DOCUMENTOS DE INVESTIGACIÓN

Facultad de Administración

No. 147, ISSN: 0124-8219
Noviembre de 2013

# Inferencia estadística
# Módulo de regresión lineal simple

Diego Fernando Cardona Madariaga
Javier Leonardo González Rodríguez
Miller Rivera Lozano
Edwin Cárdenas Vallejo

Universidad del Rosario
Facultad de Administración

# Inferencia estadística
# Módulo de regresión lineal simple

Documento de investigación No. 147

Diego Fernando Cardona Madariaga
Javier Leonardo González Rodríguez
Miller Rivera Lozano
Edwin Cárdenas Vallejo

Universidad del Rosario
Escuela de Administración
Editorial Universidad del Rosario
Bogotá D.C.
2013

Inferencia estadística módulo de regresión lineal simple / Diego Fernando Cardona Madariaga… [y otros autores]. —Bogotá: Editorial Universidad del Rosario, Escuela de Administración, 2013.
57 páginas.—(Borrador de investigación; 147)
ISSN: 0124-8219

Estadística matemática / Análisis de regresión / Probabilidades / Matemáticas / I. González Rodríguez, Javier Leonardo / II. Rivera Lozano, Miller / III. Cárdenas Vallejo, Edwin/ IV. Título / V. Serie.

519.536 SCDD 20

Catalogación en la fuente – Universidad del Rosario. Biblioteca

amv Octubre 10 de 2013

Diego Fernando Cardona Madariaga
Javier Leonardo González Rodríguez
Miller Rivera Lozano
Edwin Cárdenas Vallejo

Corrección de estilo
Claudia Ríos

Diagramación
Fredy Johan Espitia Ballesteros

Editorial Universidad del Rosario
http://editorial.urosario.edu.co

ISSN: 0124-8219



Primera edición: Noviembre de 2013
Hecho en Colombia
*Made in Colombia*

# Contenido

## Figuras



## Tablas

# Inferencia estadística
# Módulo de regresión lineal simple

Diego Fernando Cardona Madariaga*
Javier Leonardo González Rodríguez**
Miller Rivera Lozano***
Edwin Cárdenas Vallejo****

## Resumen

La utilización del modelo de regresión lineal en los procesos relacionados con el análisis de datos demanda el conocimiento objetivo e instrumentación de la relación funcional de variables, el coeficiente de determinación y de correlación y la prueba de hipótesis como pilares fundamentales para verificar e interpretar su significancia estadística en el intervalo de confianza determinado.

La presentación específica de los temas relacionados con el modelo de regresión lineal, el análisis de regresión, el uso de la ecuación de regresión como instrumento para estimar y predecir y la consideración del análisis de residuales ha sido realizada tomando como referente el estudio de problemas reales definidos en los entornos de la economía, la administración y la salud, utilizando como plataforma de apoyo la hoja de cálculo Excel®.

Se consideran en este módulo didáctico, los elementos teóricos correspondientes al análisis de regresión lineal, como técnica estadística empleada para estudiar la relación entre variables determinísticas o aleatorias que resultan de algún tipo de investigación, en la cual se analiza el comportamiento de dos variables, una dependiente y otra independiente.

* Profesor titular de carrera. Universidad del Rosario.
** Profesor principal de carrera. Universidad del Rosario.
*** Director del Laboratorio de Modelamiento y Simulación. Universidad del Rosario.
**** Profesor de la Secretaría de Educación del Distrito.

Se muestra mediante la gráfica de dispersión el posible comportamiento de las variables: lineal directa, inversa, no lineal directa o no lineal inversa, con el fin de desarrollar en el lector las competencias interpretativas y propositivas requeridas para dimensionar integralmente la importancia de la estadística inferencial en la vida del profesional en ciencias económicas, administrativas y de la salud.

# 1. Introducción

En muchas investigaciones estadísticas tendientes a la toma de decisiones de tipo profesional o personal uno de los objetivos principales es establecer relaciones que permitan pronosticar una o más variables en términos de otras. Por ejemplo, se han efectuado estudios de la reducción del peso de una persona en términos del número de semanas que ha seguido una dieta específica; también, sobre el consumo per cápita de ciertos artículos alimenticios en términos de su valor nutricional. En otro caso, una empresa de energía eléctrica en una ciudad como Cartagena o Barranquilla podría determinar la relación entre la temperatura máxima diaria y la demanda de electricidad, para predecir el consumo de energía con base en las temperaturas máximas pronosticadas para el mes siguiente.

Algunos administradores confían en su intuición para juzgar cómo se relacionan dos variables. Sin embargo, si los responsables en la toma de decisiones pueden tomar datos y utilizar un procedimiento estadístico de análisis para determinar cómo lo conocido se relaciona con el evento futuro, pueden ayudar considerablemente en el mejoramiento de los procesos que administran.

El procedimiento estadístico que se utiliza para este fin se conoce como **análisis de regresión,** el que permite establecer la relación funcional o ecuación matemática que relaciona las variables, así como la fuerza de esa relación.

El término regresión fue utilizado por primera vez como un concepto estadístico en 1877 por Sir Francis Galton, quien llevó a cabo un estudio que mostró que la estatura de los niños nacidos de padres altos tiende a retroceder o "regresar" hacia la estatura media de la población. Designó la palabra regresión como el nombre del proceso general de predecir una variable (la estatura de los niños) a partir de otra (la estatura del padre o de la madre). Más tarde, los estadísticos acuñaron el término regresión múltiple para describir el proceso mediante el cual se utilizan varias variables para predecir otra (Devore, 2005).

En la terminología de la regresión, la variable que se va a predecir se llama dependiente. La o las variables que se usan para predecir el valor de la variable dependiente se llaman variables independientes.

En general, existen cuatro posibles formas en que las variables se pueden relacionar, a saber: relación lineal directa, relación lineal inversa, relación no

lineal directa y relación no lineal inversa, cuya estructura formal y funcional permite dilucidar con objetividad las actividades orientadas a decidir qué ecuación se debe emplear, cuál ha de ser la ecuación que mejor se ajusta a los datos y cómo debe validarse la significancia de los pronósticos realizados.

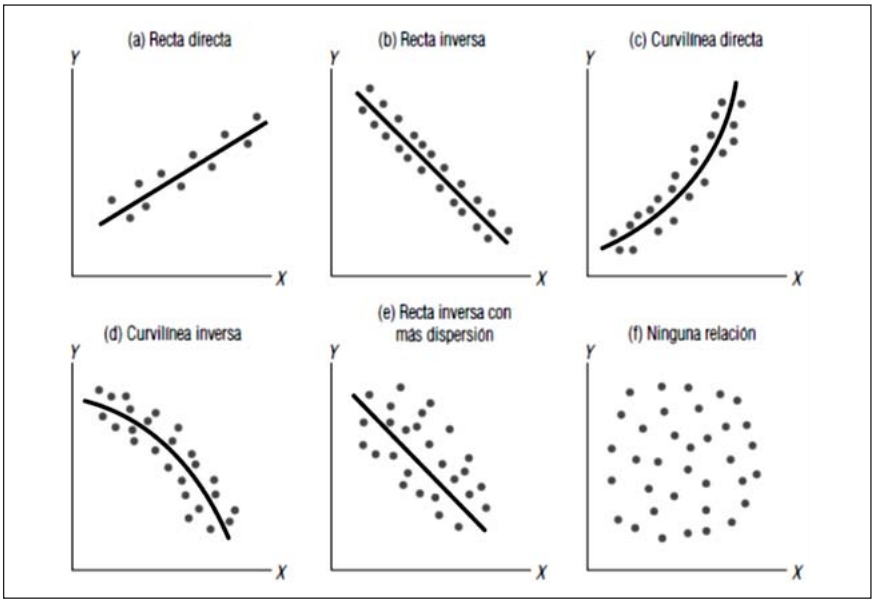
En este trabajo se describirá el análisis de regresión donde intervienen una variable dependiente y una independiente y en la cual la relación entre ellas se aproxima por medio de una línea recta. A esto se le llama regresión lineal simple.

En próximos documentos se tratará el análisis de regresión donde intervienen dos o más variables independientes, llamada regresión múltiple. De igual manera se trabajará el caso en que la relación de las variables no es lineal.

## 2. El modelo de regresión lineal simple

El atender problemas relacionados con los sistemas de representación funcional y el comportamiento de las variables demanda el estar familiarizado con cada uno de los casos que se señalan en la figura 1 y que apropiadamente se explican para orientar al lector en el proceso del cálculo de la línea de regresión, precisándose en primera instancia revisar el concepto de función lineal, para luego abordar con propiedad el modelo de regresión lineal simple, debiéndose considerar la ecuación estimada y el método de los mínimos cuadrados.

**Figura 1. Tipos de relación entre dos variables**




Fuente: Levin y Rubin, 2004.

En esta figura, el segmento (e) ilustra una relación lineal inversa con un patrón de puntos ampliamente disperso. Esta mayor dispersión indica que existe menor grado o fuerza de asociación entre las variables. El patrón de puntos señalado en el segmento (f) de la misma figura parece indicar que no existe relación entre las dos variables, por tanto, conocer el pasado referente a una variable no nos permitirá pronosticar ocurrencias futuras de la otra.

En los diagramas de dispersión que se mostraron en la figura 1 se pusieron las líneas de regresión ajustando las líneas visualmente entre los puntos de datos. En esta sección aprenderemos a calcular la línea de regresión de manera más precisa, usando una ecuación que relaciona las dos variables matemáticamente. En primera instancia se debe revisar el concepto de función lineal.

## La ecuación de la recta

La ecuación para una línea recta donde la variable dependiente Y está determinada por la variable independiente X es:

$$Y = a + bX \quad \textbf{(1)}$$

Donde a representa la "ordenada Y" porque su valor es el punto en el cual la línea de regresión cruza el eje Y, es decir, el eje vertical.

La b en la ecuación (1) es la "pendiente" de la recta. Representa qué tanto cambia la variable dependiente Y por cada unidad de incremento de la variable independiente X. También se conoce como razón de cambio.

$$b = \frac{\Delta Y}{\Delta X} \quad \textbf{(2)}$$

Tanto *a* como *b* son constantes numéricas porque para cualquier línea recta dada sus valores no cambian.

*Ejemplo 1*

Sean a = 4 y b = 2. Determinemos cuál sería Y para X igual a 5. Al sustituir los valores de a, b y X en la ecuación (1), encontramos que el valor correspondiente de Y es:

$$Y = a + bX$$

$$Y = 4 + 2(5)$$

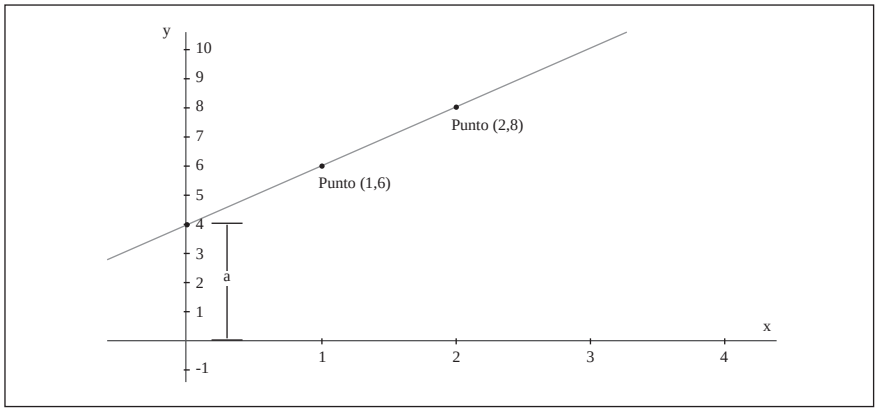
$$= 4 + 10$$

$$= 14$$

Ahora bien, ¿cómo se pueden obtener los valores de a y de b a partir de los puntos? Para ilustrar este proceso, se usará la recta de la figura 2. Para ello, se visualiza la ordenada en el origen y localizando el punto donde la recta cruza a este eje. En la figura 2, se observa, que esto sucede cuando a =4. Para encontrar la pendiente de la recta, b se debe determinar cómo cambia la variable dependiente, Y, al cambiar la variable independiente, X. Para esto se empieza por elegir dos puntos sobre la línea de la figura 12-6. Ahora, debemos encontrar los valores de X y Y (las coordenadas) de ambos puntos. Podemos llamar a las coordenadas de nuestro primer punto (X1, Y1) y (X2, Y2) a las del segundo.

La lectura de la figura 2 permite observar que (X1, Y1) = (1, 6) y (X2, Y2) = (2, 8).

**Figura 2. Ejemplo de función lineal**




Fuente: Elaboración propia.

