de un cohete de la compañía Ariane.

1995

REFORMA CONSTITUCIONAL: En enero, apenas en el segundo mes de su gestión, el gobierno zedillista decide reformar el artículo 28 constitucional con el fin de que la comunicación vía satélite deje de ser considerada legalmente como "una actividad estratégica exclusiva del Estado" --como había sido definida por el gobierno de de la Madrid en diciembre de 1982-- y pase a ser una "actividad prioritaria" en donde se permite al capital privado participar en la propiedad y operación de esos artefactos espaciales. La iniciativa presidencial es aprobada por el Congreso de la Unión con el apoyo de los partidos PRI y PAN.

LEY DE TELECOMUNICACIONES: En mayo, el Congreso de la Unión aprueba, nuevamente con el voto favorable del PRI y del PAN, una Ley Federal de Telecomunicaciones en la que se establecen las condiciones para que el capital privado participe en la comunicación por satélite, actividad que había sido hasta ese momento exclusiva del Estado. La nueva ley establece: 1) que el gobierno federal, a través de la SCT, podrá concesionar a empresas privadas la ocupación y la explotación de posiciones orbitales asignadas a México; 2) La SCT podrá otorgar concesiones para la explotación de bandas de frecuencias, a través de las cuales se puedan transmitir señales provenientes de satélites extranjeros que cubran el territorio nacional, "siempre y cuando se tengan firmados tratados en la materia con el país (de cuya nacionalidad sea el satélite) y dichos tratados contemplen la reciprocidad para los satélites mexicanos"; 3) en cualquiera de los dos casos anteriores, la inversión extranjera en las empresas concesionarias no podrá ser superior al 49 por ciento.

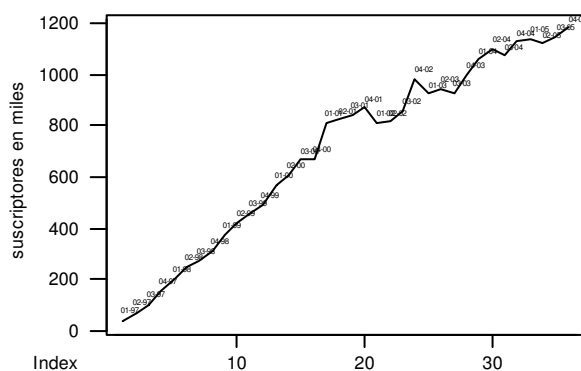
La Ley Federal de Telecomunicaciones fue aprobada por el Congreso de La unión el 18 de mayo de 1995 y publicada en el Diario Oficial de la Federación el 7 de junio del mismo año.

DIRECTV: El 9 de marzo las empresas Hughes Communications, de Estados Unidos, Organización Cisneros, de Venezuela, Televisión Abril, de nacionalidad brasileña, y MVS Multivisión, de México, dan a conocer la creación del consorcio Galaxy Latin America. Se trata de una alianza estratégica para lanzar al mercado DirecTV, un servicio de televisión vía satélite "directo al hogar" (DTH) con cobertura latinoamericana.

Este proyecto es resultado de la expansión hacia Latinoamérica de otro que, con el mismo nombre, inició en Estados Unidos la empresa Hughes Communications a finales de 1993. El 17 de diciembre de ese año Hughes colocó en el espacio un satélite --el Galaxy 4 H-- con la finalidad de prestar el servicio de "TV directa al hogar" o DTH. Para comercializar el nuevo sistema, Hughes crea la filial DirecTV que mediante una inversión de 600 millones de dólares proporciona el servicio DTH en Estados Unidos. El inicio de operaciones se produce en junio de 1994.

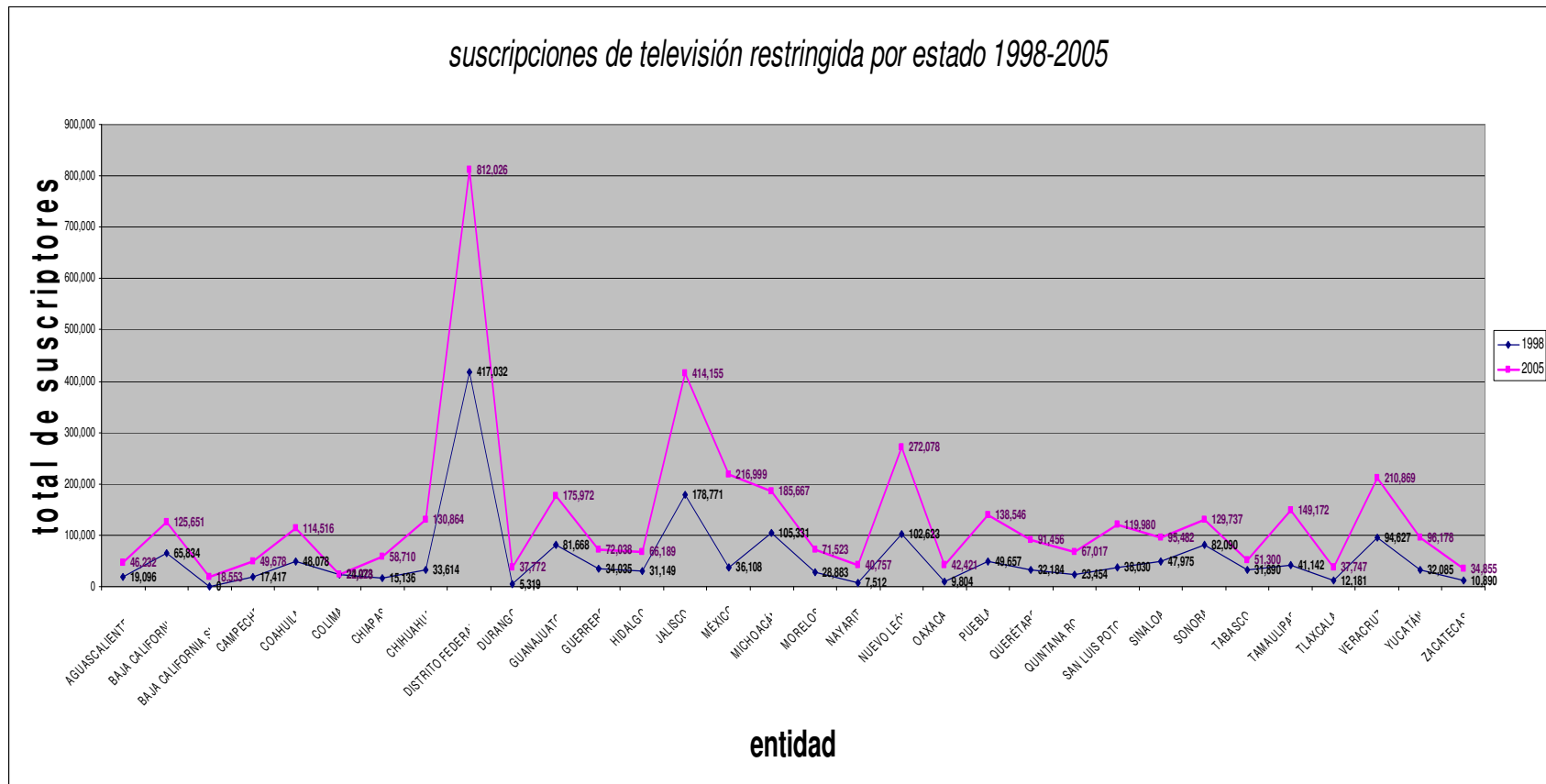
SKY: Ante la solidez que mostrada por el proyecto DirecTV, el consorcio mexicano Televisa, el conglomerado estadounidense News Corporation, la empresa brasileña O'Globo y la compañía Tele Communications Internacional Inc. (TCI), la operadora de sistemas de cable más grande de Estados Unidos, deciden unir esfuerzos y firman el 20 de noviembre de 1995 un convenio para prestar de manera conjunta el servicio de televisión directa vía satélite, o DTH, en Latinoamérica. En esta sociedad, Televisa, News Corporation y O'Globo aportan cada una el 30 por ciento de la inversión, mientras que TCI participa con el 10 por ciento restante. El servicio se proporcionará a nivel latinoamericano por medio de los satélites Panamsat, aunque en México, Televisa decide hacerlo inicialmente a través del sistema Solidaridad. El nombre que se asigna a este nuevo proyecto es el de Sky Entertainment Services, sin embargo para fines publicitarios se le conoce simplemente como Sky. Al igual que DirecTV se proyecta que Sky comience a funcionar en el segundo semestre de 1996.

La siguiente grafica nos muestra el número de suscriptores en miles del sistema SKY de Enero de 1997 a Diciembre del 2005, según información de la COMISION FEDERAL DE TELECOMUNICACIONES (**COFETEL**)³



³ http://www.COFETelvision_y_audio_restringido

Por ultimo mostramos una grafica, donde se observa el número de suscriptores por estado que existían en 1998 junto con el número de suscriptores que existían en diciembre del 2005, en la Republica Mexicana, dicha grafica nos muestra que en realidad existe una gran variación en el total de suscriptores por estado.



6.4 Indicadores demográficos.

Los indicadores demográficos varían según cada municipio debido a que cada uno cuenta con una administración diferente, por lo que nuestro análisis se realizará por municipios.

Se tomaran en cuenta distintos indicadores demográficos, los cuales fueron obtenidos por medio de las siguientes fuentes:

- INEGI : Instituto Nacional de Estadística, Geografía e Informática
XII Censo General de Población y Vivienda 2000
- CONAPO: Consejo Nacional de la Población. Fuente: estimaciones de con base en la muestra del diez por ciento del XII Censo General de Población y Vivienda 2000.
- SKY: Sistema de televisión satelital, información proporcionada
Total de suscriptores de diciembre del 2005 al 15 de abril del 2006 por municipio.

Las variables que se toman en cuenta en este análisis estadístico son las siguientes:

1. Índice de nivel de escolaridad

Descripción: Indicador de desarrollo humano propuesto para medir los logros en cuanto a la adquisición de conocimientos.

Formula: $IE = \{2(IA) + IAE\} / 3$

Donde:

IE = Índice de nivel de escolaridad

IA = Proporción de las personas de 15 años o mas alfabetos

IAE = Proporción de las personas de 6 a 24 años que van a la escuela

2. Índice de PIB per capita

Descripción: Indicador de desarrollo humano propuesto para medir los logros en cuanto al acceso a los recursos que permiten disfrutar de un nivel de vida digno y decoroso.

Formula: $IPIB = \{ \ln(PPC) - \ln(100) \} / \{ \ln(40000) - \ln(100) \}$

Donde:

IPIB = Índice del PIB per capita

PPC = PIB per capita ajustado al poder adquisitivo del dólar en Estados Unidos

100 = Valor mínimo del PPC registrado recientemente en el mundo

40000 = Valor máxima del PPC registrado recientemente en el mundo.

3. Índice de desarrollo Humano (IDH)

Descripción: Indicador que mide el logro medio de un país o región en tres dimensiones del desarrollo humano:

1. La capacidad de gozar de una vida larga y saludable, medida a través de la esperanza de vida al nacer o, en su caso, del índice de sobre vivencia infantil.
2. La capacidad de adquirir conocimientos, medida mediante una combinación del grado de alfabetismo de los adultos y el nivel de asistencia escolar conjunto de niños, adolescentes y jóvenes (de 6 a 24 años)
3. La capacidad de contar con el acceso a los recursos que permitan disfrutar de un nivel digno y decoroso, medido por el PIB per cápita ajustado al poder adquisitivo del dólar de los Estados Unidos.

Formula: $IDH = (IEV + IE + IPIB) / 3$

Donde:

IDH = Índice de desarrollo humano

IEV = Índice de esperanza de vida o, en su caso, Índice de sobre vivencia infantil.

4. Total viviendas

Descripción: Cualquier espacio delimitado que el momento del censo se utilice para alojamiento aunque haya sido construido para un fin distinto al de habitación (faros, escuelas, cuevas, bodegas, tiendas, fábricas o talleres.

5. Porcentaje de personas de 15 a 64 años

Descripción: Porcentaje del número de personas en este rango de edad respecto a la población total.

6. Densidad demográfica

Descripción: Número de personas por Kilómetro cuadrado.

Formula: $\text{Numero de personas} / \text{superficie total}$.

7. Inmigración

Descripción: Acción mediante la cual una persona llega a radicar a una unidad geográfica determinada (municipio o delegación, entidad o país), procedente de otra.

8. Tasa media de crecimiento anual

Descripción: Porcentaje de incremento de la población de un año a otro.

9. Porcentaje de personas menores de 15 años

Descripción: Porcentaje de personas de esa edad respecto a la población total.