Ahora, se puede calcular el valor de b, para ello se usa la ecuación (2), expresada de la forma:

$$b = \frac{Y_2 - Y_1}{X_2 - X_1} \quad \textbf{(3)}$$

$$b = \frac{8 - 6}{2 - 1}$$

$$b = \frac{2}{1}$$

$$b = 2$$

Así, entonces se pueden conocer los valores de las constantes numéricas a y b, y escribir la ecuación de la recta. La línea de la figura 2 puede describirse por la ecuación (1), en la que a =4 y b = 2. Por tanto,

$$Y = a + bX$$

$$Y = 4 + 2X$$

Si se sustituyen más valores de *X* en la ecuación, se observará que *Y* se incrementa al aumentar *X*. Por tanto, la relación entre las variables es directa y la pendiente es positiva.

Como se puede observar en la figura 2, todos los puntos que satisfacen la ecuación de la recta están efectivamente sobre la línea. Lo que se hace es encontrar una recta que pase "en medio" de todos los puntos, es decir, que se ajuste de la mejor manera a los puntos.

## El modelo de regresión lineal simple

Con el fin de estudiar este modelo, se emplearán los datos tomados de una muestra real, extraídos de un comunicado de prensa que revela el porcentaje de pobreza, pobreza extrema y el coeficiente de Gini (indicador de la desigualdad económica en una población) en los años 2010 y 2011 de las trece principales ciudades de Colombia., los cuales se presentan en la tabla 1 (DANE, 2012).

**Tabla 1. Porcentaje de pobreza en las principales ciudades de Colombia**

| Pobreza, pobreza extrema y Gini por ciudades, 2010-2011 | | | | | | |
|---|---|---|---|---|---|---|
| Dominio | Nueva metodología | | | | | |
| | Pobreza | | Pobreza extrema | | Gini | |
| | 2010 | 2011 | 2010 | 2011 | 2010 | 2011 |
| Pasto | 43,2 | 40,6 | 11,7 | 8,8 | 52,3 | 52,2 |
| Montería | 39,7 | 37,5 | 6,7 | 6,5 | 52,5 | 53,0 |
| Barranquilla | 39,5 | 34,7 | 7,4 | 5,3 | 49,7 | 47,2 |
| Cúcuta | 39,3 | 33,9 | 8,4 | 5,7 | 47,9 | 47,1 |
| Cartagena | 34,2 | 33,4 | 6,2 | 4,7 | 48,9 | 48,8 |

Continúa

| | | | | | | |
|---|---|---|---|---|---|---|
| Cali | 26,1 | 25,1 | 6,4 | 5,2 | 52,9 | 50,4 |
| Villavicencio | 25,4 | 23,0 | 4,8 | 4,0 | 46,7 | 46,7 |
| Ibagué | 26,6 | 22,0 | 4,3 | 2,7 | 49,5 | 44,9 |
| Pereira | 26,8 | 21,6 | 3,8 | 2,2 | 45,6 | 45,1 |
| Manizales | 23,8 | 19,2 | 4,7 | 2,3 | 49,5 | 47,1 |
| Medellín | 22,0 | 19,2 | 5,6 | 4,0 | 53,8 | 50,7 |
| Bogotá | 15,5 | 13,1 | 2,6 | 2,0 | 52,6 | 52,2 |
| Bucaramanga | 10,9 | 10,7 | 1,2 | 1,1 | 45,0 | 44,9 |

Fuente: DANE, 2012.

Para los efectos explicativos pertinentes a este documento se considera únicamente la variable pobreza, donde *X* será el porcentaje de pobreza en 2010 y *Y* será el porcentaje de pobreza en 2011. Por tanto, lo que nos interesa es mostrar si el porcentaje de pobreza en 2011 depende linealmente del porcentaje de pobreza en 2010 para estas ciudades.

El análisis de los datos de pobreza extrema y Gini se dejará como ejercicio para el lector.

La ecuación general que describe la relación entre las dos variables es:

$$y = \alpha + \beta x + \epsilon \quad \textbf{(4)}$$

En este modelo, y es una función lineal de x (la parte $\alpha + \beta x$) más $\epsilon$ (letra griega épsilon) que representa el error *y* es una variable aleatoria. El término de error explica la variabilidad en *y* que no se puede explicar con la relación lineal (Anderson, Sweeney y Williams, 2001).

## La ecuación estimada de regresión

Infortunadamente, los valores de los parámetros $\alpha$ y $\beta$ de la ecuación (4) no se conocen en la práctica y se deben estimar usando los datos de la muestra. Se calculan los estadísticos de la muestra (denotados a y b) como estimadores de los parámetros $\alpha$ y $\beta$, respectivamente. En la regresión lineal simple, la ecuación estimada de regresión se escribe:

$$\hat{y} = a + bx \quad \textbf{(5)}$$

La gráfica de la ecuación de regresión se llama línea de regresión estimada, donde a es la ordenada en el origen y b es la pendiente y $\hat{y}$ es el valor estimado de y para determinado valor de x.

## El método de lo mínimos cuadrados

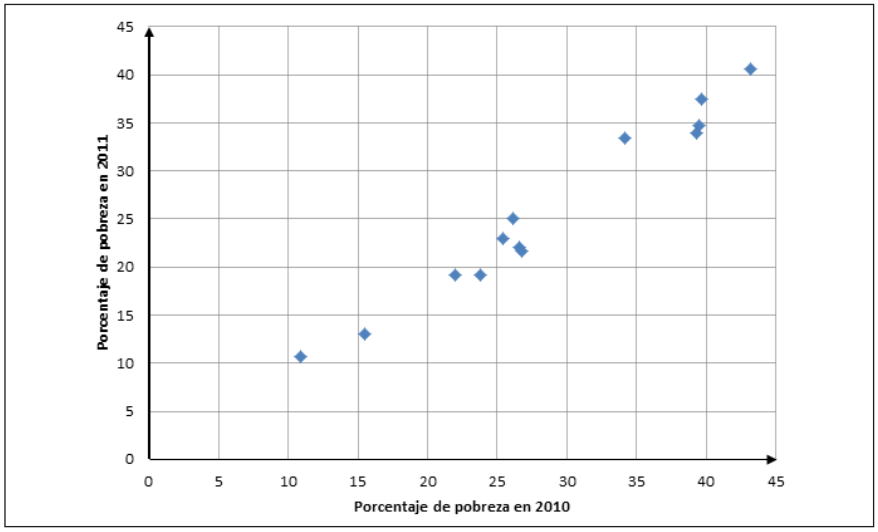
Este método es un procedimiento para encontrar la ecuación de regresión y se remonta al inicio del siglo XIX por el trabajo del matemático francés Adrien Legendre. Para ilustrarlo con el ejemplo del porcentaje de pobreza en las ciudades colombianas (tabla 1), tenemos una muestra de trece ciudades. Para la i-ésima ciudad de la muestra, $x_i$ es el porcentaje de pobreza de esa ciudad en 2010 y $y_i$ es el porcentaje de pobreza de esa ciudad en 2011. En la tabla se observa que $x_3$=39,5 y $y_3$=34,7 corresponden a los porcentajes de pobreza en la ciudad de Barranquilla en 2010 y 2011, respectivamente.

La figura 3 es un diagrama de dispersión de los datos de la tabla 1. Los valores del porcentaje de pobreza en 2010 se representan en el eje horizontal y los valores del porcentaje de pobreza en 2011 se representan en el eje vertical. El diagrama de dispersión nos permite observar gráficamente los datos y sacar conclusiones preliminares acerca de la posible relación entre las variables.

**¿Qué conclusiones se pueden entonces formular al interpretar la figura 3?** Parece que, conforme aumenta la pobreza en ciertas poblaciones en 2010, también aumenta la pobreza en el año 2011, lo cual indica una relación directa entre las variables. Además se observa que los puntos parecen aproximarse a una línea recta. En consecuencia, elegimos el modelo de regresión lineal simple para representar la relación entre las variables.

Para que la línea estimada de regresión se ajuste bien a los datos se desea que las diferencias entre los valores observados de $y$ ($y_i$) y los valores estimados de $y$ ($\hat{y}$) sean mínimas.

**Figura 3. Diagrama de dispersión**




Fuente: Elaboración propia.

## Criterio de los mínimos cuadrados

Este método emplea los datos de la muestra para determinar las características de la recta que hacen mínima la suma de los cuadrados de las desviaciones:

$$min \sum (y_i - \hat{y}_i)^2 \quad \textbf{(6)}$$

Siendo:

$y_i$ = valor observado de la variable dependiente para la i-ésima observación.

$\hat{y}_i$ = valor estimado de la variable dependiente para la i-ésima observación.

$$\sum (y_i - \hat{y}_i)^2 = \sum [y_i - (a + bx_i)]^2 \quad \textbf{(7)}$$

Minimizar el miembro derecho de la ecuación (7) implica calcular las derivadas parciales de la expresión con respecto a los coeficientes de regresión *a* y *b* e igualar a cero las dos derivadas. Al finalizar este procedimiento se llega a las siguientes ecuaciones, conocidas como ecuaciones normales (Walpole y Myers, 1999).

Ecuaciones normales

$$\sum y_i = na + b\sum x_i \textbf{ (8)}$$

$$\sum x_i y_i = a\sum x_i + b\sum x_i^2 \textbf{ (9)}$$

Donde n es el número de observaciones.

Al resolver algebraicamente el sistema de ecuaciones anterior se obtienen las soluciones para a y b.

Complementariamente, y para los fines pertinentes, se hace necesario tener presente las siguientes fórmulas.

$$\bar{x} = \frac{\sum x_i}{n} \quad \bar{y} = \frac{\sum y_i}{n}$$

$$s_{xx} = \sum x_i^2 - \frac{(\sum x_i)^2}{n} \quad s_{yy} = \sum y_i^2 - \frac{(\sum y_i)^2}{n} \ (9a)$$

$$s_{xy} = \sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}$$

Pendiente e intercepto para la ecuación estimada de regresión

$$b = \frac{s_{xy}}{s_{xx}} \textbf{ (10)}$$

$$a = \bar{y} - b\bar{x} \textbf{ (11)}$$

Algunos de los cálculos necesarios para determinar la ecuación estimada de regresión por mínimos cuadrados, en el caso de los porcentajes de pobreza en las ciudades colombianas, aparecen en la tabla 2. En este ejemplo hay trece observaciones, en consecuencia, n=13. Aplicando las ecuaciones (10) y (11) y con la información de la tabla 2 podemos determinar la pendiente y la ordenada al origen de la ecuación (5).

**Tabla 2. Cálculos para la ecuación de regresión lineal**

| Cálculos para la ecuación estimada de regresión | | | | |
|---|---|---|---|---|
| | Año 2010 | Año 2011 | | |
| Observación | $x_i$ | $y_i$ | $x_i y_i$ | $x_i^2$ |
| 1 | 43,2 | 40,6 | 1753,92 | 1866,24 |
| 2 | 39,7 | 37,5 | 1488,75 | 1576,09 |
| 3 | 39,5 | 34,7 | 1370,65 | 1560,25 |
| 4 | 39,3 | 33,9 | 1332,27 | 1544,49 |
| 5 | 34,2 | 33,4 | 1142,28 | 1169,64 |
| 6 | 26,1 | 25,1 | 655,11 | 681,21 |
| 7 | 25,4 | 23 | 584,2 | 645,16 |
| 8 | 26,6 | 22 | 585,2 | 707,56 |
| 9 | 26,8 | 21,6 | 578,88 | 718,24 |
| 10 | 23,8 | 19,2 | 456,96 | 566,44 |
| 11 | 22 | 19,2 | 422,4 | 484 |
| 12 | 15,5 | 13,1 | 203,05 | 240,25 |
| 13 | 10,9 | 10,7 | 116,63 | 118,81 |
| Totales | 373 | 334 | 10690,3 | 11878,38 |

Fuente: Elaboración propia.

$$S_{xx} = 11878{,}38 - \frac{373^2}{13}$$

$$S_{xx} = 11878{,}38 - 10702{,}23 = 1176{,}15$$

$$S_{xy} = 10690{,}3 - \frac{373 * 334}{13}$$

$$S_{xy} = 10690{,}3 - 9583{,}23 = 1107{,}069$$

$$b = \frac{1107{,}069}{1176{,}15} = 0{,}94126$$

$$\overline{x} = \frac{373}{13} = 28{,}69 \quad \overline{y} = \frac{334}{13} = 25{,}69$$

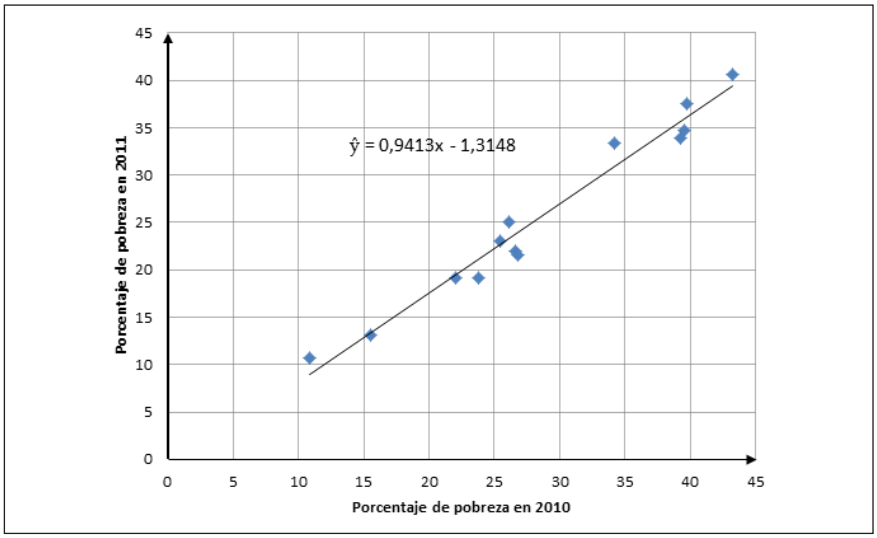
$$a = 25{,}69 - 0{,}94126 * 28{,}69 = -1{,}3148$$

Por lo anterior, la ecuación estimada de regresión es:

$$\hat{y} = -1{,}3148 + 0{,}94126\,x$$

La figura 4 muestra la gráfica de esta ecuación sobre el diagrama de dispersión.