10. Porcentaje de personas de 5 y más años que hablan una lengua indígena

Descripción: Porcentaje de personas que practican parte del conjunto de idiomas que históricamente son herencia de las diversas etnias del continente americano.

11. Porcentaje de viviendas con energía eléctrica

Descripción: Porcentaje de viviendas en las cuales existe energía eléctrica para alumbrar la vivienda, sin considerar la fuente de donde provenga.

12. Porcentaje de viviendas con agua entubada

Descripción: Accesibilidad de los ocupantes de la vivienda al uso del agua entubada dentro o fuera de la vivienda pero dentro del terreno, así como la forma de abastecimiento cuando no disponen de ella.

13. Porcentaje de viviendas con drenaje

Descripción: Sistema de tuberías mediante el cual se elimina de la vivienda las aguas negras o las aguas sucias.

Si al menos una de las instalaciones sanitarias de la vivienda (lavadero, sanitario, fregadero o regadera) disponen de un sistema de tuberías para eliminar las aguas negras o aguas sucias, se considera que tiene drenaje.

14. Ocupantes por vivienda

Descripción: Promedio de personas que habitan una vivienda.

15. Población Total

Descripción: Total de la población que habita en el municipio.

16. Superficie en Km².

Descripción: Superficie total del municipio en Km².

17. Porcentaje de personas que residen en localidades de más de 2500 habitantes.

Descripción: Porcentaje de personas que residen en localidades de más de 2500 habitantes.

18. Porcentaje de personas entre 6 y 24 años que asisten a la escuela.

Descripción: Porcentaje de personas entre 6 y 24 años que asisten a la escuela.

19. Promedio mensual de suscriptores por municipio

Descripción: Promedio de personas que están suscritos por mes en un municipio dado.

Municipio

Descripción: División territorial político administrativa de una entidad federativa.

En el caso del Distrito Federal, las 16 delegaciones políticas son equivalentes a los municipios.

7 METODOLOGIA.

Se comenzará tomando en cuenta 19 indicadores demográficos, se observará que correlación guarda con el promedio de instalaciones mensuales, se discriminaran aquellos indicadores que estén correlacionados entre si. Se usara el estadístico C-p para discriminar el mayor número de variables posibles, obteniendo el mejor modelo de regresión múltiple.

Se crea un subconjunto de los datos que incluya sólo los municipios del área metropolitana y la ciudad de México, al cual llamaremos modelo piloto, ya que se considera la hipótesis de que las variables que intervengan en el modelo piloto pueden variar con respecto al modelo completo, partiendo de la idea que la vida en el área metropolitana tiene diferencias con la vida de provincia.

Una vez creado el modelo piloto, se creara el modelo total, si existe diferencia entre las variables incluidas en el mejor modelo y si el coeficiente de determinación disminuye considerablemente, (< 0.85), se tomara la opción de crear dos modelos: el primero para el área metropolitana (ya realizado, como modelo piloto) y el segundo para el interior del país

Si nuestro modelo final tiene un coeficiente de determinación aceptable (mayor a 0.85) se tomará en cuenta realizar pronósticos con base en las variables involucradas en el modelo.

8 ANALISIS DE DATOS.

Crearemos un modelo piloto, usando solo los municipios del área metropolitana para observar si realmente puede realizarse un buen modelo de regresión con las variables consideradas.

8.1 MODELO PILOTO (Área metropolitana).

8.1.1 Correlación.

Comenzamos por ver que tan correlacionados están nuestros indicadores con la variable suscriptores, se usa la correlación de Pearson y el valor que se encuentra debajo de cada coeficiente de correlación es el valor p de la prueba de hipótesis siguiente

$H_0: r = 0$ versus $H_1: r \neq 0$ donde r es la correlación entre un par de variables.

Se incluirán en el modelo solo las variables que obtengan un coeficiente de correlación mayor a 0.3

Correlación:

	variables	sup.(Km ²)	densidad	inmigración	ocupantes	tasa media de crecimiento	residentes en localidades	de 5 y mas años	alfabetas	escolares	PIB dólares
suscripciones	r	-0.126	0.580	0.225	-0.614	-0.485	0.482	0.136	0.557	0.528	0.569
	valor P	0.392	0.000	0.124	0.000	0.000	0.001	0.355	0.000	0.000	0.000

	variables	nivel escolar	desarrollo	total de viviendas	de 15 a 64	población total	menores de 15	con energía	con agua	con drenaje
suscripciones	r	0.566	0.650	0.887	0.738	0.848	-0.647	0.247	0.448	-0.591
	valor P	0.001	0.000	0.000	0.000	0.000	0.000	0.090	0.001	0.000

Se crean los intervalos de confianza del 95 % para todas las correlaciones usando la aproximación de Ruben para observar si el coeficiente de correlación es significativo para estos casos:

r=	0.126	0.58	0.225	0.614	0.485	0.482	0.136	0.557	0.528	0.569	0.566	0.65	0.887	0.738	0.848	0.647	0.247	0.448	0.591
inf	-0.1649	0.374	-0.06	0.43	0.236	0.232	-0.15	0.338	0.295	0.356	0.352	0.495	1.344	0.692	1.095	0.49	-0.04	0.188	0.391
sup	0.1748	0.8	0.284	0.872	0.629	0.624	0.186	0.755	0.701	0.778	0.772	0.957	2.186	1.224	1.805	0.95	0.309	0.571	0.822

Los únicos intervalos de confianza que incluyen al cero son aquellos cuyo valor de correlación son menores a 0.3, es decir, no se descarta la posibilidad de que el coeficiente de correlación sea cero. Por lo tanto se descartan estas variables de nuestro primer modelo de regresión.

En conclusión nuestro primer modelo de regresión incluye las siguientes variables:

- Densidad demográfica
- Ocupantes
- Tasa media de crecimiento
- Residentes en localidades de 2500 hab.
- Alfabetas
- Escolares
- PIB
- Nivel escolar
- Desarrollo humano
- Total de viviendas
- Porcentaje de población de 15 a 64 años
- Población total
- Menores de 15 años
- Porcentaje de viviendas con agua
- Porcentaje de viviendas con drenaje

8.1.2 Mejor subconjunto de regresión.

La siguiente tabla nos muestra a los dos mejores subconjuntos de regresión que se pueden formar, usando desde una variable hasta las 14 variables que seleccionamos anteriormente. De esta forma se escogerá al subconjunto que reúna las siguientes características:

- ✓ El estadístico C-p en valor absoluto sea menor al número de las variables incluidas en el subconjunto.
- ✓ El valor de R-sq adjunta sea el mayor de todos.
- ✓ El valor de la desviación estándar S sea el mas pequeño.
- ✓ Por ultimo se considera un mejor subconjunto aquel que tenga un menor número de variables.

# de variables	R-Sq	R-Sq(adj)	C-p	S	Ocupantes	TASA MEDIA	RESIDENTES EN LOCALIDADES	PORCENTAJE DE ALFABETAS	PORCENTAJE DE PERSONAS QUE ASISTEN A LA ESCUELA	PIB	NIVEL DE ESCOLARIDAD	NIVEL DE DESARROLLO	TOTAL DE VIVIENDAS	DE 15 A 64 AÑOS	POBLACION TOTAL	MENORES DE 15 AÑOS	VIVIENDAS CON AGUA	VIVIENDAS CON DRENAJE
1	78.6	78.2	46.2	148.79									X					
1	71.9	71.3	74.6	170.68											X			
2	91.0	90.6	-4.2	97.385									X		X			
2	88.1	87.6	8.2	112.25									X			X		
3	91.6	91.0	-4.4	95.639								X	X		X			
3	91.5	90.9	-4.2	95.934						X			X		X			
4	91.9	91.1	-3.7	94.990		X				X			X		X			
4	91.8	91.1	-3.6	95.046		X						X	X		X			
5	92.0	91.0	-2.1	95.500	X	X						X	X		X			
5	91.9	91.0	-2.0	95.609	X	X				X			X		X			
6	92.0	90.9	-0.3	96.289	X	X		X				X	X		X			
6	92.0	90.8	-0.3	96.414	X	X				X			X		X		X	
7	92.1	90.7	1.4	97.072	X	X		X		X		X	X		X		X	
7	92.1	90.7	1.4	97.169	X	X		X				X	X		X		X	
8	92.1	90.5	3.2	98.007	X	X		X				X	X		X	X		X
8	92.1	90.5	3.2	98.125	X	X		X		X		X	X		X		X	
9	92.2	90.3	5.0	99.059	X	X		X			X	X	X		X	X		X
9	92.2	90.3	5.0	99.059	X	X		X	X			X	X		X	X		X
10	92.2	90.1	7.0	100.38	X	X		X			X	X	X		X	X	X	X
10	92.2	90.1	7.0	100.38	X	X		X	X			X	X		X	X	X	X

11	92.2	89.8	9.0	101.76	X	X		X			X	X	X	X	X	X	X	X
11	92.2	89.8	9.0	101.76	X	X		X		X	X	X	X		X	X	X	X
12	92.2	89.5	11.0	103.20	X	X		X		X	X	X	X	X	X	X	X	X
12	92.2	89.5	11.0	103.20	X	X	X	X			X	X	X	X	X	X	X	X
13	92.2	89.2	13.0	104.71	X	X	X	X		X	X	X	X	X	X	X	X	X
13	92.2	89.2	13.0	104.71	X	X		X	X	X	X	X	X	X	X	X	X	X
14	92.2	88.9	15.0	106.29	X	X	X	X	X	X	X	X	X	X	X	X	X	X

El mejor subconjunto es el que consta de cuatro variables dado que:

- El valor C-p igual a 3.7 es menor que el numero de variables igual a cuatro.
- El valor de R-sq(adj) igual a 91.1 es el mayor de todos.
- El valor de S igual a 94.990 es el menor de todos.

Por lo tanto las variables que incluimos en nuestro primer modelo son:

- Tasa media de crecimiento
- PIB
- Total de viviendas
- Población total

8.1.3 Modelo de regresión.

Para escoger el modelo de regresión mas adecuado primero eliminaremos la colinealidad de las variables incluidas en el modelo por medio del factor inflacionario de varianza, por lo que generamos el primer modelo de regresión y observamos el VIF de cada variable, si al menos uno es mayor de 5, procedemos a quitar esa variable, se hará este paso tantas veces como sea necesario, hasta que ninguna variable tenga $VIF > 5$.

- Primer modelo de regresión múltiple.

La ecuación de regresión es

Suscripciones = - 38.3 + 10.6 Tasa media de crecimiento anual + 0.00414 PIB dólares + 0.0144 total viviendas - 0.00261 Población total

variables	Coefficiente	SE Coeficiente	T	P	VIF
constante	-38.32	41.09	-0.93	0.356	
Tasa media	10.628	7.640	1.39	0.171	1.8
PIB dólares	0.004139	0.003098	1.34	0.189	2.8
Total de viviendas	0.014400	0.002698	5.34	0.000	316.5
Población total	-0.0026143	0.0006020	-4.34	0.000	299.2

S = 95.92 R-Sq = 91.7% R-Sq(adj) = 90.9%

Se observa que las variables cuyo valor VIF es mayor a 5 son las siguientes:

- Total de viviendas
- Población total

Por lo que procedemos a quitar la variable “*Población total*” dado que la variable “*total de viviendas*” tiene una mayor correlación con los suscriptores.

➤ Segundo modelo de regresión múltiple.

La ecuación de regresión es

Suscripciones = - 68.8 - 1.26 Tasa media de crecimiento anual + 0.0134 PIB dólares + 0.00270 total viviendas

variables	Coefficiente	SE Coeficiente	T	P	VIF
constante	-68.75	48.01	-1.43	0.159	
Tasa media	-1.263	8.457	-0.15	0.882	1.6
PIB dólares	0.013368	0.002673	5.00	0.000	1.5
Total de viviendas	0.0027043	0.0001959	13.81	0.000	1.2

S = 113.7 R-Sq = 88.1% R-Sq(adj) = 87.2%

Dado que en este modelo ninguna variable obtiene un valor VIF mayor a 5 podemos concluir que estas variables no son colineales entre si.

Observemos ahora que el valor p nos indica cuando una variable contribuye de forma significativa en el modelo y cuando no es útil. Por lo que si el valor p es mayor a 0.005 entonces no se rechaza la hipótesis de que la variable X_j no mejora significativamente el modelo, es decir, existe evidencia suficiente para considerar que esta variable no es significativa para el modelo de regresión.