**Figura 4. Gráfica de la ecuación de regresión lineal**




Fuente: Elaboración propia.

La pendiente de la recta es positiva, lo que implica que en las ciudades donde se observó mayor pobreza en 2010 también se notó mayor pobreza en 2011. Pero como la pendiente es un número entre cero y uno, significa que el incremento en el porcentaje de pobreza en 2011 entre una ciudad y otra es menor que en 2010.

Ahora bien, si creemos que esta ecuación describe de la mejor forma posible la relación entre x e y, parece razonable usarla para predecir el porcentaje de pobreza de una ciudad en 2011 si se conoce el valor de 2010. Por ejemplo, si se supiera que Armenia presentó en 2010 un nivel de pobreza del 25,3%; entonces, podríamos estimar el nivel de pobreza en 2011.

$$\hat{y} = -1{,}3148 + 0{,}94126(25{,}3) = 22{,}5\%$$

Sin embargo es necesario verificar y evaluar con otros métodos lo adecuado de esta ecuación para estimar y predecir.

# 3. Análisis de regresión

En el capítulo anterior se empleó la ecuación estimada de regresión, tratando los pronósticos como promedios o valores esperados, por lo tanto se exige entonces ahora el responder estas preguntas:

1. ¿Qué tan buenos son los valores obtenidos para *a* y *b* en la ecuación de regresión $\hat{y}$ = 0,9413x - 1,3148?
2. ¿Cómo podemos estar seguros de que la estimación $\hat{y}$ = 22,5% para el nivel esperado de pobreza en la ciudad de Armenia en el año 2011 será realmente buena?

## Coeficiente de determinación

Con respecto a la primera pregunta, el coeficiente de determinación es una medida de la bondad de ajuste para una ecuación de regresión.

Para la i-ésima observación de la muestra, la desviación entre el valor observado de la variable dependiente $y_i$ y el valor estimado de la variable dependiente $\hat{y}_i$, se llama i-ésimo residual. Representa el error que se comete al usar $\hat{y}_i$ para estimar $y_i$. La suma de los cuadrados de esos residuales es lo que se minimiza en el método de mínimos cuadrados. También se le conoce como la suma de los cuadrados debidos al error (SSE):

$$SSE = \sum(y_i - \hat{y}_i)^2 \quad \textbf{(12)}$$

El valor de SSE es una medida del error que se comete al usar la ecuación de regresión para calcular los valores de la variable dependiente en la muestra.

Otro valor de importancia es la medida del error incurrido al usar $\bar{y}$ para estimar $y_i$, llamado suma total de cuadrados (SST):

$$SST = \sum(y_i - \bar{y})^2 \quad \textbf{(13)}$$

La tabla 3 proporciona los cálculos necesarios para encontrar estas medidas.

Para saber cuánto se desvían los valores de $\hat{y}_i$ medidos en la línea de regresión de los valores de $\bar{y}$, se calcula otra suma de cuadrados. A esa suma se le llama suma de cuadrados debida a la regresión, y se representa por SSR.

$$SSR = \sum(\hat{y}_i - \bar{y})^2 \quad \textbf{(14)}$$

Existe una relación entre las tres sumas:

$$SST = SSR + SSE \quad \textbf{(15)}$$

Ahora bien, es posible entender cómo se pueden emplear las tres sumas de cuadrados para suministrar una medida de la bondad de ajuste para la ecuación de regresión. Esa ecuación tendría un ajuste perfecto si cada valor observado de la variable independiente estuviera sobre la línea de regresión. En este caso, cada diferencia $y_i - \hat{y}_i$ sería cero, por tanto, SSE=0. De la ecuación (15) se tendría que SST=SSR y, por consiguiente, la relación SSR/SST sería igual a 1 como el máximo ajuste. De manera análoga, los ajustes menos perfectos darán como resultado mayores valores de SSE. En consecuencia, de (15) se deduce que el máximo valor de SSE se tiene cuando SSR es cero.

**Tabla 3. Cálculos para el análisis de regresión**

| | Año 2010 | Año 2011 | | Residuales | | | |
|---|---|---|---|---|---|---|---|
| Observación | $x_i$ | $y_i$ | $\hat{y}_i$ | $y_i - \hat{y}_i$ | $(y_i - \hat{y}_i)^2$ | $y_i - \bar{y}$ | $(y_i - \bar{y})^2$ |
| 1 | 43,2 | 40,6 | 39,35 | 1,252 | 1,568 | 14,908 | 222,239 |
| 2 | 39,7 | 37,5 | 36,05 | 1,447 | 2,092 | 11,808 | 139,422 |
| 3 | 39,5 | 34,7 | 35,87 | -1,17 | 1,358 | 9,0077 | 81,1385 |
| 4 | 39,3 | 33,9 | 35,68 | -1,78 | 3,158 | 8,2077 | 67,3662 |
| 5 | 34,2 | 33,4 | 30,88 | 2,523 | 6,368 | 7,7077 | 59,4085 |
| 6 | 26,1 | 25,1 | 23,25 | 1,848 | 3,414 | -0,592 | 0,35083 |
| 7 | 25,4 | 23 | 22,59 | 0,407 | 0,165 | -2,692 | 7,24852 |
| 8 | 26,6 | 22 | 23,72 | -1,72 | 2,968 | -3,692 | 13,6331 |
| 9 | 26,8 | 21,6 | 23,91 | -2,31 | 5,341 | -4,092 | 16,747 |
| 10 | 23,8 | 19,2 | 21,09 | -1,89 | 3,562 | -6,492 | 42,1501 |
| 11 | 22 | 19,2 | 19,39 | -0,19 | 0,037 | -6,492 | 42,1501 |
| 12 | 15,5 | 13,1 | 13,27 | -0,17 | 0,031 | -12,59 | 158,566 |
| 13 | 10,9 | 10,7 | 8,945 | 1,755 | 3,08 | -14,99 | 224,769 |
| Totales | 373 | 334 | | | SSE=33,14 | | SST=1075,19 |
| Promedio | 28,69231 | 25,69231 | | | | | SSR=1042 |

Fuente: Elaboración propia.

La relación SSR/SST, que asume valores entre cero y uno, se usa para evaluar la bondad de ajuste de la ecuación de regresión. A esta relación se le llama coeficiente de determinación y se representa por $r^2$.

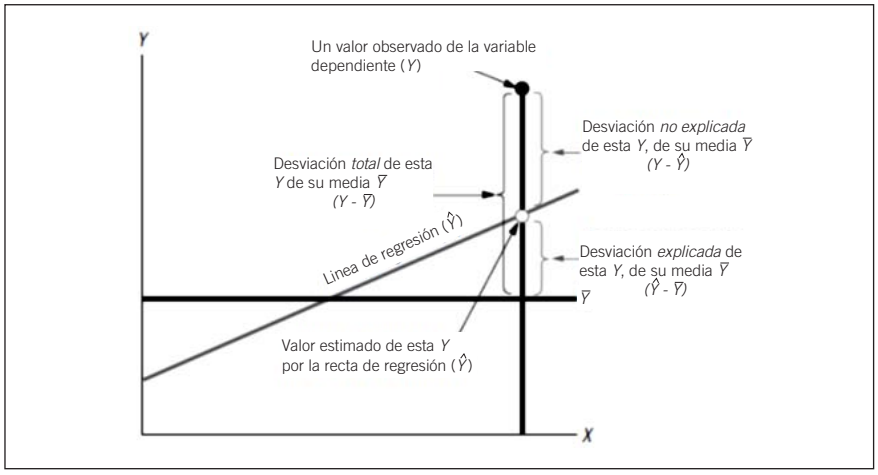
$$r^2 = \frac{SSR}{SST} \quad \textbf{(16)}$$

Para este ejemplo, r=1042/1075=0,9693

Expresando este valor como un porcentaje, se puede interpretar a $r^2$ como el porcentaje de la variación de los valores de la variable independiente que se puede explicar con la ecuación de regresión (Levin y Rubin, 2004) —figura 5—. Se puede decir entonces que el 96,93% de los valores de pobreza en 2011 para esas ciudades son explicados por medio de la ecuación de regresión encontrada.

El coeficiente de determinación es la principal forma en que se puede medir el grado, o fuerza, de la asociación que existe entre dos variables, *X* y *Y*.

**Figura 5. Desviaciones con respecto a la línea de regresión**




Fuente: Levin y Rubin, 2004.

## Coeficiente de correlación

El coeficiente de correlación es la segunda medida que se usa para describir qué tan bien explica una variable a la otra. El coeficiente de correlación de la muestra se denota por r y es la raíz cuadrada del coeficiente de determinación:

$$r = (signo\ de\ b)\ \sqrt{(r^2)} \quad \textbf{(17)}$$

El signo del coeficiente indica si la relación es directa o inversa.

Para el ejemplo, r=0,9844. Esto indica que existe una fuerte asociación lineal positiva entre las variables.

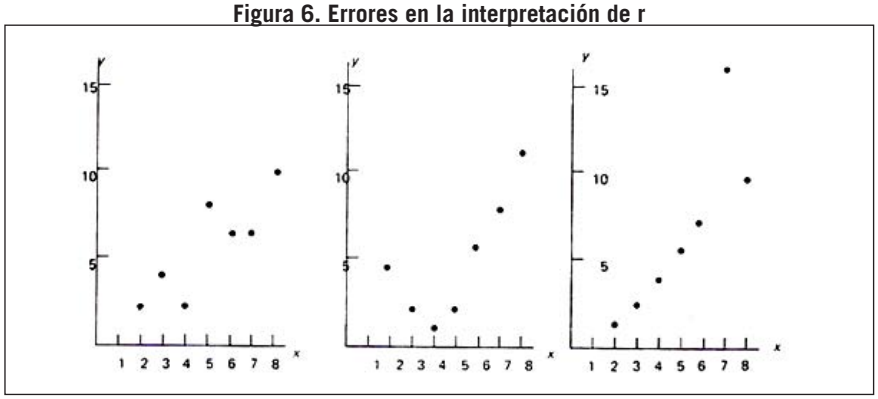
En el caso de una relación lineal entre dos variables, el coeficiente de determinación y el de correlación permiten tener medidas de la intensidad de la relación. El coeficiente de determinación da una medida entre 0 y 1, mientras que el coeficiente de correlación da una medida entre -1 y 1.

Es importante resaltar que **el coeficiente de correlación solo mide la fuerza de asociación en una relación lineal**, el coeficiente de determinación se puede usar en relaciones no lineales (obviamente, teniendo como ecuación de regresión una función no lineal) y en relaciones con dos o más variables independientes. En este sentido, el coeficiente de determinación tiene mayor aplicabilidad (Walpole y Myers, 1999); debe siempre tenerse en cuenta que:

> Para las condiciones normales que se encuentran en las ciencias sociales, con frecuencia se consideran útiles valores de $r^2$ tan bajos como 0,25. En las ciencias naturales, se manejan valores de 0,60 o más. De hecho, en algunos casos se encuentran valores mayores que 0,90. En aplicaciones de negocios, los valores de $r^2$ varían mucho, dependiendo de las características específicas de cada aplicación (Anderson, Sweeney y Williams, 2001).

La interpretación equívoca del parámetro r lleva a validar con detenimiento la figura 6, en la cual se pueden observar tres conjuntos de datos, para los cuales r=0,75 indicaría una asociación fuerte entre las variables *X* y *Y*. Sin embargo, ésta es una medida significativa de la fuerza de la relación solo en el primer caso. En el segundo hay una relación curvilínea muy evidente entre las dos variables y en el tercer caso, seis de los siete puntos en realidad caen en la línea recta, pero el séptimo punto está tan alejado que sugiere la posibilidad de un grave error de cálculo o un error en el registro de los datos. Así, antes de calcular r se deben graficar los datos para verificar si hay algún motivo para pensar que la relación es, de hecho, lineal.