➤ Tercer modelo de regresión múltiple

La ecuación de regresión es

Suscripciones = - 74.5 + 0.0136 PIB dólares + 0.00271 total viviendas

variables	Coefficiente	SE Coeficiente	T	P	VIF
constante	-74.55	27.95	-2.67	0.011	
PIB dólares	0.013571	0.002278	5.96	0.000	1.1
Total de Viviendas	0.0027118	0.0001873	14.48	0.000	1.1

S = 112.5 R-Sq = 88.1% R-Sq(adj) = 87.5%

Se observa que el valor p de las variables restantes es menor que 0.005, por lo tanto se considera este modelo para realizar el análisis residual.

8.1.4 Análisis de residuos.

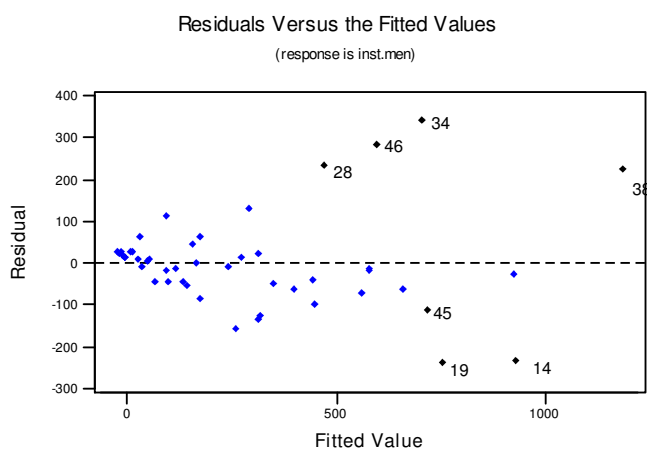
Análisis de varianza:

Fuente	DF	SS	MS	F	P
Regresión	2	4196034	2098017	95.67	0.000
Error Residual	45	569406	12653		
Total	47	4765440			

El valor p del modelo de regresión es menor de 0.005 por lo tanto se rechaza la hipótesis de que los coeficientes de la variable sean cero, es decir, al menos una de las variables contribuye de manera significativa al modelo de regresión.

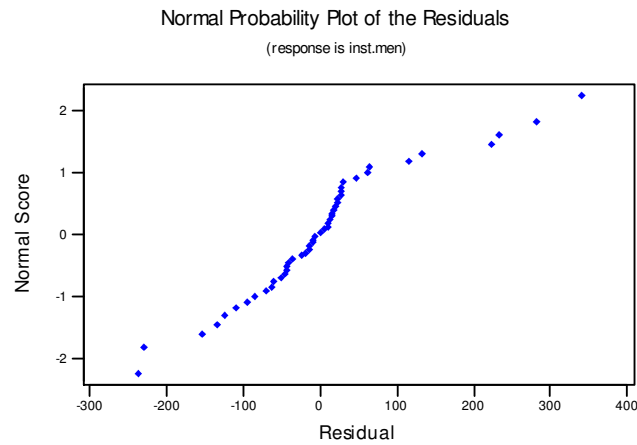
Fuente	DF	Seq SS
PIB dólares	1	1543583
Población total	1	2652451

1. Residuos contra valores ajustados.



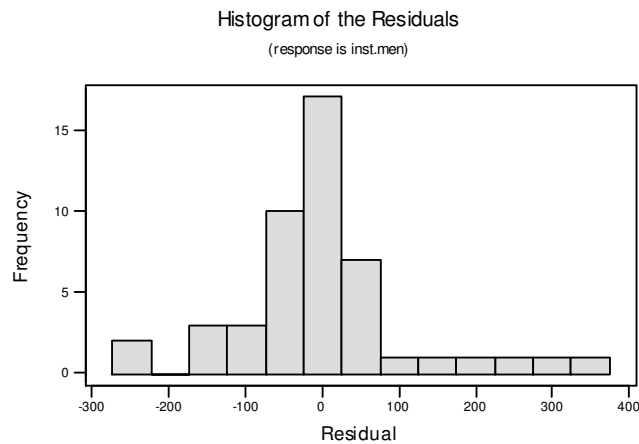
Se observa que los residuos tienden a alejarse del centro, por lo que se podría pensar que la varianza no es constante, sin embargo, los puntos extremos que se encuentran señalados en esta grafica son los que se alejan del centro, mientras que los demás residuos si se comportan de forma aleatoria, por lo tanto, no se descarta la idea de que sea un modelo adecuado.

2. Grafica de la distribución de los residuos



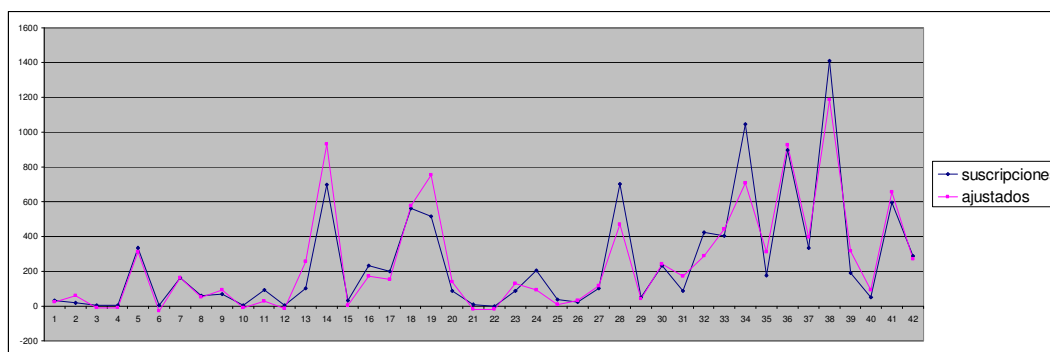
Se observa que los residuos tienen un comportamiento lineal y centrado en el cero.

3. Histograma de los residuos.



El histograma nos muestra que efectivamente los residuos tienden a comportarse como una variable aleatoria normal con media cero y varianza constante.

8.1.5 Puntos extremos.



En la grafica anterior se observa que tanto se acerca nuestro modelo de regresión lineal a las suscripciones reales, también se observan los puntos extremos u observaciones extrañas, las cuales corresponden a los siguientes municipios:

Obs.	Municipio	Total de viviendas	PIB	Suscriptores	Ajustado	SE ajuste	Residuo	St Residuo	
14	Ecatepec de Morelos	346922	4621	697.0	928.9	56.2	-231.9	-2.38	RX
19	Nezahualcóyotl	274984	6151	517.0	754.6	42.1	-237.6	-2.28	R
28	Tlalnepantla de Baz	166006	6812	701.0	468.1	24.5	232.9	2.12	R
34	Coyoacán	163036	24943	1046.0	706.1	37.0	339.9	3.20	R
38	Iztapalapa	403922	12184	1409.0	1186.1	60.4	222.9	2.35	RX
45	Benito Juárez	113741	35594	605.0	716.9	59.9	-111.9	-1.18	X
46	Cuauhtémoc	1471818	20018	879.0	596.2	28.0	282.8	2.60	R

R denota una observación con un gran residuo estandarizado.
X denota una observación cuyo valor X influye demasiado.

Se concluye que el modelo es adecuado y dado que su valor de determinación $R_{sq(adj)}$ es igual a 87.5% , se considera la opción de crear un modelo completo de regresión múltiple utilizando estas variables,

8.2 MODELO COMPLETO.

En esta sección se utilizarán todos los datos que se obtuvieron para crear un modelo de regresión lineal, dichos municipios corresponden solo a las siguientes entidades:

- Sinaloa
- Veracruz
- Estado de México
- Distrito Federal

En total se consideran los municipios donde existe por lo menos un suscriptor en un periodo de dos meses, por lo tanto, en total se consideraron 122 municipios.

8.2.1 Correlaciones incluyendo todos los datos

Comenzamos por ver que tan correlacionados están nuestros indicadores con la variable suscriptores, se usa la correlación de Pearson y el valor que se encuentra debajo de cada coeficiente de correlación es el valor p de la prueba de hipótesis siguiente

$H_0: r = 0$ versus $H_1: r \neq 0$ donde r es la correlación entre un par de variables.

Se incluirán en el modelo solo las variables que obtengan un coeficiente de correlación mayor a 0.25

Correlación:

	variables	sup.(Km ²)	densidad	inmigración	ocupantes	tasa media de crecimiento	residentes en localidades	de 5 y mas años	alfabetas	escolares	PIB dólares
suscripciones	r	-0.172	0.586	-0.028	-0.296	-0.004	0.462	-0.082	0.447	0.438	0.783
	valor P	0.062	0.000	0.758	0.001	0.962	0.000	0.374	0.000	0.000	0.000

	variables	nivel escolar	desarrollo	total de viviendas	de 15 a 64	población total	menores de 15	con energía	con agua	con drenaje
suscripciones	r	0.454	0.594	0.811	0.614	0.783	-0.587	0.309	0.375	0.445
	valor P	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000

Se crean los intervalos de confianza del 95 % para todas las correlaciones usando la aproximación de Ruben para observar si el coeficiente de correlación es significativo para estos casos:

r=	0.172	0.586	0.028	0.296	0.004	0.462	0.082	0.447	0.438	0.783	0.454	0.594	0.811	0.614	0.783	0.58	0.309	0.445
inf	-0.01	0.51	-0.15	0.125	-0.18	0.324	-0.1	0.304	0.293	0.984	0.314	0.523	1.095	0.559	0.984	0.499	0.14	0.302
sup	0.193	0.757	0.045	0.331	0.02	0.548	0.1	0.526	0.513	1.316	0.536	0.772	1.45	0.813	1.316	0.745	0.347	0.523

Los únicos intervalos de confianza que incluyen al cero son aquellos cuyo valor de correlación son menores a 0.25, es decir, no se descarta la posibilidad de que el coeficiente de correlación sea cero. Por lo tanto se incluyen solo las variables cuyo valor de correlación sea mayor a 0.25, las cuales son:

- Densidad demográfica
- Ocupantes
- Residentes en localidades de 2500 hab.
- Porcentaje de personas alfabetas mayores de 15 años
- Porcentaje de personas de 6 a 24 años que asisten a la escuela.
- PIB en dólares.
- Nivel escolar
- Nivel de desarrollo humano
- Total de viviendas
- Porcentaje de población de 15 a 64 años
- Población total
- Porcentaje de personas menores de 15 años
- Porcentaje de viviendas con energía
- Porcentaje de viviendas con agua
- Porcentaje de viviendas con drenaje

8.2.2 Mejor subconjunto de regresión.

# de variables	R-Sq	R-Sq(adj)	C-p	S	DENSIDAD	OCUPANTES	RESIDENTES EN LOCALIDADES	PORCENTAJE DE ALFABETAS	PORCENTAJE DE PERSONAS QUE ASISTEN A LA ESCUELA	PIB	NIVEL DE ESCOLARIDAD	NIVEL DE DESARROLLO	TOTAL DE VIVIENDAS	DE 15 A 64 AÑOS	POBLACION TOTAL	MENORES DE 15 AÑOS	CON ENERGIA	VIVIENDAS CON AGUA	VIVIENDAS CON DRENAJE
1	63.8	63.0	42.9	201.87									X						
1	58.1	57.2	56.5	217.13											X				
2	80.0	79.1	6.0	151.77						X			X						
2	79.8	78.9	6.5	152.50						X					X				
3	82.2	80.9	2.8	144.95						X		X	X						
3	82.0	80.8	3.1	145.42				X		X			X						
4	82.9	81.4	2.9	143.34						X		X	X	X					
4	82.9	81.3	3.0	143.42						X	X	X	X						
5	84.2	82.3	2.0	139.82						X		X	X	X		X			
5	83.8	81.8	2.9	141.46		X				X		X	X	X					
6	84.7	82.5	2.7	138.98		X				X		X	X	X		X			
6	84.7	82.4	2.8	139.18						X	X	X	X	X		X			
7	85.2	82.6	3.6	138.55		X		X		X	X		X	X		X			
7	85.1	82.5	3.7	138.75		X		X		X	X	X	X	X					
8	85.6	82.6	4.6	138.43		X		X		X	X	X	X		X				
8	85.4	82.4	5.2	139.48		X		X		X	X		X	X	X	X			
9	85.9	82.5	5.9	138.73		X		X		X	X	X	X	X	X	X			X
9	85.7	82.4	6.2	139.41		X		X	X	X	X		X	X	X	X			X
10	86.0	82.2	7.5	139.89		X		X	X	X	X	X	X	X	X	X			
10	86.0	82.2	7.6	139.93	X	X		X		X	X	X	X	X	X	X			X
11	86.2	82.0	9.1	140.89		X	X	X	X	X	X	X	X	X	X	X			X
11	86.2	81.9	9.2	141.07	X	X		X	X	X	X	X	X	X	X	X			X
12	86.2	81.5	11.0	142.71	X	X	X	X	X	X	X	X	X	X	X	X			X
12	86.2	81.5	11.1	142.83		X	X	X	X	X	X	X	X	X	X	X	X		X
13	86.2	81.0	13.0	144.75	X	X	X	X	X	X	X	X	X	X	X	X	X		X
13	86.2	81.0	13.0	144.77	X	X	X	X	X	X	X	X	X	X	X	X		X	X
14	86.3	80.4	15.0	146.91	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

El mejor subconjunto es el que consta de 7 variables dado que:

- El valor C-p igual a 3.6 es menor que el numero de variables igual a 7.
- El valor de R-sq(adj) igual a 82.6% es el mayor de todos.
- El valor de S igual a 138.55 es de los más pequeños.