**Figura 6. Errores en la interpretación de r**




Fuente: Freund y Simon, 1994.

En la deducción de la ecuación de regresión por mínimos cuadrados, y en el cálculo del coeficiente de determinación, no hicimos pruebas estadísticas de significancia de la relación entre *X* y *Y*. Los valores mayores de $r^2$ simplemente implican que la línea de regresión da un mejor ajuste con los datos, esto es, que las observaciones están agrupadas más estrechamente cerca de la recta. Pero si solo usamos el coeficiente de determinación, no llegaremos a la conclusión acerca de si la relación es estadísticamente significativa. Esa conclusión se debe basar en consideraciones donde intervengan el tamaño de la muestra y las propiedades de las distribuciones muestrales adecuadas de los estimadores de los mínimos cuadrados.

## Hipótesis del modelo

Al efectuar un análisis de regresión se comienza proponiendo una hipótesis acerca del modelo adecuado de la relación entre las variables. Para el caso de la regresión lineal simple el modelo es:

$$y = \alpha + \beta x + \epsilon$$

Luego se aplica el método de los mínimos cuadrados para determinar los valores de *a* y *b*, que son los estimados de $\alpha$ y $\beta$, respectivamente. La ecuación estimada de regresión resultante es:

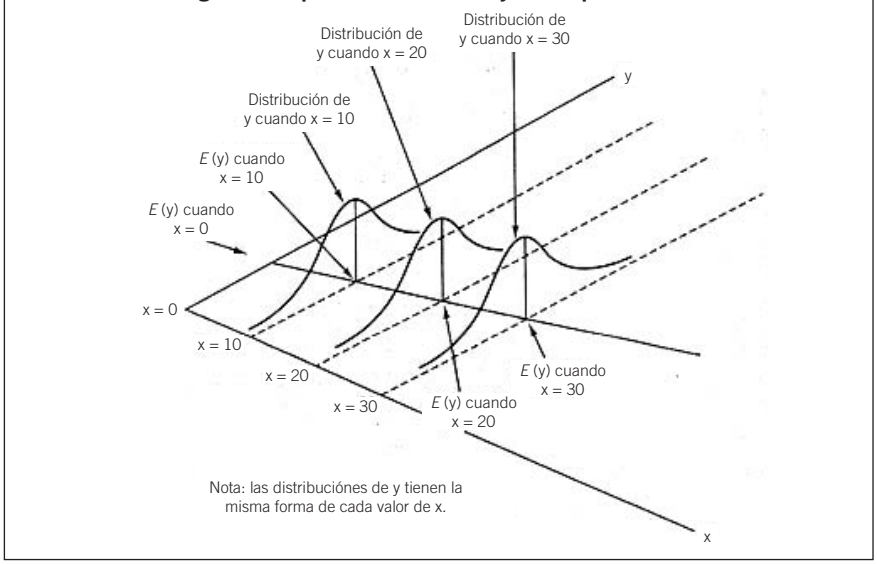
$$\hat{y} = a + bx$$

Un paso importante en la determinación de si es adecuado el modelo supuesto implica determinar el significado (o importancia estadística) de la relación. Las pruebas de significancia en el análisis de regresión se basan en los siguientes supuestos acerca del término de error $\epsilon$.

Supuestos acerca del término de error en el modelo de regresión

1. El término de error es una variable aleatoria con media o valor esperado igual a cero; $E(\epsilon) = 0$
2. La varianza de $\epsilon$, representada por $\sigma^2$, es igual para todos los valores de $x$. Esto implica que la varianza de $y$ es igual a $\sigma^2$ y es la misma para todos los valores de $x$.
3. Los valores de $\epsilon$ son independientes. El valor de $\epsilon$ para un determinado valor de x no se relaciona con el valor de $\epsilon$ para cualquier otro valor de $x$, así, el valor de $y$ para determinado valor de $x$ no se relaciona con el valor de $y$ para cualquier otro valor de $x$.
4. El término de error, $\epsilon$, es una variable aleatoria con distribución normal (Anderson, Sweeney y Williams, 2001).

**Figura 7. Supuestos del modelo y sus implicaciones**



Fuente: Sweeney y Williams, 2001.

## Pruebas de significancia

La ecuación de regresión lineal simple indica que el valor medio esperado de *y* es una función lineal de *x*:

$$E(y) = \alpha + \beta x \quad \textbf{(18)}$$

Si β=0, entonces E(y)=α. En este caso el valor medio de *y* no depende del valor de *x* y se concluye que no existe relación lineal entre las variables. En forma análoga, si el valor de β no es igual a cero, se concluye que las dos variables se relacionan. Así, para probar si hay alguna relación importante de regresión debemos efectuar una prueba de hipótesis para determinar si el valor de β es cero. Existen dos pruebas que se usan con más frecuencia y para ellas se necesita un estimado de la varianza del error en el modelo de regresión.

## Estimado de $\sigma^2$

La varianza de $\epsilon$ también representa la varianza de los valores de *y* respecto a la línea de regresión. Así, la suma de los residuales al cuadrado, SSE, es una medida de la variabilidad de las observaciones reales respecto a la línea de regresión. Cada suma de cuadrados tiene asociado un número que llamamos grados de libertad. Se ha demostrado que SSE tiene n – 2 grados de libertad, porque se deben estimar dos parámetros α y β.

El error cuadrado medio ($s^2$) es el estimado de $\sigma^2$. Se calcula mediante la ecuación:

$$s^2 = \frac{SSE}{n - 2} \quad \textbf{(19)}$$

## Desviación estándar de la estimación

El error típico o desviación estándar del estimado se calcula como la raíz cuadrada de la varianza del estimado.

$$s = \sqrt{\frac{SSE}{n - 2}} \quad \textbf{(20)}$$

De la tabla 3 se tiene que el valor de SSE=33,14
En consecuencia, 3,0129 y s=1,7358

## Prueba t

En el modelo de regresión lineal, si las variables tienen una relación lineal, debe suceder que β≠0. El objetivo de la prueba t es ver si se puede concluir que β≠0. Se usan los datos de la muestra para probar las siguientes hipótesis:

$$H_0 : \beta = 0$$

$$H_a : \beta \neq 0$$

Si se rechaza $H_0$ la conclusión será que β≠0 y que hay una relación estadísticamente significativa entre las dos variables. En este caso, las propiedades de la distribución de b, el estimador de β por mínimos cuadrados, son la base de esta prueba de hipótesis.

Las propiedades de la distribución muestral de b son las siguientes:

Valor esperado

$$E(b) = \beta$$

Desviación estándar estimada

$$s_b = \frac{S}{\sqrt{S_{xx}}} \qquad \textbf{(21)}$$

Forma de la distribución: Normal

Para el ejemplo de las ciudades colombianas, s = 1,7358 y $S_{xx}$=1176,15

$$s_b = \frac{1,7358}{\sqrt{1176,15}} = 0,0506$$

La prueba t de la significancia de la relación se basa en el hecho de que el estadístico de prueba

$$t = \frac{b - \beta}{s_b}$$

tiene una distribución t con n-2 grados de libertad.

Pasos para la prueba de hipótesis:

$$H_0 : \beta = 0$$

$$H_a : \beta \neq 0$$

$$t = \frac{b}{s_b} \textbf{ (22)}$$

Rechazar $H_0$ si $t < -t_{\alpha/2}$ 0 si $t > t_{\alpha/2}$

En donde $t_{\alpha/2}$ se basa en una distribución t con n-2 grados de libertad.

Haciendo esta prueba para el ejemplo: el estadístico de prueba (22) es:

$$t = \frac{b}{s_b} = \frac{0{,}94126}{0{,}0506} = 18{,}6$$

De acuerdo con la tabla 2 del apéndice, se observa que el valor bilateral de t que corresponde a $\alpha$=0,01 y n–2=13–2=11 grados de libertad es $t_{0.005}$=3,106.

Como 18,6>3,106, se rechaza $H_0$ y se concluye que, a un nivel de significancia de 0,01, $\beta$ no es cero. La evidencia estadística es suficiente para concluir que hay una relación importante entre las variables.

## Prueba F

También se puede usar una prueba basada en la distribución F de probabilidades, para probar si la regresión es significativa. Como solo hay una variable independiente, la prueba F debe indicar la misma conclusión que la prueba t, pero cuando hay más de una variable independiente solo se puede usar la prueba F (Miller, 2000).

Pasos de la prueba:

$$H_0 : \beta = 0$$

$$H_a : \beta \neq 0$$

$$F = \frac{SSR}{s^2} \textbf{ (23)}$$

Rechazar $H_0$ si $F > F_\alpha$

En donde $F_\alpha$ se basa en una distribución F con un grado de libertad en el numerador y n-2 grados de libertad en el denominador.

Haciendo la prueba F para el ejemplo del porcentaje de pobreza: el estadístico de prueba F es:

$$F = \frac{SSR}{s^2} = \frac{1042}{3{,}0129} = 345{,}85$$

En la tabla 3 del apéndice observamos que el valor de F que corresponde a $\alpha$=0,01, con un grado de libertad en el numerador y n-2= 11 grados de libertad en el denominador, es $F_{0.01}$= 9,65. Como 345,85>9,65, rechazamos $H_0$ y se concluye que, a un nivel de significancia del 0,01, $\beta$ no es cero.

## 4. Uso de la ecuación de regresión para estimar y predecir

Si el análisis de la ecuación de regresión obtenida con los datos demuestra que existe una relación estadísticamente significativa entre las variables, y si el ajuste que proporciona la ecuación es bueno, esa ecuación podría usarse para estimaciones y predicciones.

### Estimación de intervalo

Los estimados puntuales, como el que hicimos con respecto a la ciudad de Armenia, no dan idea alguna de la precisión asociada con el valor estimado.

Para ese fin se deben determinar estimaciones de intervalo. El primer tipo de estimado es el de *intervalo de confianza*, que es un estimado del valor medio de *y* para determinado valor de *x*. El segundo tipo es el *estimado de intervalo de predicción*, que se usa cuando deseamos un estimado de intervalo de valor individual de *y* que corresponda a determinado valor de *x*. Con la estimación puntual se obtiene el mismo valor, sea que se esté estimando el valor medio de *y* o prediciendo un valor individual de *y*, pero con los estimados de intervalo se obtienen valores distintos (Freund y Simon, 1994).

### Estimado del intervalo de confianza del valor medio de y

Al estimar el porcentaje promedio de pobreza en 2011 de todas las ciudades que en 2010 mostraron un índice de pobreza del 25,3% el estimado de $E(y_p)$, el valor medio desconocido, es :

$$\hat{y}_p = -1{,}3148 + 0{,}94126\ (25{,}3) = 22{,}5$$

Donde $\hat{y}_p$ es el estimado del valor particular de *y*.

Dado que no se puede esperar que $\hat{y}_p$ sea exactamente igual a $E(yp)$. Entonces es necesario considerar la varianza de los estimados basados en la ecuación de regresión. La fórmula para estimar la desviación estándar de $\hat{y}_p$ dado un valor particular de *x*, $x_p$, es:

$$s_{\hat{y}p} = s\sqrt{\frac{1}{n} + \frac{(x_p - \bar{x})^2}{S_{xx}}} \quad (24)$$

Entonces para el ejemplo se tiene:

$$s_{\hat{y}p} = 1{,}7358\sqrt{\frac{1}{13} + \frac{(25{,}3 - 28{,}6923)^2}{1176{,}15}} \quad (24)$$

$$s_{\hat{y}p} = 1{,}7358\sqrt{0{,}0867} = 0{,}5111$$

La ecuación general para un estimado del intervalo de confianza de $E(y_p)$ dado un valor particular de $x$ es:

$$\hat{y}_p \pm t_{\alpha/2} \cdot s_{\hat{y}p} \quad (25)$$

En donde el coeficiente de confianza es 1–α y $t_{\alpha/2}$ se basa en una distribución t con n-2 grados de libertad.

Al usar la ecuación (25) para determinar un estimado de intervalo de confianza del 95% para el porcentaje promedio de pobreza en 2011 de todas las ciudades que en 2010 mostraron un índice de pobreza del 25,3%, necesitamos el valor de t para α/2=0.025 y n–2= 11 grados de libertad. De acuerdo con la tabla 2 del apéndice, vemos que $t_{0.025}$=2,201. Así, con $\hat{y}_p$ = 22,5 y $s_{\hat{y}p}$ = 0,5111, tenemos:

$$22{,}5 \pm 2{,}201 \cdot 0{,}5111$$

$$22{,}5 \pm 1{,}125$$

Entonces, con una confianza del 95% se puede decir que el porcentaje promedio de pobreza en 2011 de todas las ciudades que en 2010 mostraron un índice de pobreza del 25,3% está entre el 21,375% y el 23,625%. Obsérvese que la desviación estándar estimada de $x_p$ expresada en la ecuación (24) es mínima cuando $x_p = \bar{x}$. Esto implica que podemos hacer el mejor estimado, o el más preciso, del valor medio de y siempre que estemos usando el valor medio de $x$. Como resultado de ello, los intervalos de confianza para el valor medio de y se ensanchan a medida que $x_p$ se aleja de $\bar{x}$.

## Estimado del intervalo de predicción para un valor particular de y

Para este análisis se supone que en vez de estimar el valor medio del porcentaje de pobreza, deseamos estimar el porcentaje de pobreza en 2011 para la ciudad de Armenia con un índice de pobreza del 25,3% en 2010.

El estimado para ese valor particular por medio de la ecuación de regresión es:

$$\hat{y}_p = -1{,}3148 + 0{,}94126(25{,}3) = 22{,}5$$

Que es el mismo valor que el estimado puntual para el porcentaje promedio.

Para determinar un estimado del intervalo de predicción debemos determinar primero la varianza asociada al empleo de $\hat{y}_p$ como estimado de un valor individual de *y*. Esta varianza está formada por la suma de dos componentes:

La varianza de los valores individuales de *y* respecto del promedio, cuyo estimado es $s^2$

La varianza asociada al uso de $\hat{y}_p$ para estimar $E(y_p)$, cuyo estimado es $s_{\hat{y}p}$.

Así, el estimado de la varianza de un valor individual es:

$$s^2_{ind} = s^2 + s_{\hat{y}p}$$

Por consiguiente, un estimado de la desviación estándar de un valor un individual de $\hat{y}_p$ es:

$$s_{ind} = s\sqrt{1 + \frac{1}{n} + \frac{(xp - \bar{x})^2}{s_{xx}}} \quad (26)$$

Para el ejemplo que se ha tratado:

$$s_{ind} = 1{,}7358\sqrt{1 + \frac{1}{13} + \frac{(25{,}3 - 28{,}6923)^2}{1176{,}15}} \quad (26)$$

$$s_{ind} = 1{,}7358\sqrt{1{,}0867} = 1{,}8095$$

La ecuación general para un estimado del intervalo de predicción para un valor individual de *y* dado un valor particular de x es:

$$\hat{y}_p \pm t_{\alpha/2} \cdot s_{ind} \quad (27)$$

En donde el coeficiente de confianza es 1–α y $t_{\alpha/2}$ se basa en una distribución t con n-2 grados de libertad.

Al usar la ecuación (27) para determinar un estimado de intervalo de predicción del 95% para el porcentaje de pobreza en 2011 de la ciudad de Armenia, que en 2010 mostró un índice de pobreza del 25,3%, se necesita el valor de t para α/2=0.025 y n-2= 11 grados de libertad. De acuerdo con la tabla 2 del apéndice, se observa que $t_{0.025}$=2,201. Así, con $\hat{y}_p$=22,5 y $s_{ind}$=1,8095, tenemos:

$$22{,}5 \pm 2{,}201 \cdot 1{,}8095$$

$$22{,}5 \pm 3{,}9827$$

Entonces, con una confianza del 95% se puede decir que el porcentaje de pobreza en 2011 de la ciudad de Armenia, que en 2010 tenía un porcentaje de pobreza del 25,3%, está entre el 18,52% y el 26,48%.

De acuerdo con lo anterior, el intervalo de predicción es mayor que el intervalo de confianza.

## Estimación de los parámetros del modelo de regresión lineal

Uno de los conceptos fundamentales sobre el cual se ha basado este análisis consiste en que la ecuación de regresión lineal obtenida a partir de los datos de la muestra es un estimado de los parámetros del modelo para la población. Por lo tanto, es posible determinar intervalos de confianza para los coeficientes de la ecuación de regresión:

$$\alpha = a \pm t_{\alpha/2} \cdot s \sqrt{\frac{1}{n} + \frac{(\bar{x})^2}{s_{xx}}} \quad (28)$$

$$\beta = b \pm t_{\alpha/2} \frac{S}{\sqrt{S_{xx}}} \quad (29)$$

Siguiendo con el ejemplo, al realizar los cálculos de los intervalos de confianza de los parámetros del modelo se tiene:

$$\alpha = -1{,}3148 \pm 2{,}201 \cdot 1{,}7358 \sqrt{\frac{1}{13} + \frac{(28{,}6923)^2}{1176{,}15}} \quad (28)$$

$$\alpha = -1{,}3148 \pm 3{,}3675$$

$$\beta = 0{,}94126 \pm 2{,}201 \cdot \frac{1{,}7358}{\sqrt{1176{,}15}}$$

$$\beta = 0{,}94126 \pm 0{,}1114$$

La estimación y la inferencia son herramientas estadísticas que, cuando se utilizan de forma correcta, pueden prestar una ayuda significativa a las personas que toman decisiones. Infortunadamente, con frecuencia se utilizan de manera incorrecta o sencillamente no se usan. Como resultado, los responsables de la toma de decisiones a menudo hacen predicciones inexactas y toman decisiones menos que deseables.

Un error común es suponer que la línea de regresión, así el ajuste sea muy bueno (valor de $r^2$ muy alto), puede aplicarse en cualquier intervalo de valores. Aun cuando una relación se cumpla para el intervalo de puntos de la muestra, puede existir una relación completamente distinta para un intervalo diferente. Por ejemplo, la relación edad y talla puede ser lineal para cierto intervalo del crecimiento de los niños en su primera infancia —como se verá en el capítulo siguiente— pero en la adolescencia esa relación ya no es lineal. Recuérdese que una ecuación de estimación es válida solo para el mismo rango dentro del cual se tomó la muestra inicialmente (Levin y Rubin, 2004).

Otro error que se suele cometer al utilizar el análisis de regresión es suponer que un cambio en una variable es "ocasionado" por un cambio en la otra variable. Como se vio, los análisis de regresión y correlación no pueden, de ninguna manera, determinar la causa y el efecto. Al decir que existe una correlación entre los porcentajes de pobreza en los años 2010 y 2011 para las trece ciudades capitales de nuestro país no se está diciendo que uno ocasiona al otro. "La validez de una conclusión de tipo causa y efecto requiere de una justificación teórica, o del buen juicio por parte del analista" (Anderson, Sweeney, y Williams, 2001).

## 5. Solución de problemas de regresión con Excel

Hacer un análisis de regresión puede ser muy engorroso si no se cuenta con un computador. En esta sección describiremos cómo se pueden automatizar los cálculos por medio del programa Microsoft EXCEL® 2007 o 2010. Se escogió este programa porque le resulta familiar a todo profesional que no haya tenido experiencia alguna con programas netamente estadísticos.

Para hacer uso de esta hoja de cálculo, se analizará la relación entre las variables edad y talla de niños entre 6 y 60 meses de edad de una muestra proveniente de 4.014 niños y niñas de Ciudad Bolívar (Bogotá) en un trabajo de investigación en el área de Salud (Ducuara Mora, 2012).

**Tabla 4. Edad y talla en niños de 6 a 60 meses**

| Edad (meses) | Talla (cms) | Edad (meses) | Talla (cms) |
|---|---|---|---|
| 6 | 65 | 34 | 87,4590361 |
| 8 | 72,25 | 36 | 89,6215054 |
| 10 | 78 | 38 | 90,7149533 |
| 12 | 71,4166667 | 40 | 94,8675 |
| 14 | 72,08 | 42 | 93,7096154 |
| 16 | 74,6736842 | 44 | 95,310219 |
| 18 | 77,8125 | 46 | 96,3507246 |
| 20 | 77,9958333 | 48 | 97,1337838 |
| 22 | 81,9057143 | 50 | 99,1140741 |
| 24 | 81,5162162 | 52 | 99,7460317 |
| 26 | 82,7729167 | 54 | 100,651095 |
| 28 | 85,5116279 | 56 | 101,551799 |
| 30 | 85,5852941 | 58 | 103,880488 |
| 32 | 85,7066667 | 60 | 107,5592 |

Fuente: Elaboración propia.

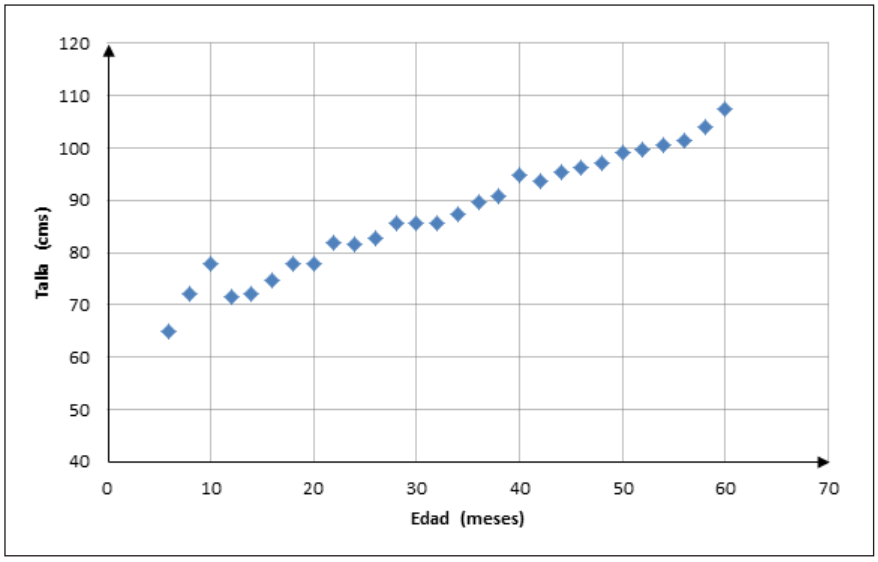
Se ingresan estos datos en la hoja de cálculo y se procede a hacer la gráfica de dispersión en el menú insertar—gráficos—dispersión (figura 8).

**Figura 8. Instrucciones para la gráfica de dispersión en EXCEL**


Fuente: Elaboración propia.

La gráfica de dispersión nos sugiere que existe una relación lineal entre la variable independiente edad y la variable dependiente talla (figura 9).

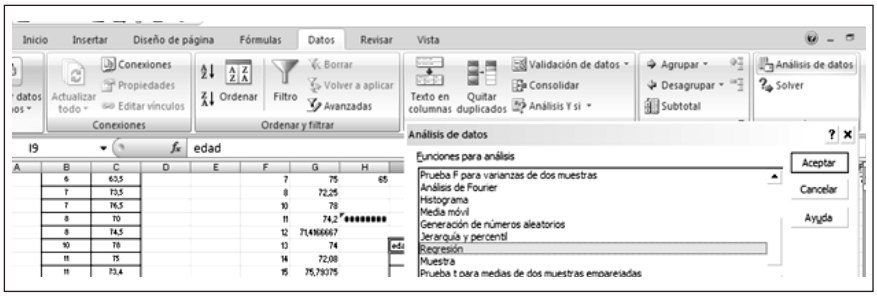
**Figura 9. Gráfica de dispersión de los datos de la Tabla 4**




Fuente: Elaboración propia.

Ahora se procede a encontrar la ecuación estimada de regresión y a hacer el análisis de ésta con respecto a los datos por medio de la instrucción Regresión del menú de Análisis de datos (figura 10).

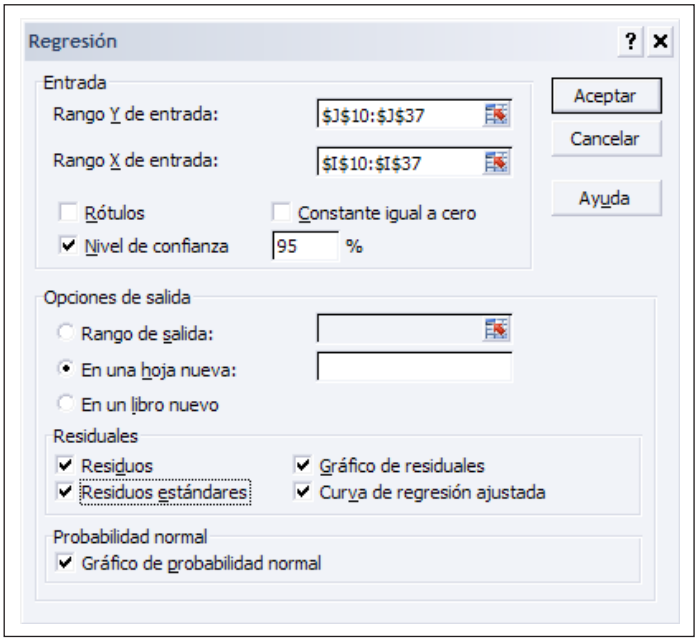
**Figura 10. Instrucciones para el análisis de regresión en Excel**




Fuente: Elaboración propia.

Luego se ingresan los rangos de las variables y se fija el nivel de confianza para el intervalo estimado de los parámetros α y β (figura 11).