Sin embargo, existe otro subconjunto el cual obtuvo también un valor de R-sq adjunta igual a 82.6%, además el valor de C-p es igual a 4.6, el cual es menor que el numero de variables igual a 8 y un valor S igual a 138.43 que es menor que 138.55. Pero dado que este subconjunto involucra a una variable más que el otro, se considera mejor subconjunto aquel que tenga menor numero de variables.

Por lo tanto las variables que incluimos en nuestro primer modelo son:

- Ocupantes
- Porcentaje de personas mayores de 15 años alfabetas.
- PIB
- Nivel escolar
- Total de viviendas
- Porcentaje de personas de 15 a 64 años
- Menores de 15 años

8.2.3 Modelo de regresión.

- Primer modelo de regresión.

La ecuación de regresión es

Suscripciones = - 348 + 33.3 Ocupantes - 8.47 alfabetas+15 + 0.0199 PIB dólares + 756 nivel escolar + 0.00200 total de viviendas + 5.68 De 15 a 64 años(%) - 1.74 menores de 15 años

variables	Coficiente	SE Coficiente	T	P	VIF
constante	-348.5	523.3	-0.67	0.507	
Ocupantes	33.33	38.35	0.87	0.387	2.9
Alfabetas mayores de 15 años	-8.473	2.816	-3.01	0.003	6.0
PIB dólares	0.019873	0.001990	9.99	0.000	2.2
Nivel escolar	756.0	477.9	1.58	0.116	6.3
Total de viviendas	0.0019972	0.0001604	12.45	0.000	1.6
De 15 a 64 años	5.680	6.207	0.92	0.362	7.3
Menores de 15 años	-1.745	6.350	-0.27	0.784	10.2

S = 96.92 R-Sq = 85.9% R-Sq(adj) = 85.0%

Se observa que las variables cuyo valor VIF es mayor a 5 son las siguientes:

- Alfabetas mayores de 15 años
- Nivel escolar

- De 15 a 64 años
- Menores de 15 años

Por lo que procedemos a quitar la variable cuyo valor VIF sea el mas alto, es decir, la variable “*menores de 15 años*”.

- Segundo modelo de regresión múltiple.

La ecuación de regresión es

Suscripciones = - 452 + 27.7 Ocupantes - 8.11 Porcentajes de personas alfabetas mayores de 15 años + 0.0201 PIB dólares + 734 nivel escolar + 0.00199 total de viviendas + 6.62 De 15 a 64 años(%)

variables	Coficiente	SE Coficiente	T	P	VIF
constante	-451.6	363.2	-1.24	0.216	
Ocupantes	27.67	32.22	0.86	0.392	2.1
Alfabetas mayores de 15 años	-8.109	2.475	-3.28	0.001	4.7
PIB dólares	0.020095	0.001810	11.10	0.000	1.8
Nivel escolar	734.1	469.3	1.56	0.120	6.2
Total de viviendas	0.0019924	0.0001588	12.54	0.000	1.6
De 15 a 64 años	6.616	5.169	1.28	0.203	5.1

S = 96.53 R-Sq = 85.9% R-Sq(adj) = 85.2%

Se observa que las variables cuyo valor VIF es mayor a 5 son las siguientes:

- Nivel escolar
- De 15 a 64 años

Por lo que procedemos a quitar la variable cuyo valor VIF sea el mas alto, es decir, la variable “*nivel escolar*”.

- Tercer modelo de regresión múltiple.

La ecuación de regresión es

Suscripciones = - 298 + 24.4 OCUPANTES - 5.21 alfabetas+15 + 0.0195 PIB dólares + 0.00198 total de viviendas + 10.2 De 15 a 64 años(%)

variables	Coficiente	SE Coficiente	T	P	VIF
constante	-298.0	351.8	-0.85	0.399	
Ocupantes	24.43	32.36	0.76	0.452	2.1
Alfabetas mayores de 15 años	-5.209	1.650	-3.16	0.002	2.1
PIB dólares	0.019493	0.001780	10.95	0.000	1.7
Total de viviendas	0.0019764	0.0001595	12.39	0.000	1.6
De 15 a 64 años	10.158	4.675	2.17	0.032	4.1

$S = 97.13$ $R\text{-}Sq = 85.6\%$ $R\text{-}Sq(\text{adj}) = 85.0\%$

Dado que en este modelo ninguna variable obtiene un valor VIF mayor a 5 podemos concluir que estas variables no son colineales entre sí, es decir, la correlación entre ellas es muy pequeña.

Ahora nos dedicaremos a discriminar aquellas variables cuyo valor p sea mayor a 0.005, dado que la prueba de hipótesis es la siguiente:

$H_0: \beta_j = 0$ contra $H_1: \beta_j \neq 0$.

Donde β_j es el coeficiente de la variable j en el modelo.

Así, observamos que las siguientes variables presentan un valor p mayor a 0.005:

- Ocupantes
- Porcentaje de personas de 15 a 64 años.

Por lo que creamos un nuevo modelo eliminando a la variable “ocupantes”.

- Cuarto modelo de regresión múltiple.

La ecuación de regresión es

Suscripciones = - 70 - 5.04 Porcentaje de personas alfabetos mayores de 15 años + 0.0196 PIB dólares + 0.00200 total de viviendas + 7.86 De 15 a 64 años(%)

variables	Coeficiente	SE Coeficiente	T	P	VIF
constante	-70.1	180.5	-0.39	0.698	
Alfabetas mayores de 15 años	-5.038	1.632	-3.09	0.003	2.0
PIB dólares	0.019618	0.001769	11.09	0.000	1.7
Total de viviendas	0.0019974	0.0001568	12.74	0.000	1.5
De 15 a 64 años	7.855	3.537	2.22	0.028	2.4

$S = 96.95$ $R\text{-}Sq = 85.5\%$ $R\text{-}Sq(\text{adj}) = 85.0\%$

Así, observamos que la variable de 15 a 64 años presenta un valor p mayor a 0.005, por lo tanto, la descartamos del modelo.

- Quinto modelo de regresión múltiple.

La ecuación de regresión es

Suscripciones = 228 - 3.15 Porcentaje de personas alfabetas mayores de 15 años + 0.0207 PIB dólares + 0.00209 total de viviendas

variables	Coeficiente	SE Coeficiente	T	P	VIF
constante	227.7	122.8	1.85	0.066	
Alfabetas mayores de 15 años	-3.155	1.147	-2.23	0.028	1.5
PIB dólares	0.020694	0.001729	11.97	0.000	1.6

Total de viviendas	0.0020873	0.0001539	13.56	0.000	1.4
--------------------	-----------	-----------	-------	-------	-----

$S = 98.55$ $R\text{-}Sq = 84.9\%$ $R\text{-}Sq(\text{adj}) = 84.5$

La variable “*alfabetas mayores de 15 años*” presenta un valor p mayor de 0.005, por lo cual la retiramos de nuestro modelo.

➤ Sexto modelo de regresión múltiple

La ecuación de regresión es

Suscripciones = - 44.5 + 0.0192 PIB dólares + 0.00200 total de viviendas

variables	Coeficiente	SE Coeficiente	T	P	VIF
constante	-44.53	11.53	-3.86	0.000	
PIB dólares	0.019202	0.001620	11.85	0.000	1.3
Total de viviendas	0.0020025	0.0001516	13.21	0.000	1.3

$S = 100.2$ $R\text{-}Sq = 84.3\%$ $R\text{-}Sq(\text{adj}) = 84.0\%$

Las dos variables incluidas en este último modelo no presentan ningún valor p mayor a 0.005, por lo tanto procedemos a hacer el análisis residual.

8.2.4 Análisis residual.

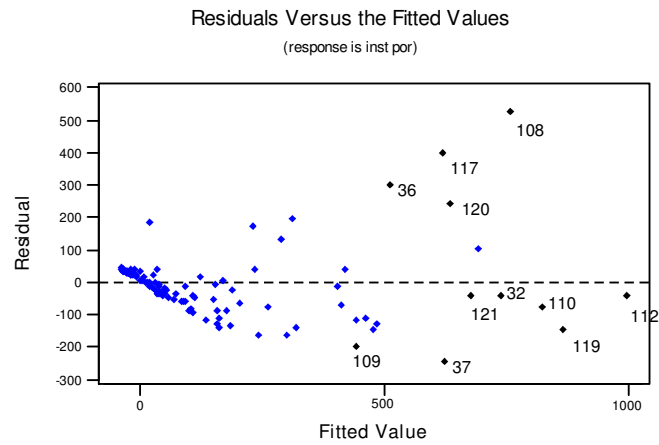
Análisis de varianza:

Fuente	DF	SS	MS	F	P
Regresión	2	6403155	3201578	319.02	0.000
Error Residual	119	1194257	10036		
Total	121	7597412			

El valor p del modelo de regresión es menor de 0.005 por lo tanto se rechaza la hipótesis de que los coeficientes de las variables sean cero, es decir, al menos una de las variables contribuye significativamente al modelo de regresión.

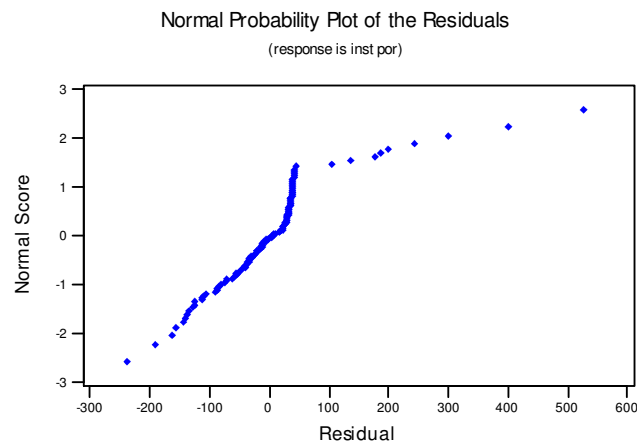
Fuente	DF	Seq SS
PIB dólares	1	465299
Total de viviendas	1	1750556

4. Residuos contra valores ajustados.



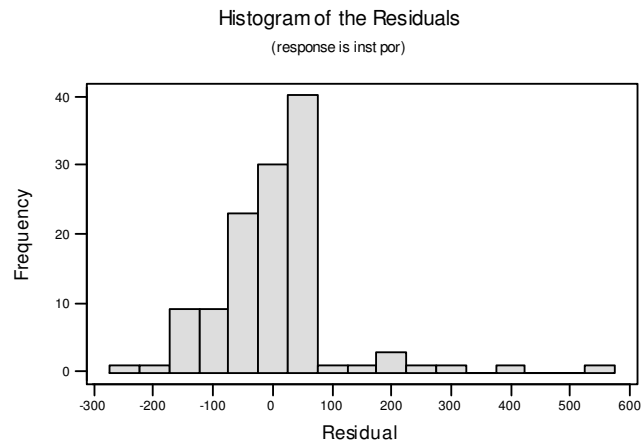
Los residuos tienden a formar un patrón definido lo que indica que la varianza no es muy constante, pero dado que los puntos extremos son los que forman este patrón, se considera la opción de analizar los puntos extremos.

5. Distribución normal de los residuos



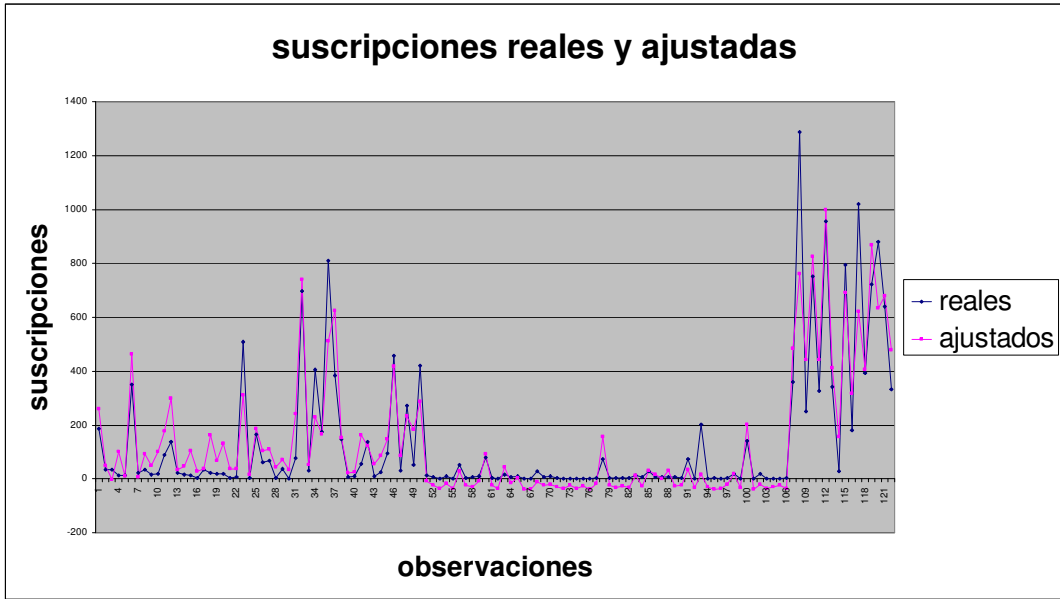
Dado que no se forma una línea recta se considera que la varianza no tiende a ser constante por lo que existe la opción de descartar este modelo.

6. Grafica del total de viviendas contra el promedio mensual de suscriptores.



El histograma nos muestra que efectivamente los residuos no se comportan como una variable normal con media cero y varianza constante por lo que se considera que este modelo no es adecuado.

8.2.5 Puntos extremos.



Observaciones inusuales:

Obs.	Municipio	total de viviendas	PIB	Suscriptores	Ajustado	SE ajuste	Residuo	St Residuo	
32	Ecatepec de Morelos	346922	4621	697	783.9	46.36	-41.9	-0.47	X
36	Naucalpan de Juárez	199026	8211	811.0	511.0	22.64	299.32	3.07	R
37	Nezahualcóyotl	274984	6151	384	624.23	34.59	-240.23	-2.56	RX
108	Coyoacán	163036	24943	1287.00	760.89	29.96	526.11	5.50	RX
109	Cuajimalpa de Morelos	33163	21927	249.00	442.92	30.68	-193.92	-2.03	RX
110	Gustavo A. Madero	295329	14556	752.00	826.36	33.88	-74.36	-0.79	X
112	Iztapalapa	403922	12184	957.00	998.26	49.90	-41.26	-0.48	X
117	Tlalpan	140148	20015	1020.00	620.43	23.53	399.57	4.10	R
119	Benito Juárez	113741	35594	723.00	866.70	46.94	-143.70	-1.62	X
120	Cuauhtémoc	147181	20018	879.00	634.58	23.32	244.42	2.51	R
121	Miguel Hidalgo	94475	27819	641.00	678.83	35.83	-37.83	-0.40	X

R denota una observación con un gran residuo estandarizado.

X denota una observación cuyo valor X influye demasiado.

Se observa que todos los puntos extremos corresponden a municipios del área metropolitana, de hecho la mitad de las delegaciones del Distrito Federal se encuentran como puntos extremos, además el modelo piloto del área metropolitana incluye las mismas variables que este modelo, por lo tanto se considera la opción de crear dos modelos;

- ✓ Un modelo de regresión múltiple para el área metropolitana (modelo piloto ya creado).
- ✓ Un modelo de regresión múltiple para los municipios de provincia.

8.3 MODELO DE REGRESION LINEAL PARA PROVINCIA.

8.3.1 Correlación.

Comenzamos por ver que tan correlacionados están nuestros indicadores con la variable suscriptores, se usa la correlación de Pearson y el valor que se encuentra debajo de cada coeficiente de correlación es el valor p de la prueba de hipótesis siguiente

$H_0: r = 0$ versus $H_1: r \neq 0$ donde r es la correlación entre un par de variables.

Se incluirán en el modelo solo las variables que obtengan un coeficiente de correlación mayor a 0.25

	variables	sup(Km ²)	densidad	inmigración	ocupantes	tasa media de crecimiento	residentes en localidades	de 5 y mas años	alfabetas	escolares	PIB dólares
suscripciones	r	0.427	0.157	0.098	-0.184	0.157	0.149	0.156	0.272	0.158	0.450
	valor P	0.000	0.190	0.405	0.117	0.182	0.213	0.191	0.019	0.179	0.000

	variables	nivel escolar	desarrollo	total de viviendas	de 15 a 64	población total	menores de 15	con energía	con agua	con drenaje
suscripciones	r	0.208	0.418	0.871	0.268	0.885	-0.297	0.228	0.281	0.260
	valor P	0.075	0.000	0.000	0.021	0.000	0.010	0.051	0.015	0.025

Se crean los intervalos de confianza del 95 % para todas las correlaciones usando la aproximación de Ruben para observar si el coeficiente de correlación es significativo para estos casos:

r=	0.427	0.157	0.098	0.184	0.157	0.149	0.156	0.272	0.158	0.45	0.208	0.418	0.871	0.268	0.885	0.297	0.228	0.281	0.26
inf	0.2235	-0.07	-0.13	-0.05	-0.07	-0.08	-0.08	0.046	-0.07	0.253	-0.02	0.212	1.332	0.042	1.437	0.073	-0	0.056	0.033
sup	0.515	0.19	0.128	0.219	0.19	0.182	0.189	0.318	0.191	0.548	0.245	0.502	1.922	0.313	2.065	0.347	0.268	0.328	0.304

Se observa que los intervalos de confianza que incluyen al cero son aquellos cuyo valor de correlación son menores a 0.25, es decir, no se descarta la posibilidad de que el coeficiente de correlación sea cero. Por lo tanto se incluyen solo las variables cuyo valor de correlación sea mayor a 0.25.

Por lo tanto las variables que se incluyen en el análisis del mejor subconjunto son:

- sup.(Km²)
- Alfabetas
- PIB dólares
- Desarrollo humano
- Total de viviendas
- Porcentaje de población de 15 a 64 años
- Población total
- Menores de 15 años
- Porcentaje de viviendas con agua
- Porcentaje de viviendas con drenaje.

8.3.2 Mejor subconjunto.

# de variables	R-Sq	R-Sq(adj)	C-p	S	SUP(Km ²)	PORCENTAJE DE ALFABETAS	PIB	NIVEL DE DESARROLLO	TOTAL DE VIVIENDAS	DE 15 A 64 AÑOS	POBLACION TOTAL	MENORES DE 15 AÑOS	VIVIENDAS CON AGUA	VIVIENDAS CON DRENAJE
1	78.3	78.0	14.3	26.47							X			
1	75.8	75.5	23.6	27.94					X					
2	81.6	81.0	3.9	24.55					X		X			
2	80.0	79.5	9.7	25.56		X					X			
3	82.1	81.3	3.9	24.37		X			X		X			
3	82.0	81.2	4.3	24.44			X		X		X			
4	83.3	82.3	1.5	23.72		X			X		X		X	
4	82.9	81.9	3.0	24.01		X		X	X		X			
5	83.7	82.4	2.1	23.64		X		X	X		X		X	
5	83.7	82.4	2.1	23.65	X	X			X		X		X	
6	83.8	82.3	3.5	23.72	X	X			X		X	X	X	
6	83.8	82.3	3.7	23.74	X	X			X		X		X	X
7	83.9	82.1	5.2	23.84	X	X			X		X	X	X	X
7	83.9	82.1	5.3	23.86	X	X	X		X		X	X	X	
8	84.0	81.9	7.0	24.00	X	X	X		X		X	X	X	X
8	84.0	81.9	7.1	24.00	X	X			X	X	X	X	X	X
9	84.0	81.6	9.0	24.18	X	X	X		X	X	X	X	X	X
9	84.0	81.6	9.0	24.19	X	X		X	X	X	X	X	X	X
10	84.0	81.3	11.0	24.38	X	X	X	X	X	X	X	X	X	X

El mejor subconjunto es el que consta de 5 variables dado que:

- El valor C-p igual a 2.1 es menor que el numero de variables igual a 5.
- El valor de R-sq(adj) igual a 82.4% es el mayor de todos.

- El valor de S igual a 23.64 es el más pequeño.

Por lo tanto las variables que incluimos en nuestro primer modelo son:

- Porcentaje de personas mayores de 15 años alfabetas.
- Nivel de desarrollo
- Total de viviendas
- Población total
- Porcentajes de viviendas con agua.

8.3.3 Modelo de regresión.

- Primer modelo de regresión.

La ecuación de regresión es

Suscripciones = 71.1 - 1.59 Porcentaje de personas alfabetas mayores de 15 años + 50.7 desarrollo humano - 0.00448 total de viviendas + 0.00149 población total + 0.315 Porcentaje de viviendas con agua entubada

variables	Coefficiente	SE Coeficiente	T	P	VIF
constante	71.09	38.40	1.85	0.068	
Alfabetas mayores de 15 años	-1.5867	0.6304	-2.52	0.014	3.0
Desarrollo humano	50.73	50.55	1.00	0.319	3.2
Total de viviendas	-0.004475	0.001383	-3.24	0.002	251.0
Población total	0.0014876	0.0003373	4.41	0.000	246.9
Con agua	0.3148	0.17868	1.78	0.079	2.0

S = 23.33 R-Sq = 83.5% R-Sq(adj) = 82.3%

Se observa que las variables cuyo valor VIF es mayor a 5 son las siguientes:

- Total de viviendas
- Población total

Por lo que procedemos a quitar la variable cuyo valor VIF sea el mas alto, es decir, la variable “*total de viviendas*”.

- Segundo modelo de regresión.

La ecuación de regresión es

Suscripciones = 103 - 1.95 Porcentaje de personas alfabetas mayores de 15 años + 57.1 desarrollo humano + 0.000399 población total + 0.264 Porcentaje de viviendas con agua entubada

variables	Coefficiente	SE Coeficiente	T	P	VIF
constante	103.14	39.56	2.61	0.011	
Alfabetas mayores de 15 años	-1.9531	0.6613	-2.95	0.004	2.9

años					
Desarrollo humano	57.09	53.87	1.06	0.293	3.2
Población total	0.00039900	0.00002651	15.05	0.000	1.3
Con agua	0.2641	0.1878	1.41	0.164	2.0

S = 24.88 R-Sq = 81.0% R-Sq(adj) = 79.8%

Dado que en este modelo ninguna variable obtiene un valor VIF mayor a 5 podemos concluir que estas variables no son colineales entre si, es decir, la correlación entre ellas es muy pequeña.

Ahora nos dedicaremos a discriminar aquellas variables cuyo valor p sea mayor a 0.005, dado que la prueba de hipótesis es la siguiente:

H0: $\beta_j = 0$ contra H1: $\beta_j \neq 0$.

Donde β_j es el coeficiente de la variable j en el modelo.

Así, observamos que las siguientes variables presentan un valor p mayor a 0.005:

- Desarrollo humano.
- Porcentaje de viviendas con agua.

Comenzamos por descartar la variable cuyo valor p sea mayor, es decir, la variable “*desarrollo humano*”.

- Tercer modelo de regresión.

La ecuación de regresión es

Suscripciones = 101 - 1.56 Porcentaje de personas alfabetas mayores de 15 años + 0.000407 población total + 0.325 Con agua entubada a

variables	Coeficiente	SE Coeficiente	T	P	VIF
constante	100.74	39.53	2.55	0.13	
Alfabetas mayores de 15 años	-1.5588	0.5472	-2.85	0.006	2.0
Población total	0.00040734	0.00002533	16.08	0.000	1.2
Con agua	0.3254	0.1788	1.82	0.073	1.8

S = 24.90 R-Sq = 80.6% R-Sq(adj) = 79.8%

La variable Porcentaje de viviendas con agua tiene un valor 0.073 mayor a 0.005, por lo tanto, se descarta del modelo.

- Cuarto modelo de regresión.

La ecuación de regresión es

Suscripciones = 70.8 - 0.939 alfabetas+15 +0.000409 población total

variables	Coeficiente	SE Coeficiente	T	P	VIF
constante	100.74	39.53	2.55	0.13	
Alfabetas mayores de 15	-0.9391	0.4352	-2.16	0.034	1.2

años					
Población total	0.00040939	0.00002572	15.92	0.000	1.2

S = 25.30 R-Sq = 79.7% R-Sq(adj) = 79.2%

La variable “*Alfabetas mayores de 15 años*” tiene un valor 0.034 mayor a 0.005, por lo tanto, se descarta del modelo.

➤ Quinto modelo de regresión.