**Figura 11. Cuadro de diálogo para el análisis de regresión en EXCEL**




Fuente: Elaboración propia.

Las opciones de residuales y gráfico de probabilidad se estudiarán en el siguiente capítulo.

Los resultados que arroja el programa son los siguientes:

**Tabla 5. Estadísticos de la regresión lineal con Excel**

| Estadísticas de la regresión | |
|---|---|
| Coeficiente de correlación múltiple | 0,986569761 |
| Coeficiente de determinación R^2 | 0,973319893 |
| R^2 ajustado | 0,972293735 |
| Error típico | 1,878895123 |
| Observaciones | 28 |
| Análisis de varianza | |

| | Grados de libertad | Suma de cuadrados |
|---|---|---|
| Regresión | 1 | 3348,470355 |
| Residuos | 26 | 91,78641898 |
| Total | 27 | 3440,256774 |
| | | |
| Promedio de los cuadrados | F | Valor crítico de F |
| 3348,470355 | 948,5088339 | 5,44709 E-22 |
| 3,530246884 | | |

Fuente: Elaboración propia.

**Tabla 6. Regresores estimados de la ecuación lineal**

| | Coeficientes | Error típico | Estadístico t | Probabilidad | Inferior 95% | Superior 95% |
|---|---|---|---|---|---|---|
| Intercepción | 65,1586551 | 0,807551362 | 80,68670075 | 9,6913E-33 | 63,49870952 | 66,8186007 |
| Variable X 1 | 0,676899136 | 0,021978764 | 30,79787061 | 5,44709E-22 | 0,63172114 | 0,72207713 |

Fuente: Elaboración propia.

De la información obtenida se deduce:

La ecuación estimada de regresión $\hat{y} = 65{,}159 + 0{,}6769\ x$

$R^2$= 0,973319893, es decir que la ecuación tiene un muy buen ajuste pues explica la variación de *y* en un 97,33%

El estadístico de prueba t= 30,798 para el estimado de la pendiente de la recta es mayor que el valor de $t_{\alpha/2}$ con $\alpha$=0.005 con n-2= 26 grados de libertad t =2,779. Por eso se rechaza $H_0$ y se dice que la relación es significativa;

además, esto se observa en el valor de probabilidad que aparece en la celda del lado derecho del estadístico (en la hoja de Excel) que es un valor prácticamente igual a cero. Lo mismo sucede con el estadístico F.

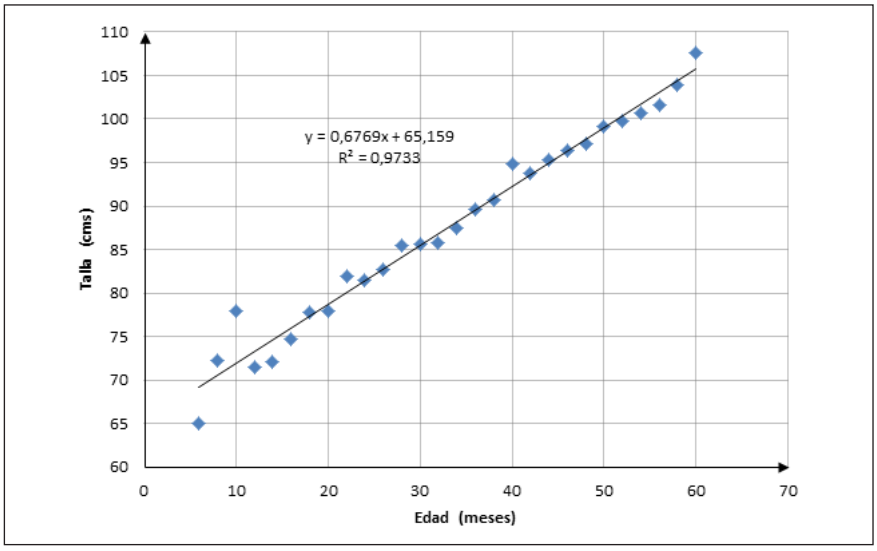
Finalmente, aparecen los intervalos de confianza del 95% de los parámetros del modelo de regresión lineal α y β.

$$\beta = 0{,}6769 \pm 0{,}045177995$$

$$\alpha = 65{,}159 \pm 1{,}65994558$$

Para hacer la gráfica de la ecuación de regresión se seleccionan los puntos que corresponden a los datos en la gráfica de dispersión y se hace clic derecho sobre alguno de ellos. Luego aparece un cuadro de diálogo en el cual se selecciona la opción Línea de tendencia, una vez allí se escoge la opción lineal y se elige la posibilidad de que aparezca la ecuación y el valor de $r^2$ en el gráfico (figura 12).

**Figura 12. Gráfica de la ecuación de regresión con Excel**




Fuente: Elaboración propia.

## 6. Análisis de residuales

Como se explicó anteriormente, el residual en la observación i-ésima es la diferencia entre el valor observado de la variable independiente ($y_i$) y el valor estimado de esa variable ($\hat{y}_i$). En otras palabras, el i-ésimo residual es el error debido al uso de la ecuación de regresión para predecir el valor de $y_i$. Un análisis de esos residuales ayudará a determinar si son adecuados los supuestos que se hicieron sobre el modelo de regresión; de hecho, ofrecen la mejor información con respecto a $\epsilon$ (Anderson, Sweeney y Williams, 2001).

Recuérdese que los supuestos sobre el modelo de regresión forman la base teórica de las pruebas t y F que se usan para determinar si la relación entre las variables es significativa y para los estimados de los intervalos de confianza y predicción que se describieron en el capítulo 3. Si hay duda acerca de esos supuestos sobre el término de error, podrían no ser válidas las pruebas de hipótesis acerca de la significancia estadística de la relación de regresión y de la estimación de intervalos.

El análisis de residuales se basa en el examen de varias gráficas, a saber:

- Gráfica de los residuales en función de la variable independiente
- Gráfica de residuales estandarizados
- Gráfica de probabilidad normal

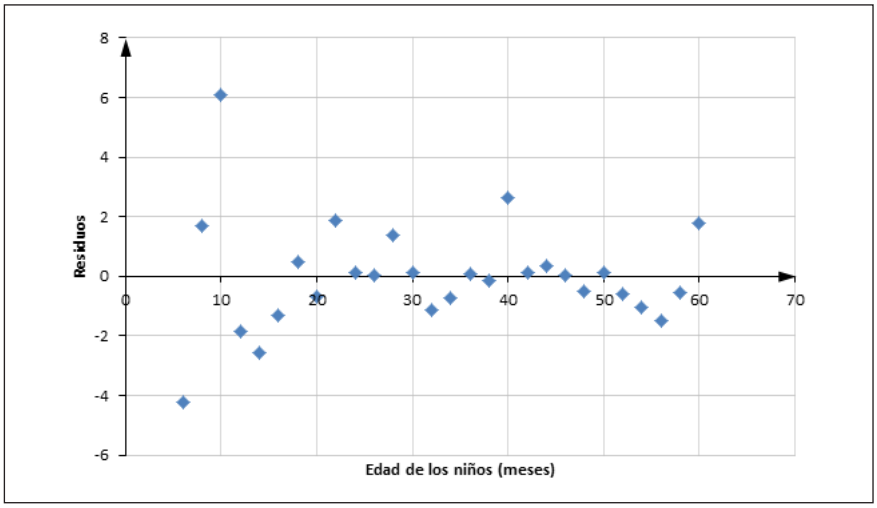
### Gráfica de residuales en función de x

Ésta es una gráfica en la que los valores de la variable independiente se representan en el eje horizontal y los valores de los residuales correspondientes, en el eje vertical. Se grafica un punto para cada residual.

También es usual presentar la gráfica de residuales con respecto a los valores de la variable dependiente ($\hat{y}_i$) estimados por la ecuación. Para la regresión lineal simple, la gráfica de residuales en función de $x$ y la de residuales en función de $\hat{y}$ muestran la misma información; mientras que, para la regresión lineal múltiple, la gráfica de residuales en función de $\hat{y}$ se usa con más frecuencia, porque se maneja más de una variable independiente. (Anderson, Sweeney y Williams, 2001).

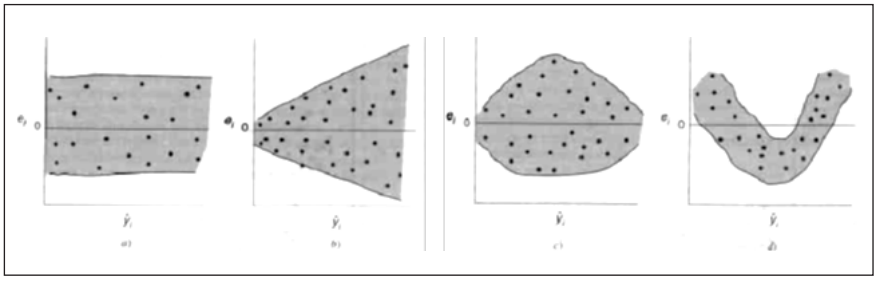
La figura 13 muestra los residuales de los datos del ejemplo de la talla y la edad de niños de la localidad Ciudad Bolívar (Bogotá).

**Figura 13. Gráfica de residuales de la relación edad y talla de los niños**




Antes de interpretar los resultados de esta gráfica analicemos algunos patrones que se pueden presentar en cualquier gráfica de este tipo.

**Figura 14. Posibles patrones de distribución de los residuales**



Fuente: Lopera, 2002.

En la figura 14 se observan cuatro posibles resultados para la gráfica de residuales. Si es cierta la hipótesis de que la varianza de $\epsilon$ es igual para todos los valores de x y si el modelo de regresión lineal es una representación adecuada de la relación entre las variables, entonces, la gráfica debe mostrar un patrón muy similar a una franja horizontal de puntos (figura 14a). Ahora bien, si la varianza de $\epsilon$ no es constante —por ejemplo, si la varianza aumenta

conforme aumenta el valor de la variable independiente— se puede observar un patrón como el de la figura 14b. Sin embargo, no es la única forma en que se puede dar que la varianza de $\epsilon$ no sea constante como en la figura 14c. Otra posibilidad es la de la figura 14d, en este caso se concluiría que el modelo lineal no representa adecuadamente la relación entre las variables y entonces se pensaría en modelos curvilíneos o de regresión múltiple.

Volviendo nuevamente a la gráfica de residuales de la relación edad y talla de los niños, se observa una distribución parecida a una franja horizontal. Por lo tanto, se concluye que la gráfica no muestra evidencia que justifique una duda sobre el supuesto de que la varianza de $\epsilon$ es constante.

## Gráfica de residuales estandarizados

La mayoría de las gráficas de residuales que se obtienen con el uso de programas estadísticos u hojas de cálculo muestran una versión estandarizada de los residuales. Estandarizar una variable aleatoria significa restarle su media y dividir el resultado entre su desviación estándar. Como sabemos, el promedio de los residuales es cero debido al método de los mínimos cuadrados. Por tanto, para obtener un residual estandarizado basta con dividir el residual entre su desviación estándar.

### Estimado de la desviación estándar del i-ésimo residual

El estimado de la desviación estándar del residual *i* depende del error estándar del estimado *s* y el valor correspondiente de la variable independiente $x_i$, así:

Donde $s_{ei} = s\sqrt{1 - h_i}$ (31)

$$h_i = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum (x_i - \bar{x})^2}$$

La cantidad $h_i$ es conocida como influencia de la observación *i* (Devore, 2005).

Una vez calculada la desviación estándar de cada residual, se procede a calcular el residual estandarizado.

$$e_{zi} = \frac{y_i - y_{\hat{i}}}{s_{ei}}$$

La tabla 5 muestra los residuales y los residuales estandarizados del estudio de edad y talla de los niños. Estos datos se obtuvieron mediante el análisis de regresión con la hoja de cálculo Excel®.

Posteriormente se grafican estos residuales estandarizados con respecto a los valores de *x*.

La gráfica de residuales estandarizados nos brinda información acerca de la hipótesis de que el término de error tiene distribución normal. Si es cierta la hipótesis, cabe esperar que, aproximadamente, el 95% de los residuales estandarizados estén entre –2 y 2. Observando la gráfica de residuales (figura 15) notamos que solo dos de los residuales están fuera del intervalo mencionado.