La ecuación de regresión es

Suscripciones = - 7.63 +0.000386 población total

variables	Coefficiente	SE Coeficiente	T	P
constante	-7.630	3.694	-2.07	0.042
Población total	0.00038581	0.00002387	16.17	0.000

S = 25.94 R-Sq = 78.4% R-Sq(adj) = 78.1%

La variable incluida en este último modelo no presenta valor p mayor a 0.0005, por lo tanto procedemos a hacer el análisis residual.

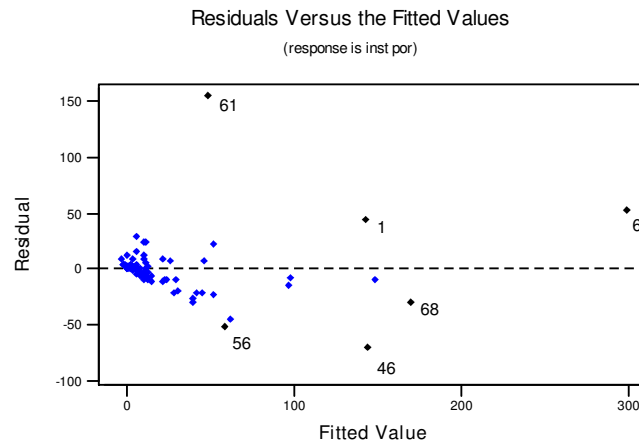
8.3.4 Análisis residual.

Análisis de varianza:

Fuente	DF	SS	MS	F	P
Regresión	1	175821	175821	261.32	0.000
Error Residual	72	48443	673		
Total	73	224264			

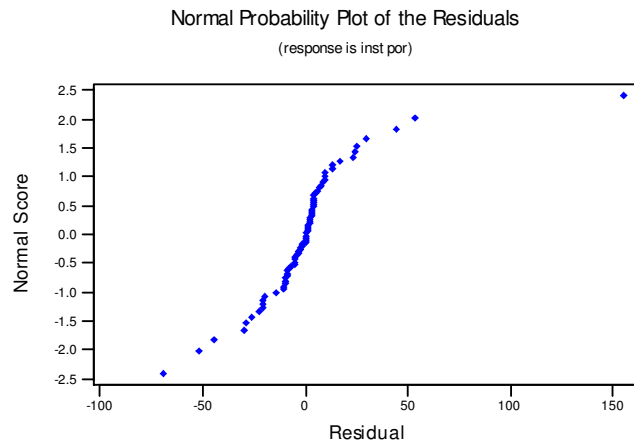
El valor p del modelo de regresión es menor de 0.005 por lo tanto se rechaza la hipótesis de que los coeficientes de la variable y la constante sean cero, es decir, el modelo de regresión es adecuado.

7. Residuos contra valores ajustados.



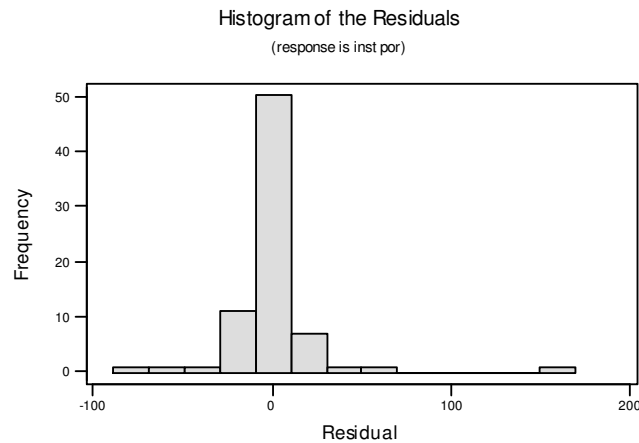
Se observa que los residuos se encuentran muy cercanos al cero y luego tienden a alejarse lo cual indica que la varianza no es muy constante, sin embargo dado que los datos que dan la impresión de alejarse corresponde a los puntos extremos, no se descarta la idea de que el modelo sea adecuado.

8. Distribución normal de los residuos.



Esta grafica nos muestra que los residuos tienden a una distribución normal con media cero.

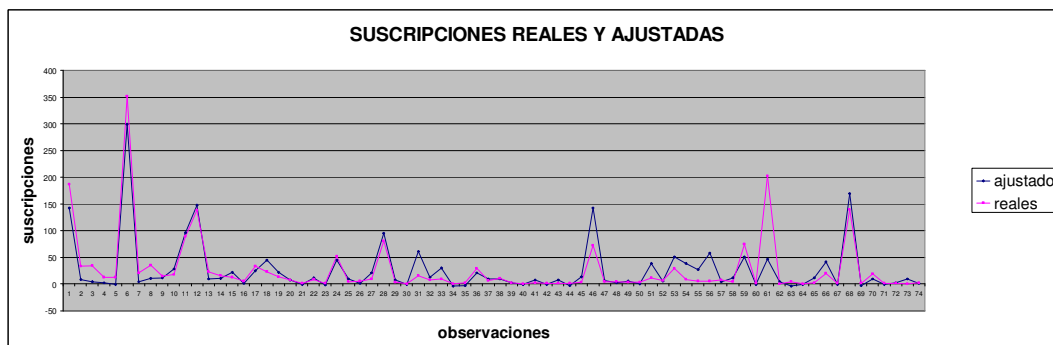
9. Histograma de los residuos.



El histograma nos muestra que efectivamente los residuos están muy cercanos al cero a excepción de muy pocos residuos que corresponden a los puntos extremos, por lo que no se descarta la idea de que sea un modelo adecuado de regresión múltiple.

8.3.5 Puntos extremos.

Por ultimo presentamos la grafica de las suscripciones reales junto con las suscripciones ajustadas.



Las observaciones extrañas pertenecen a los siguientes municipios:

obs. Extrañas	municipio	suscriptores	población total
1	Ahome	186.585	388344
6	Culiacán	351.951	793730
12	Mazatlán	138.293	403888
46	Xalapa	73	390590
56	Papantla	6	170304
61	San Andrés Tuxtla	202.5	142343
68	Veracruz	139.25	457377

Obs.	municipio	Población total	Suscriptores	Ajustado	SE ajuste	Residuo	St Residuo	
1	Ahome	388344	186.59	142.20	7.75	44.39	1.79	X
6	Culiacán	793730	351.95	298.60	17.08	53.35	2.73	RX
12	Mazatlán	403888	138.29	148.19	8.09	-9.90	-0.40	X
46	Xalapa	390590	73.00	143.06	7.80	-70.06	-2.83	RX
56	Papantla	170304	6.00	58.08	3.58	-52.08	-2.03	R
61	San Andrés Tuxtla	142343	202.50	47.29	3.27	155.21	6.03	R
68	Veracruz	457377	139.25	168.83	9.29	-29.58	-1.22	X

R denota una observación con un gran residuo estandarizado.

X denota una observación cuyo valor X influye demasiado

9 Conclusiones.

A partir de este análisis se puede concluir que si se puede predecir el número de suscriptores que en promedio pueden existir mensualmente, para los municipios del área metropolitana, con un nivel de confiabilidad del 87.5% a partir del siguiente modelo:

Suscriptores = $-74.5 + 0.0136 \text{ PIB en dólares} + 0.00200 \text{ Total de viviendas}$.

Lo cual indica que:

- ✓ Por cada 1000 viviendas que existen en un municipio hay 2 suscripciones en promedio mensuales.
- ✓ Cada 1000 dólares del PIB existen 13 suscriptores.
- ✓ Existen en promedio 74.5 suscripciones por municipio que no se pueden predecir con estas variables.

Se observó que los siguientes factores se encuentran directamente relacionados con el promedio de suscriptores por municipio:

- ✓ Población total
- ✓ El porcentaje de personas de 15 a 64 años.
- ✓ El nivel de desarrollo humano.
- ✓ El nivel escolar.
- ✓ El porcentaje de personas alfabetas mayores de 15 años
- ✓ El porcentaje de personas de 6 a 24 años que asisten a la escuela.
- ✓ El porcentaje de residentes en localidades de más de 2500 habitantes
- ✓ El porcentaje de viviendas con agua entubada

Y los factores que se encuentran inversamente relacionados son los siguientes:

- ✓ El promedio de ocupantes por vivienda.
- ✓ La tasa media de crecimiento.
- ✓ El porcentaje de personas menores de 15 años.
- ✓ El porcentaje de viviendas con drenaje.
- ✓

Por ultimo para el área metropolitana se recomienda que:

- ✓ El número de supervisiones sea proporcional al número de suscriptores ajustado por el modelo.
- ✓ Contar con un inventario mínimo mensual igual al número de suscriptores ajustado por el modelo.
- ✓ Se considera que existen factores que intervienen de forma directa en el numero de suscriptores y que no se consideraron en este estudio para los siguientes municipios: Tlalnepantla, Coyoacan, Iztapalapa y Cuauhtémoc, por lo que es conveniente invertir en estos municipios

considerados como puntos extremos y que se encuentran por encima del valor ajustado

- ✓ De la misma forma se considera que existen factores que se encuentran relacionados indirectamente con el número de suscriptores en los siguientes municipios: Ecatepec, Nezahualcoyotl y Benito Juárez, por lo cual no se considera importante un aumento de viviendas en estos municipios.

El modelo de regresión múltiple para los municipios de provincia, no se considera un modelo adecuado, dado que:

- ✓ Los residuos forman un patrón definido.
- ✓ Los residuos no tienen una distribución normal con media cero y varianza constante.
- ✓ El valor de determinación igual a 78.1% no se considera suficiente para predecir.

También se observó que las variables que se encuentran relacionadas directamente con el número de suscriptores son:

- ✓ La población Total.
- ✓ El total de viviendas.
- ✓ El PIB.
- ✓ La superficie del municipio en Km².
- ✓ El nivel de desarrollo humano.
- ✓ El porcentaje de viviendas con drenaje.
- ✓ El porcentaje de viviendas con agua.
- ✓ El porcentaje de personas de 15 a 64 años.
- ✓ El porcentaje de personas alfabetas mayores de 15 años.

Los cuales se encuentran enlistados de mayor a menor relación.

Y la variable que se encuentra relacionada indirectamente con el número de suscriptores es el porcentaje de personas menores de 15 años.

Por lo que se concluye que los factores que se encuentran relacionados con el número de suscriptores en el área metropolitana, difieren con los factores que intervienen en provincia.

10 Referencias.

- i. Giovanni Sartori .: ***Homo videns. La sociedad teledirigida.*** 1997
- ii. Lohr Sharon, Thomson.: ***Muestreo diseño y análisis,*** México 2004
- iii. Perez Lopez Cesar.: ***Técnicas de muestreo estadístico,*** Alfaomega, México 2003
- iv. Freund, J.E .: ***Estadística Matemática con Aplicaciones,*** et al., Pearson Educación, México.
- v. Triola, M.F.: ***Estadística.*** Pearson Educación, México, 2004
- vi. Meyer , Paul L.: ***Probabilidad y Aplicaciones estadísticas.*** México: Fondo Educativo Interamericano. 1973.
- vii. Mendenhall, Scheaffer, Wackerly: ***Estadística Matemática con Aplicaciones.*** México: Editorial Iberoamericana.
- viii. <http://www.rrppnet.com.ar/efectostv.htm> Lorenzo Vilches ***“Usos de la televisión”: diversión o entretenimiento, utilidad social e información.*** Consultada en Mayo 2006.
- ix. http://revista.amec.com.mx/num_8_2004/Jasa_Graciela_Frias_Martha2.htm. Consultada en Mayo 2006.
- x. http://www.COFE_Television_y_audio_restringido. Consultada en Mayo 2006.
- xi. <http://www.inegi.org.mx> Consultada en Mayo 2006.

11. Anexo

Estado	municipio																				
		Ocupantes por vivienda	Con drenaje (%)	Con agua entubada a	Con energía eléctrica (%)	menores de 15 años	población total	Personas de 15 a 64 años(%)	total de viviendas	desarrollo humano	nivel escolar	PIB dólares	escolares de 6-24	alfabetas+15	de 5 y mas años que hablan	residentes en localidades de 2500 hab.	Tasa media de crecimiento	inmigración	Densidad demográfica	sup (KM2)	Suscripciones
Sinaloa	Ahome	4.0	89.0	95.2	98.6	29.0	388344	62.0	93944	0.798	0.855	6038	66.8	94.8	1.33	78.4	1.39	0.55247	96.9	4007	186.6
	Angostura	3.9	86.4	93.0	98.4	28.6	42445	61.8	10563	0.759	0.836	3804	66.8	92.0	1.10	33.2	-0.56	0.32589	23.8	1787	32.9
	Badiraguato	4.5	41.5	33.1	54.2	36.5	32295	55.4	7165	0.636	0.744	1441	56.4	83.4	0.26	11.0	-2.72	0.38300	6.8	4747	34.4
	Concordia	4.0	71.7	86.9	94.4	30.2	27001	59.3	6647	0.772	0.794	6840	60.5	88.9	0.20	30.8	-0.52	0.66207	12.6	2148	12.4
	Cosalá	4.9	65.8	83.6	88.6	36.6	17813	55.9	3558	0.692	0.764	2472	62.7	83.3	0.15	38.3	0.55	0.06754	8	2223	12.4
	Culiacán	4.1	92.2	93.0	97.9	29.7	793730	61.6	186001	0.803	0.841	6970	66.1	93.2	0.85	85.1	1.11	0.24541	126.8	6261	352.0
	Choix	4.3	54.6	64.9	83.4	33.4	31763	56.7	7227	0.671	0.75	1883	59.2	82.9	1.56	25.6	1.40	0.02674	9.9	3211	21.2
	Elota	4.3	64.9	80.2	97.6	35.0	46462	57.5	10631	0.713	0.685	6009	46.7	79.4	9.42	27.9	-1.10	0.52794	28.2	1647	35.1
	Escuinapa	4.1	85.8	85.8	97.7	32.2	49655	59.1	11794	0.763	0.841	3669	68.6	91.9	0.61	80.7	-0.28	0.56635	31.5	1574	14.6
	Fuerte, El	4.2	59.3	87.3	96.9	28.9	92585	60.2	21154	0.763	0.81	5368	62.9	90.1	6.00	40.1	0.60	0.27807	22.1	4188	18.3
	Guasave	4.2	78.4	82.9	99.0	30.3	270260	61.4	63254	0.766	0.816	4907	63.3	90.8	0.99	63.6	-0.46	0.03063	92.6	2919	88.5
	Mazatlán	3.8	94.4	96.3	98.4	28.3	403888	63.0	103534	0.815	0.865	7019	67.3	96.1	0.53	92.4	1.06	0.57934	162.8	2481	138.3

Estado de México.	Mocorito	22.7	2797	15.8	0.28518	-2.17	25.8	1.22	85.7	59.9	3030	0.771	0.711	10696	59.4	44217	29.2	97.6	71.9	75.3	4.1
	Rosario	16.1	2632	18	0.35059	-0.20	40.7	0.46	90.8	65.7	3496	0.824	0.748	11629	59.6	47394	30.0	96.3	86.3	81.3	4.0
	Salvador Alvarado	12.4	788	97.1	0.14690	0.76	87.5	0.37	94.3	69.7	5828	0.861	0.800	18701	61.6	76537	29.8	99.3	94.6	96.1	4.0
	San Ignacio	5.1	5138	4.5	0.38382	-2.37	33.2	1.27	87.5	57.5	3161	0.775	0.714	5511	56.3	23355	32.4	95.2	84.1	73.1	4.0
	Sinaloa	32.9	6313	13.5	0.07623	-0.02	19.3	1.22	84.9	61.6	2362	0.771	0.693	18858	57.8	85017	33.1	93.5	79.4	51.3	4.4
	Navolato	23.4	2473	54.9	0.49427	-1.24	50.7	3.74	87.0	52.3	7409	0.754	0.764	31555	60.7	135681	31.9	96.9	82.8	80.3	4.2
	Acolman	18.0	82	747	0.75157	3.56	91.0	1.00	95.5	65.3	4546	0.855	0.785	12799	59.9	61250	31.2	98.6	95.1	90.4	4.5
	Amecameca	18.0	181.7	249.1	0.43338	2.24	74.5	0.48	94.7	63.4	8274	0.843	0.813	9107	59.4	45255	31.6	98.3	98.1	83.0	4.7
	Apaxco	5.0	79	300.4	0.77697	2.54	79.7	0.39	92.1	63.2	3762	0.825	0.767	4852	58.7	23734	33.7	97.2	95.3	85.0	4.7
	Atenco	6.0	136	253.2	0.68093	5.00	76.6	1.24	95.6	63.3	3528	0.848	0.768	6736	60.0	34435	32.4	99.4	90.1	81.3	4.9
	Atizapán de Zaragoza	509.0	84	5570	0.44715	4.06	99.9	2.13	96.2	67.1	7619	0.865	0.822	104778	63.6	467886	29.2	95.2	98.8	98.8	4.3
	Axapusco	3.0	196	104.7	0.65804	2.66	52.9	0.83	90.0	61.8	2707	0.806	0.729	4230	54.2	20516	32.6	99.7	94.3	68.5	4.5
	Coacalco de Berriozábal	166.0	38	6646	0.60425	5.24	99.9	0.88	98.5	74.2	6238	0.904	0.830	55917	61.5	252555	27.3	99.6	99.0	99.4	4.1
	Cuautitlán	62.0	27	2809	0.78882	4.53	91.4	1.22	96.9	68.7	6083	0.875	0.816	16077	59.1	75836	28.9	98.3	98.8	97.5	4.3
	Chalco	67.0	224	973.1	0.61583	-2.59	96.3	2.86	93.2	64.3	3612	0.836	0.761	43051	55.6	217972	34.2	98.3	92.6	84.8	4.6
	Chiautla	5.0	20	981	0.70876	2.91	49.1	0.77	95.7	64.2	4171	0.852	0.777	3766	59.8	19620	32.3	99.1	89.7	84.6	5.0
	Chicoloapan	36.0	34	2282	0.77796	3.10	99.4	1.77	94.3	62.2	4381	0.836	0.774	16300	60.6	77579	32.5	98.1	97.9	95.2	4.5
	Chiconcuac	1.0	5	3594	0.78568	2.42	95.2	1.25	96.1	63.9	3789	0.853	0.780	3022	58.9	17972	31.2	98.3	96.9	91.4	5.5
	Chimalhuacán	76.0	56	8764	0.59404	7.37	98.9	4.20	93.5	61.4	4462	0.828	0.773	99372	57.6	490772	36.5	99.2	87.7	82.3	4.7
	Huehuetoca	31.0	105	366.3	0.72578	4.21	81.7	1.31	94.5	60.6	4315	0.832	0.776	7570	59.5	38458	33.6	99.0	92.9	89.8	4.7
	Huixquilucan	405.0	143	1353	0.11594	3.93	86.0	2.58	95.5	63.2	10066	0.847	0.830	40417	62.2	193468	29.6	98.6	94.6	92.3	4.5
	Ixtapaluca	175.0	276	1078	0.82139	8.10	97.3	2.11	95.2	66	4598	0.855	0.787	61310	55.0	297570	33.0	99.5	93.0	89.5	4.3
	Naucalpan de Juárez	811.0	151	5687	0.11365	0.89	98.4	3.32	95.8	63	8211	0.849	0.820	199026	64.3	858711	28.3	99.4	98.1	97.9	4.2
	Nezahualcóyotl	384.0	70	17514	0.55771	-0.24	99.9	1.71	96.1	65.9	6151	0.861	0.811	274984	64.0	1225972	29.0	98.7	98.8	99.0	4.3
	Nicolás Romero	147.0	224	1203	0.34300	3.91	94.5	1.79	93.9	61.6	4389	0.832	0.777	56838	60.3	269546	32.2	96.3	94.6	91.3	4.5
	Otumba	8.0	198	147	0.67408	2.93	65.5	0.54	91.2	59.2	2936	0.805	0.735	5825	56.2	29097	33.4	98.0	94.7	71.8	4.7

	Ozumba	9.0	48	491.5	0.67974	2.73	75.9	0.51	93.9	62.7	3098	0.835	0.748	4506	55.8	23592	34.4	98.5	90.0	69.3	5.0
	La Paz	54.0	34	6256	0.70008	4.70	99.3	2.63	95.1	62.6	6031	0.843	0.799	45367	59.9	212694	33.2	98.8	91.1	88.1	4.5
	Tecámac	139.0	152	1137	0.23463	3.47	97.7	1.64	95.5	64.5	4915	0.852	0.793	36443	59.8	172813	31.4	99.1	97.5	94.7	4.5
	Teotihuacán	9.0	76	587.5	0.20289	3.92	83.9	1.67	94.7	63.6	4321	0.844	0.775	9262	59.0	44653	32.3	98.9	91.5	90.4	4.6
	Tepotztlán	26.0	195	319.4	0.69362	4.65	88.5	0.97	94.4	65.6	5345	0.848	0.795	13194	59.4	62280	31.9	99.0	94.8	87.3	4.5
	Texcoco	95.0	404	505.2	0.70406	3.84	94.9	2.16	95.5	68.8	5706	0.866	0.800	41473	61.3	204102	29.9	99.6	93.9	88.0	4.5
	Tlalnepantla de Baz	457.0	20	36071	0.57709	0.26	100.0	1.70	96.3	66.5	6812	0.864	0.818	166006	65.3	721415	26.7	98.9	98.3	98.1	4.2
	Tultepec	31.0	28	3331	0.75761	7.07	93.6	1.33	95.6	66.3	4838	0.858	0.791	19466	58.6	93277	33.2	99.4	98.8	93.1	4.5
	Tultitlán	270.0	13	33242	0.01368	5.82	99.1	1.36	96.7	68.2	4855	0.872	0.802	91633	59.8	432141	30.5	99.2	98.3	95.6	4.4
	Valle de Chalco Solidaridad	53.0	49	6601	0.46592	0.00	99.8	4.25	93.2	60.1	4890	0.821	0.777	66901	58.9	323461	35.6	99.6	99.0	95.0	4.6
	cuautiltan izcalli	422.0	112	4047	0.46426	3.35	99.4	0.87	97.5	71.2	6816	0.887	0.828	100213	64.0	453298	27.8	99.8	98.1	97.9	4.3
Veracruz	Acayucan	14.0	726	107.8	0.45886	1.12	70.8	6.32	82.8	62.8	62.8	0.84	0.600	17804	59.5	78243	35.2	90.2	67.0	77.7	4.4
	Actopan	7.8	857	45.9	0.69126	-0.30	17.3	0.43	82.1	57.8	57.8	0.811	0.605	10226	62.2	39354	29.8	97.0	91.2	81.6	3.8
	Acultzingo	1.0	167	106.5	0.62257	2.42	31.9	22.66	80.2	58.3	58.3	0.761	0.523	3507	54.2	17785	40.0	83.0	60.1	25.2	5.1
	Alvarado	9.0	838	59.1	0.56398	0.09	58.6	0.83	86.2	64.3	64.3	0.842	0.649	12536	63.8	49499	28.7	92.5	69.7	78.4	3.8
	Banderilla	1.3	22	747	0.19157	-2.94	93.8	0.63	91.7	64.1	64.1	0.866	0.670	3838	61.9	16433	33.0	98.3	95.2	92.1	4.2
	Boca del Río	51.5	39	3482	0.57661	-0.63	99.2	1.12	95.5	68.4	68.4	0.886	0.830	34984	67.3	135804	26.0	99.3	97.2	91.7	3.8
	Catemaco	4.5	671	67.6	0.27392	1.13	52.1	0.98	77.4	60.8	60.8	0.81	0.548	9638	57.5	45383	36.9	87.3	87.6	64.4	4.7
	Cerro Azul	5.8	*	*	0.16960	-1.29	87.0	2.25	91.7	75.6	75.6	0.859	0.682	6302	62.0	24729	29.5	93.8	94.0	85.6	3.9
	Coatepec	9.5	198	371.4	0.70813	1.77	80.1	0.43	89.8	60.3	60.3	0.841	0.665	17206	63.4	73536	29.6	97.4	97.4	92.8	4.2
	Coatzacoalcos	80.8	308	867.6	0.66996	1.38	97.5	3.19	93.6	68.2	68.2	0.874	0.722	67668	64.9	267212	30.5	97.5	90.9	94.4	3.9
	Coatzintla	4.0	235	166.8	0.73533	1.37	66.4	10.88	88.8	69.9	69.9	0.837	0.635	9262	60.9	39189	33.0	90.3	76.6	71.2	4.2
	Comapa	0.3	320	53.4	0.59415	1.37	25.4	0.11	75.4	52.4	52.6	0.729	0.436	3187	54.8	17094	39.7	85.4	59.7	30.7	5.4
	Córdoba	16.0	139	1276	0.18123	1.67	85.1	1.97	92.8	65.6	65.6	0.869	0.687	42744	63.4	177288	30.0	98.0	85.9	90.7	4.1
	Cosamaloapan	7.8	581	93.3	0.80582	-3.45	70.7	4.56	86.4	64.1	64.1	0.848	0.646	14043	62.6	54185	29.3	94.8	76.7	83.1	3.8
	Cosoleacaque	10.0	234	416.4	-	7.68	0.0	44.51	89.2	67.6	67.6	0.854	0.680	22965	61.4	97437	34.6	91.9	57.4	88.0	4.2