**Tabla 5. Residuales y residuales estandarizados para el ejemplo de edad y talla de los niños**

| Edad | Talla | Residuos | Residuos estándares |
|---|---|---|---|
| $x_i$ | $y_i$ | $y_i - \hat{y}_i$ | $e_{zi}$ |
| 6 | 65,0 | -4,22004992 | -2,288812715 |
| 8 | 72,3 | 1,67615181 | 0,909088195 |
| 10 | 78,0 | 6,07235353 | 3,293439707 |
| 12 | 71,4 | -1,86477807 | -1,011392718 |
| 14 | 72,1 | -2,55524301 | -1,385877609 |
| 16 | 74,7 | -1,31535707 | -0,713405303 |
| 18 | 77,8 | 0,46966044 | 0,254727981 |
| 20 | 78,0 | -0,70080449 | -0,380092716 |
| 22 | 81,9 | 1,85527819 | 1,006240301 |
| 24 | 81,5 | 0,11198184 | 0,060735175 |
| 26 | 82,8 | 0,01488402 | 0,008072592 |
| 28 | 85,5 | 1,39979699 | 0,759202666 |
| 30 | 85,6 | 0,11966493 | 0,064902221 |
| 32 | 85,7 | -1,11276079 | -0,603523916 |
| 34 | 87,5 | -0,71418959 | -0,38735234 |
| 36 | 89,6 | 0,09448137 | 0,051243508 |

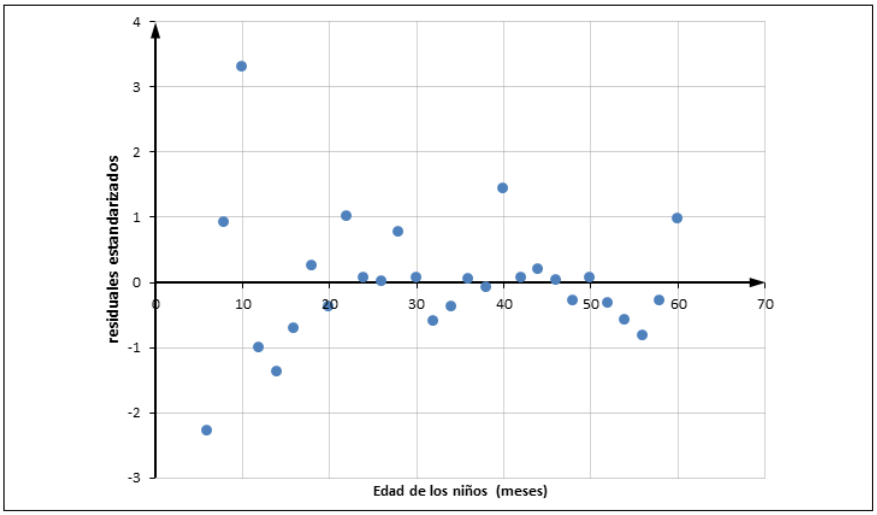
Continúa

| | | | |
|---|---|---|---|
| 38 | 90,7 | -0,16586901 | -0,089961754 |
| 40 | 94,9 | 2,63287945 | 1,427984994 |
| 42 | 93,7 | 0,12119656 | 0,065732927 |
| 44 | 95,3 | 0,36800188 | 0,199591807 |
| 46 | 96,4 | 0,05470927 | 0,029672463 |
| 48 | 97,1 | -0,51602986 | -0,279877185 |
| 50 | 99,1 | 0,11046216 | 0,059910951 |
| 52 | 99,7 | -0,61137844 | -0,33159104 |
| 54 | 100,7 | -1,06011357 | -0,574969835 |
| 56 | 101,6 | -1,51320817 | -0,820713062 |
| 58 | 103,9 | -0,5383172 | -0,291965087 |
| 60 | 107,6 | 1,78659673 | 0,968989794 |

Fuente: Elaboración propia.

Puesto que se está trabajando con 28 observaciones, decir que dos de ellas están fuera del intervalo de dos desviaciones estándar implica que aproximadamente el 95% de los datos está dentro del intervalo y no habría razón suficiente para dudar de que el término de error tenga distribución normal.

**Figura 15. Gráfica de residuales estandarizados para el ejemplo de edad y talla de los niños**




Fuente: Elaboración propia.

## Gráfica de probabilidad normal

Se necesita a nivel documentativo señalar la tendencia de la probabilidad normal tal como se visualiza en el figura 16, para lo cual es necesario utilizar el análisis de regresión realizado con Excel, parametrizando que la construcción de esta gráfica requiere considerar los siguientes pasos:

1. Ordenar los *n* residuos de menor a mayor.
2. Estimar el porcentaje empírico de residuos menor que el residuo específico que se está considerando así:

$$P_e = \frac{i - 0.5}{n}$$

   Donde i: es el número de orden de cada dato.
   n: es el total de datos.

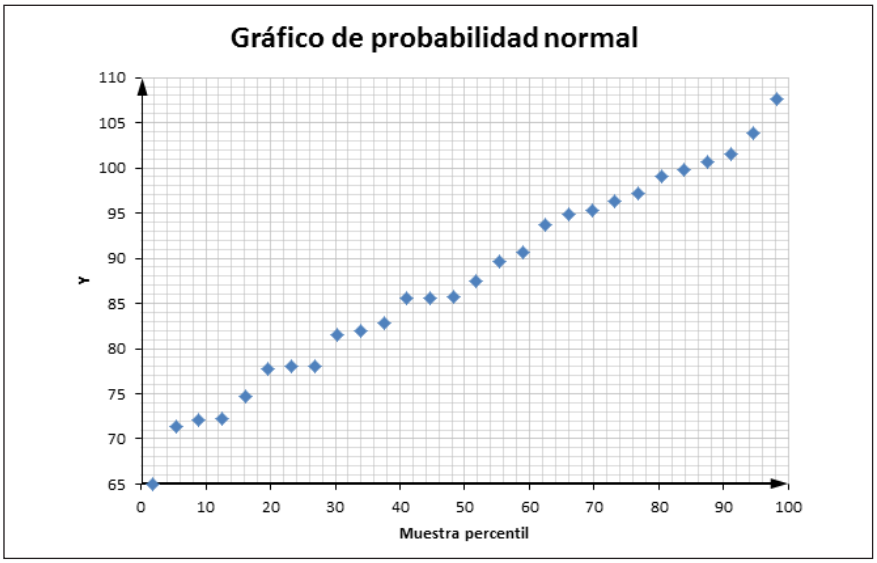
3. Calcular el porcentaje teórico de residuos menor que el residuo específico usando la tabla de distribución normal (tabla 1 del Apéndice), es decir:

$$F(residuo) = P(Z<residuo\ estándar)$$

4. Grafique la pareja $(F(residuo), P_e)$

Si los puntos parecen ajustarse a una línea recta (de la forma $y = x$), indicaría que los datos provienen de una distribución normal, pero hay que tener en cuenta que, en algunos casos, aunque los puntos se ajusten a una línea recta puede que los datos no sean generados por una distribución normal, por ello, es recomendable siempre utilizar como métodos de referencia y validación las pruebas de Shapiro-Wilks y Kolmogorov Smirnov (Muñoz R., 2006).

**Figura 16. Gráfica de probabilidad normal del ejemplo de edad y talla de los niños**




Fuente: Elaboración propia.

## Detección de valores atípicos

Un dato atípico (outlier) es un registro mayor o menor de lo esperado que se detecta por tener un residuo que es un valor "inusual", muy grande o muy pequeño en relación con la distribución asociada a los residuos.

Dado que los residuos estandarizados $e_{zi}$ son una muestra aleatoria de una distribución normal con media cero y desviación estándar uno, N(0,1), se verifica que aproximadamente un 68% de los $e_{zi}$ deben estar entre -1 y1, y alrededor del 95% entre -2 y 2 y prácticamente todos entre -3 y 3. Por ello, un residuo estandarizado que diste más de 3 o 4 unidades del 0 corresponde, potencialmente, con una observación atípica (Vilar, 2006).

Los valores atípicos representan observaciones de alguna manera sospechosas y que requieren de un examen cuidadoso. Pueden representar datos erróneos; en este caso se deben corregir los datos. Pueden evidenciar una violación de los supuestos del modelo; de ser así, se debe buscar otro modelo. También, pueden ser valores poco usuales que han sucedido por casualidad y se deben conservar (Anderson, Sweeney y Williams, 2001).

Esta última situación que se menciona, se evidencia en la tabla 5 en ejemplo de la relación edad y talla de los niños que fueron objeto de estudio en la investigación de Ducuara. Allí podemos encontrar que para la tercera observación se tiene un residual estandarizado mayor a 3, pero, al no haber error en la medición, el dato debe conservarse.

## Detección de observaciones influyentes

Existen situaciones en las cuales una o más observaciones tienen una gran influencia sobre los resultados obtenidos. Una observación es influyente si tiene un impacto notable sobre los coeficientes de regresión ajustados porque "jalan" al modelo en su dirección. Se caracterizan por tener un valor moderadamente inusual tanto en las predictoras como en la respuesta (Lopera, 2002).

La siguiente tabla muestra el porcentaje de niños entre 5 y 10 años que presentan desnutrición en un estudio realizado en México (Instituto Nacional de Salud Pública, 2006).

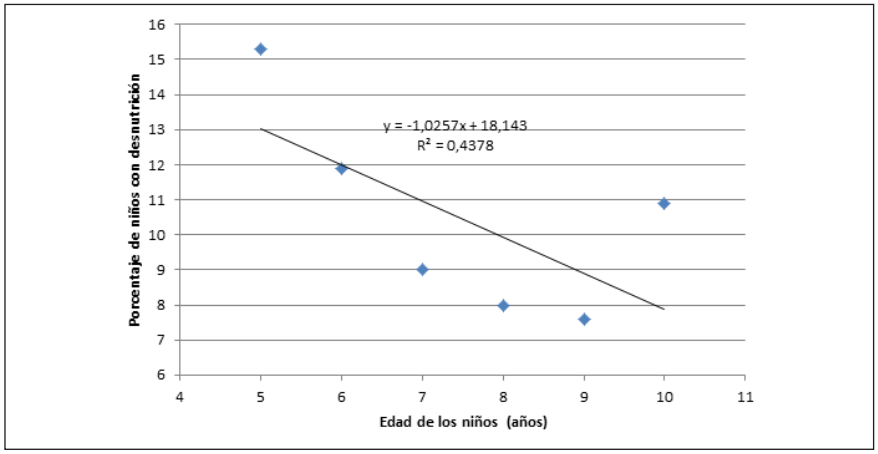
**Tabla 6. Porcentaje de niños en estado de desnutrición**

| | Masculino | | | |
|---|---|---|---|---|
| Edad en años | Muestra número | Número (miles) | Expansión | |
| | | | % | IC95% |
| 5 | 988 | 1 117.9 | 15.3 | (11.2, 20.53) |
| 6 | 961 | 1 037.6 | 11.9 | (9.02, 15.62) |
| 7 | 1001 | 1 014.5 | 9.0 | (6.65, 12.05) |
| 8 | 1099 | 1 075.1 | 8.0 | (5.91, 10.79) |
| 9 | 1170 | 1 200.3 | 7.6 | (5.53, 10.32) |
| 10 | 1194 | 1 268.6 | 10.9 | (8.38, 14.1) |

Fuente: Instituto Nacional de Salud Pública.

La gráfica de dispersión y de la línea estimada de regresión para estos datos aparece en la figura 17.

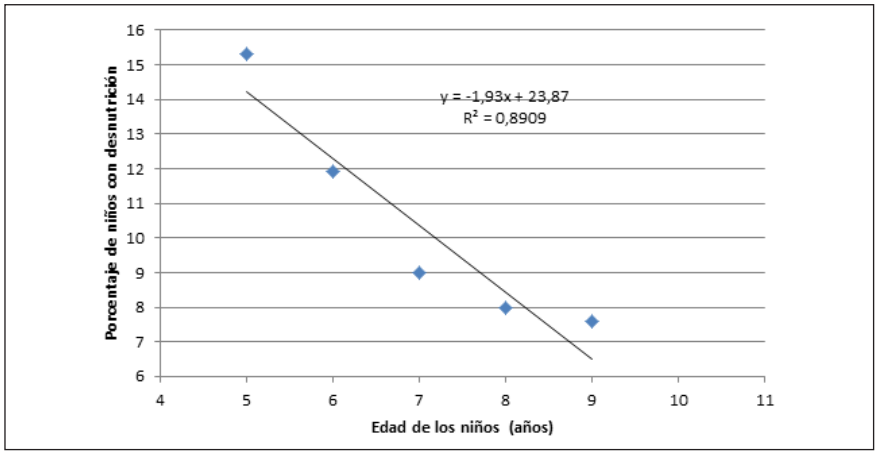
**Figura 17. Gráfica de la ecuación de regresión para el ejemplo de los niños en México**




Fuente: Elaboración propia.

Se observa que el coeficiente de determinación es bajo (43,78%) debido a la última observación. Si este valor se elimina, obtendríamos una ecuación de regresión más ajustada (figura 18).

**Figura 18. Gráfica de regresión sin la observación influyente**




Fuente: Elaboración propia.

Como se pudo notar, las observaciones influyentes tienen un efecto grande sobre la ecuación de regresión. Si la observación influyente es válida y no un error en el registro, como en este caso; entonces es necesario replantear el modelo y no usar la regresión lineal.