					0.71621																
					-																
Cuichapa	0.5	70	155	0.74800	0.44	56.7	5.53	84.8	61.9	61.9	0.813	0.564	2362	57.5	10849	36.5	93.6	89.2	73.2	4.6	
				-																	
Chalma	2.3	199	64.8	0.37208	-0.64	32.1	0.44	71.0	70.3	70.3	0.751	0.497	2748	56.4	12902	36.6	85.3	14.5	24.3	4.7	
				-																	
Las Choapas	29.5	3158	23.1	0.64922	-0.51	81.7	1.41	80.9	56.8	56.8	0.8	0.627	15375	53.5	73077	41.7	69.7	37.6	68.7	4.7	
Emiliano Zapata	6.5	421	105.9	1.67481	2.07	54.6	3.87	90.2	57.8	57.8	0.847	0.627	10638	61.3	44580	32.0	96.4	95.3	77.3	4.2	
				-																	
Fortín	10.5	73.2	629.1	0.78108	2.26	23.0	5.44	91.3	63.7	63.7	0.852	0.693	10936	63.1	46053	30.3	95.7	81.2	89.9	4.2	
				-																	
Gutiérrez Zamora	2.5	238	111	0.55883	-0.48	35.9	21.99	87.9	68.9	68.9	0.816	0.581	6367	59.8	26413	32.4	89.7	51.3	65.0	4.1	
				-																	
Huayacocotla	0.3	562	32.2	0.75971	-0.02	27.0	1.04	74.7	60.2	60.2	0.763	0.521	3978	53.4	18093	39.0	59.3	57.3	27.8	4.5	
Hueyapan de Ocampo	2.3	711	56	0.23319	0.39	61.9	4.00	74.2	61.5	61.5	0.779	0.643	9358	57.3	39795	36.7	82.2	65.0	44.2	4.2	
				-																	
Ignacio de la Llave	1.3	397	44.7	0.44807	-1.02	67.5	3.01	73.7	65.2	65.2	0.764	0.535	4462	57.5	17753	34.3	83.5	31.2	61.0	4.0	
				-																	
Isla	1.8	933	41.6	0.64761	2.20	73.1	8.89	81.9	56	56	0.827	0.617	8938	59.1	38847	35.9	90.0	73.5	74.6	4.3	
				-																	
Ixhuatlán del Sureste	1.3	157	84.7	0.73948	1.05	98.1	0.79	86.9	54.9	64.5	0.808	0.608	3005	59.6	13294	36.1	84.9	46.6	80.4	4.4	
				-																	
Ixtaczoquitlán	3.5	115	494.7	0.61255	2.68	80.7	2.12	88.2	62.3	62.3	0.835	0.619	12709	61.6	56896	33.2	94.1	92.4	79.4	4.5	
				-																	
Xalapa	73.0	118	3310	0.82242	3.10	23.9	5.41	94.7	70.5	70.5	0.873	0.734	99875	65.3	390590	28.4	98.9	98.4	94.8	3.9	
				-																	
Jáltipan	4.5	318	118.8	0.63903	-0.24	44.8	4.21	87.7	67.5	67.5	0.851	0.598	9154	60.8	37764	33.1	91.4	52.7	87.9	4.1	
				-																	
Jesús Carranza	4.3	1382	18.4	0.67309	-0.98	22.4	0.37	80.4	59.9	59.9	0.788	0.521	5339	53.1	25424	41.6	86.5	39.4	55.0	4.7	
				-																	
Juan Rodríguez Clara	4.0	986	34	0.64662	0.04	67.2	1.47	79.3	58	58	0.803	0.575	7750	57.6	33495	36.2	88.5	72.0	60.2	4.3	
				-																	
Manlio Fabio Altamirano	3.3	225	91.5	0.66861	0.63	57.0	0.51	86.1	57.6	57.6	0.807	0.555	5261	61.4	20580	31.1	93.9	77.8	43.7	3.9	
				-																	
Martínez de la Torre	11.5	815	146.2	0.59954	1.47	73.3	4.27	88.1	59.6	59.6	0.848	0.640	28709	61.3	119166	32.4	93.2	54.7	76.6	4.1	
				-																	
Medellín	6.3	370	95.1	0.78885	1.86	100.0	2.58	88.8	59.3	59.3	0.825	0.641	8953	62.6	35171	31.2	91.8	52.1	74.2	3.9	
				-																	
Minatitlán	29.0	2116	72.3	0.54408	-2.44	54.4	2.13	91.2	68.2	68.2	0.863	0.654	36850	61.6	153001	32.6	88.7	65.8	82.2	4.1	
				-																	
Orizaba	8.3	30	3953	0.85490	0.38	48.8	24.57	95.4	71.6	71.6	0.885	0.711	29921	64.7	118593	26.1	98.9	98.7	97.4	3.9	
				-																	
Pánuco	6.0	3177	28.5	0.54269	0.33	30.5	0.24	89.7	62.1	62.1	0.831	0.616	21381	60.2	90657	33.4	83.6	69.6	44.8	4.2	
				-																	
Papantla	6.0	1491	114.2	0.54995	0.76	34.3	35.68	84.4	66.9	66.9	0.793	0.575	36892	58.8	170304	35.1	81.9	38.9	51.1	4.6	
				-																	
Paso de Ovejas	8.0	388	79.4	0.77756	0.73	99.1	2.56	88.0	59.7	59.7	0.832	0.616	7625	61.6	30791	30.3	96.4	91.9	64.6	4.0	
				-																	
Playa Vicente	4.3	1567	31.5	0.65342	-0.04	25.6	0.40	75.0	63.6	63.6	0.765	0.531	10486	54.4	49388	39.9	86.0	30.2	43.7	4.7	

	Poza Rica de Hidalgo	74.8	47	3252	0.59993	0.07	49.3	0.70	94.5	71.3	71.3	0.878	0.712	38398	65.9	152838	26.8	97.1	87.2	95.7	4.0
	Puente Nacional	1.3	133	142.8	0.60858	0.69	24.8	1.33	88.8	62.1	62.1	0.848	0.658	4837	60.6	18999	32.3	96.4	92.8	68.3	3.9
	San Andrés Tuxtla	202.5	958	148.6	0.46297	1.35	0.0	77.04	75.0	56.7	56.7	0.79	0.538	30134	57.5	142343	36.7	88.5	75.0	51.1	4.7
	San Juan Evangelista	0.8	*	*	0.52209	-0.14	45.0	3.38	79.7	58.1	58.1	0.774	0.486	7058	56.4	32645	37.3	87.9	40.7	64.1	4.6
	Texcatepec	4.3	*	*	0.67674	1.51	17.5	21.07	51.0	54	54	0.644	0.410	1681	52.5	9051	43.2	52.6	19.6	0.2	5.4
	Texistepec	0.3	447	42.7	0.52625	0.43	73.9	2.04	81.3	60.9	60.9	0.785	0.481	4061	56.5	19066	37.9	78.3	44.6	64.8	4.7
	Tezonapa	2.5	522	97.7	0.50745	-0.26	65.7	0.62	77.0	58.6	58.6	0.741	0.520	10926	54.5	51006	39.4	71.4	29.2	29.9	4.7
	Túxpam	20.3	958	132.2	0.59846	0.67	96.1	1.11	92.2	69.4	69.4	0.849	0.664	30518	61.8	126616	31.7	92.0	60.4	64.7	4.1
	Vega de Alatorre	1.5	340	55.2	0.72186	0.36	25.0	2.29	84.8	59.8	59.8	0.834	0.595	4802	59.4	18771	33.0	93.9	89.8	82.8	3.9
	Veracruz	139.3	243	1882	0.57065	3.39	85.9	1.46	95.5	68.1	68.1	0.883	0.734	122498	66.2	457377	26.9	98.4	92.2	88.9	3.7
	Yecuatla	1.5	154	81.2	0.70185	-0.75	43.4	1.33	77.9	57	57	0.772	0.504	2644	54.5	12500	38.5	84.1	79.4	67.6	4.7
	Agua Dulce	19.0	373	118.2	0.61715	-0.69	95.2	3.76	91.2	69.9	69.9	0.857	0.663	10727	61.1	44100	33.0	93.3	55.9	92.6	4.1
	El Higo	2.0	462	39.9	0.65283	-1.49	53.8	9.21	86.4	65.8	65.8	0.838	0.623	4309	60.0	18446	32.4	88.9	94.6	44.6	4.3
	Nanchital de Lázaro Cárdenas del Río	1.5	64	425.3	0.56138	0.19	0.0	38.99	94.1	72.8	72.8	0.881	0.700	6710	64.3	27218	31.4	97.2	92.5	97.1	4.0
	Tres Valles	0.3	541	81.7	0.52220	0.38	*	*	83.9	61.2	61.2	0.818	0.619	10599	58.0	44215	35.7	90.8	56.3	75.7	4.1
	Uxpanapa	2.8	1938	12.1	0.64520	0.00	*	*	79.0	59.1	59.1	0.745	0.446	4238	51.6	23461	44.9	71.9	50.3	36.9	5.5
Distrito Federal	Azcapotzalco	359.0	34	12971	0.67905	-0.74	100.0	1.27	97.7	72.4	16203	0.893	0.873	109233	67.5	441008	24.2	99.7	99.3	99.1	4.0
	Coyoacán	1287.0	54	11860	0.61936	0.01	100.0	1.92	97.8	74	24943	0.899	0.901	163036	69.5	640423	22.6	99.4	99.2	99.0	3.9
	Cuajimalpa de Morelos	249.0	70	2160	0.60763	2.38	99.0	1.52	96.4	66.3	21927	0.864	0.874	33163	64.3	151222	29.6	99.7	98.1	97.2	4.4
	Gustavo A. Madero	752.0	88	14040	0.59748	-0.26	100.0	1.52	97.0	69.8	14556	0.879	0.860	295329	66.7	1235542	25.9	99.4	99.0	99.0	4.1
	Iztacalco	327.0	23	17884	-0.6516	-0.86	100.0	1.45	97.4	71.4	15027	0.887	0.866	98234	66.6	411321	25.2	99.5	99.0	99.0	4.1
	Iztapalapa	957.0	114	15556	0.61859	1.77	100.0	2.04	96.4	67	12184	0.866	0.846	403922	65.5	1773343	29.6	99.4	98.1	98.5	4.3
	La Magdalena Contreras	340.0	64	3470	0.64518	1.31	99.8	1.87	96.5	67.7	18356	0.869	0.870	51831	65.6	222050	27.5	99.0	97.7	96.4	4.2
	Milpa Alta	28.0	287	337.2	0.74729	4.31	93.8	4.53	94.4	67.4	8206	0.854	0.815	21350	61.9	96773	32.7	99.5	93.0	85.7	4.5
	Álvaro Obregón	796.0	96	7157	0.66575	0.67	100.0	1.68	96.6	67.5	21315	0.869	0.881	163481	66.9	687020	26.4	99.5	98.7	99.0	4.2

Tláhuac	181.0	86	3521	-	0.67799	3.92	99.5	1.51	96.5	68.8	11582	0.873	0.844	69564	63.6	302790	31.6	99.4	98.6	97.1	4.3
Tlalpan	1020.0	310	1877	-	0.61989	1.85	99.0	2.10	96.9	70.5	20015	0.881	0.882	140148	67.3	581781	26.8	99.2	90.9	97.0	4.1
Xochimilco	393.0	119	3108	-	0.71882	3.17	98.6	2.66	96.2	69.9	14806	0.875	0.862	82078	65.1	369787	29.0	99.8	94.1	91.8	4.4
Benito Juárez	723.0	27	13351	-	0.61986	-1.23	100.0	1.80	98.9	77.6	35594	0.918	0.930	113741	69.3	360478	17.7	98.9	99.0	99.3	3.1
Cuauhtémoc	879.0	33	15644	-	0.65708	-1.44	100.0	2.06	97.9	71.5	20018	0.891	0.888	147181	66.9	516255	22.4	99.8	98.3	98.3	3.4
Miguel Hidalgo	641.0	46	7666	-	0.56226	-1.43	100.0	1.82	98.1	71	27819	0.891	0.904	94475	67.6	352640	20.8	99.7	99.0	99.1	3.6
Venustiano Carranza	333.0	34	13612	-	0.68519	-1.16	100.0	1.39	97.7	71.2	15032	0.889	0.866	116986	66.2	462806	24.9	99.7	99.2	99.1	3.9