# 7. Conclusiones

El análisis de regresión es una técnica estadística empleada para el estudio de la relación entre variables determinísticas o aleatorias que provienen de un proceso investigativo, el caso más sencillo de estudio se conoce como modelo de regresión lineal simple, caracterizado porque solo hay dos variables, una independiente y una dependiente, y la gráfica de dispersión muestra que se relacionan por medio de una recta, cuya ecuación es $y = a + bx$.

El modelo utilizado es $y = \alpha + \beta x + \epsilon$ que corresponde a la recta que representa el comportamiento de la población. Por medio del método de mínimos cuadrados se llega al valor medio del modelo $E(y)$ o ecuación estimada de regresión $\hat{y} = a + bx$ a partir de la muestra.

Una vez obtenida la ecuación se debe determinar la medida de la bondad de ajuste y la fuerza de esa relación por medio de los coeficientes de determinación y correlación. Verificado lo anterior, se comprueba si la relación es estadísticamente significativa por medio de las pruebas t y F y se validan los supuestos del modelo con respecto al término de error $\epsilon$ haciendo un análisis de residuales.

Cuando se tiene certeza de la pertinencia y validez de la ecuación, el investigador puede hacer uso de ella para hacer predicciones y estimaciones. Sin embargo, solo puede hacerlas utilizando valores dentro del rango de los datos observados.

Para hacer predicciones fuera del rango, el investigador o experto debe tener la seguridad de que, fuera del intervalo, la tendencia o relación entre los datos se mantiene. Esta técnica, en ningún caso, determina una relación causa-efecto entre las variables, por muy bueno que sea el ajuste de la recta con respecto a las observaciones.

Cuando los datos son muy grandes o se incurre en demasiada demora en su manejo, es preciso utilizar las hojas de cálculo o los programas estadísticos, para asegurar así efectividad, oportunidad y trasparencia en la interpretación y el análisis de los resultados.

En la práctica, es poco común encontrar dos variables que se relacionen efectivamente en forma lineal. De hecho, el investigador en ocasiones recurre a hacer transformaciones en los datos para que la relación resulte

lineal. Por tal motivo, si se encuentran observaciones atípicas o influyentes que no puedan obviarse y la recta de regresión no explica de manera adecuada la variación en *y* de acuerdo con los incrementos en *x* será necesario abandonar este modelo y pensar en uno no lineal.

## Bibliografía


Anderson, D. R., Sweeney, D. J. y Williams, T. A. (2001). *Estadística para administración y economía* (7a ed., Vol. II). México: Thomson.

DANE. (17 de mayo de 2012). "Pobreza en Colombia". *Comunicado de prensa*, 6.

Devore, J. L. (2005). *Probabilidad y estadística para ingeniería y ciencias* (6a ed.). México: Thomson Learning.

Ducuara Mora, P. E. (2012). "Determinantes socio-económicas de la desnutrición global infantil en la localidad de Ciudad Bolívar en el Año 2011". Bogotá, Colombia. Trabajo de grado

Evans, M. y Rosenthal, J. S. (2005). *Probabilidad y estadística. La ciencia de la incertidumbre.* Barcelona: Reverté.

Freund, J. E. y Simon, G. A. (1994). *Estadística elemental* (8a ed.). México: Prentice Hall.

Instituto Nacional de Salud Pública. (2006). "Encuesta de salud y nutrición" (2a ed.). México.

Levin, R. I. y Rubin, D. S. (2004). *Estadística para administración y economía.* México: Pearson Educación.

Lopera, C. M. (2002). "Análisis de residuales", en Universidad Nacional de Colombia: http://www.docentes.unal.edu.co/cmlopera/docs/Estad2/2_RLM/2.(Complemento)Análisis de Residuales y Otros en RLM.pdf

Mendoza, H., Vargas, J., López, L. y Bautista, G. (2002). "Métodos de regresión", en Universidad Nacional de Colombia: http://www.virtual.unal.edu.co/cursos/ciencias/2007315/

Miller, I. (2000). *Estadística matemática con aplicaciones.* (6a ed.). México: Pearson Educación.

Muñoz R., L. A. (2006). "Comprobación de los supuestos del modelo de regresión lineal", en Universidad Autónoma de Occidente: http://augusta.uao.edu.co/moodle/file.php/284/18_supuestos_de_la_regresion_lineal.pdf

Pacheco, P. (2012). "Validación de supuestos" en Universidad Nacional de Colombia: http://www.virtual.unal.edu.co/cursos/ciencias/dis_exp/und_3/pdf/validaciondesupuestosunidad 3b[1].pdf

Vilar, J. (2006). "Identificación de valores atípicos y observaciones influyentes, en Universidad de La Coruña: http://www.udc.es/dep/mate/estadistica2/sec4_6.html

Walpole, R. E. y Myers, R. H. (1999). *Probabilidad y estadística para ingenieros* (6a ed.). México: Prentice Hall.

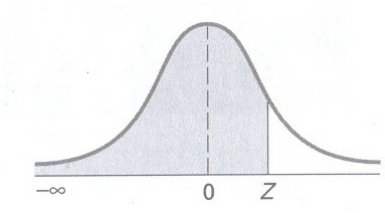
# Apéndice

**Tabla 1. Distribución normal estandarizada**

| Z | 0,00 | 0,01 | 0,02 | 0,03 | 0,04 | 0,05 | 0,06 | 0,07 | 0,08 | 0,09 |
|---|---|---|---|---|---|---|---|---|---|---|
| -3,0 | 0,0013 | 0,0013 | 0,0013 | 0,0012 | 0,0012 | 0,0011 | 0,0011 | 0,0011 | 0,0010 | 0,0010 |
| -2,9 | 0,0019 | 0,0018 | 0,0018 | 0,0017 | 0,0016 | 0,0016 | 0,0015 | 0,0015 | 0,0014 | 0,0014 |
| -2,8 | 0,0026 | 0,0025 | 0,0024 | 0,0023 | 0,0023 | 0,0022 | 0,0021 | 0,0021 | 0,0020 | 0,0019 |
| -2,7 | 0,0035 | 0,0034 | 0,0033 | 0,0032 | 0,0031 | 0,0030 | 0,0029 | 0,0028 | 0,0027 | 0,0026 |
| -2,6 | 0,0047 | 0,0045 | 0,0044 | 0,0043 | 0,0041 | 0,0040 | 0,0039 | 0,0038 | 0,0037 | 0,0036 |
| -2,5 | 0,0062 | 0,0060 | 0,0059 | 0,0057 | 0,0055 | 0,0054 | 0,0052 | 0,0051 | 0,0049 | 0,0048 |
| -2,4 | 0,0082 | 0,0080 | 0,0078 | 0,0075 | 0,0073 | 0,0071 | 0,0069 | 0,0068 | 0,0066 | 0,0064 |
| -2,3 | 0,0107 | 0,0104 | 0,0102 | 0,0099 | 0,0096 | 0,0094 | 0,0091 | 0,0089 | 0,0087 | 0,0084 |
| -2,2 | 0,0139 | 0,0136 | 0,0132 | 0,0129 | 0,0125 | 0,0122 | 0,0119 | 0,0116 | 0,0113 | 0,0110 |
| -2,1 | 0,0179 | 0,0174 | 0,0170 | 0,0166 | 0,0160 | 0,0158 | 0,0154 | 0,0150 | 0,0146 | 0,0143 |
| -2,0 | 0,0228 | 0,0222 | 0,0217 | 0,0212 | 0,0207 | 0,0202 | 0,0197 | 0,0192 | 0,0188 | 0,0183 |
| -1,9 | 0,0287 | 0,0281 | 0,0274 | 0,0268 | 0,0262 | 0,0256 | 0,0250 | 0,0244 | 0,0239 | 0,0233 |
| -1,8 | 0,0359 | 0,0351 | 0,0344 | 0,0336 | 0,0329 | 0,0322 | 0,0314 | 0,0307 | 0,0301 | 0,0294 |
| -1,7 | 0,0446 | 0,0436 | 0,0427 | 0,0418 | 0,0409 | 0,0401 | 0,0392 | 0,0384 | 0,0375 | 0,0367 |
| -1,6 | 0,0548 | 0,0537 | 0,0526 | 0,5160 | 0,0505 | 0,0495 | 0,0485 | 0,0475 | 0,0465 | 0,0455 |
| -1,5 | 0,0668 | 0,0655 | 0,0643 | 0,0630 | 0,0618 | 0,0606 | 0,0594 | 0,0582 | 0,0571 | 0,0559 |
| -1,4 | 0,0808 | 0,0793 | 0,0778 | 0,0764 | 0,0749 | 0,0735 | 0,0721 | 0,0708 | 0,0694 | 0,0681 |
| -1,3 | 0,0968 | 0,0951 | 0,0934 | 0,0918 | 0,0901 | 0,0855 | 0,0869 | 0,0853 | 0,0838 | 0,0823 |
| -1,2 | 0,1151 | 0,1131 | 0,1112 | 0,1093 | 0,1075 | 0,1056 | 0,1038 | 0,1020 | 0,1003 | 0,0985 |
| -1,1 | 0,1357 | 0,1335 | 0,1314 | 0,1292 | 0,1271 | 0,1251 | 0,1230 | 0,1210 | 0,1190 | 0,1170 |
| -1,0 | 0,1587 | 0,1562 | 0,1539 | 0,1515 | 0,1492 | 0,1469 | 0,1446 | 0,1423 | 0,1401 | 0,1379 |
| -0,9 | 0,1841 | 0,1814 | 0,1788 | 0,1762 | 0,1736 | 0,1711 | 0,1685 | 0,1660 | 0,1631 | 0,1611 |
| -0,8 | 0,2119 | 0,2090 | 0,2061 | 0,2033 | 0,2005 | 0,1977 | 0,1949 | 0,1922 | 0,1894 | 0,1867 |
| -0,7 | 0,2420 | 0,2389 | 0,2358 | 0,2327 | 0,2296 | 0,2266 | 0,2236 | 0,2206 | 0,2177 | 0,2148 |
| -0,6 | 0,2743 | 0,2709 | 0,2676 | 0,2643 | 0,2611 | 0,2578 | 0,2546 | 0,2514 | 0,2483 | 0,2451 |
| -0,5 | 0,3085 | 0,3050 | 0,3015 | 0,2981 | 0,2946 | 0,2912 | 0,2877 | 0,2843 | 0,2810 | 0,2776 |

Continúa

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| -0,4 | 0,3446 | 0,3409 | 0,3372 | 0,3336 | 0,3300 | 0,3264 | 0,3228 | 0,3192 | 0,3156 | 0,3121 |
| -0,3 | 0,3821 | 0,3783 | 0,3745 | 0,3707 | 0,3669 | 0,3632 | 0,3594 | 0,3557 | 0,3520 | 0,3483 |
| -0,2 | 0,4207 | 0,4168 | 0,4129 | 0,4090 | 0,4052 | 0,4013 | 0,3974 | 0,3936 | 0,3897 | 0,3859 |
| -0,1 | 0,4602 | 0,4562 | 0,4522 | 0,4483 | 0,4443 | 0,4404 | 0,4364 | 0,4325 | 0,4286 | 0,4247 |
| 0,0 | 0,5000 | 0,4960 | 0,4920 | 0,4880 | 0,4840 | 0,4801 | 0,4761 | 0,7210 | 0,4681 | 0,4641 |
| | | | | | | | | | | |



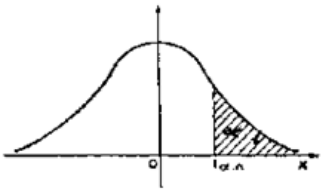

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | |
| Z | 0,00 | 0,01 | 0,02 | 0,03 | 0,04 | 0,05 | 0,06 | 0,07 | 0,08 | 0,09 |
| 0,0 | 0,5000 | 0,5040 | 0,5080 | 0,5120 | 0,5160 | 0,5199 | 0,5239 | 0,5279 | 0,5319 | 0,5359 |
| 0,1 | 0,5398 | 0,5438 | 0,5478 | 0,5517 | 0,5557 | 0,5596 | 0,5636 | 0,5675 | 0,5714 | 0,5753 |
| 0,2 | 0,5793 | 0,5832 | 0,5871 | 0,5910 | 0,5948 | 0,5987 | 0,6026 | 0,6064 | 0,6103 | 0,6141 |
| 0,3 | 0,6179 | 0,6217 | 0,6255 | 0,6293 | 0,6331 | 0,6368 | 0,6406 | 0,6443 | 0,6480 | 0,6517 |
| 0,4 | 0,6554 | 0,6591 | 0,6628 | 0,6664 | 0,6700 | 0,6736 | 0,6772 | 0,6808 | 0,6844 | 0,6879 |
| 0,5 | 0,6915 | 0,6950 | 0,6985 | 0,7019 | 0,7054 | 0,7088 | 0,7123 | 0,7157 | 0,7190 | 0,7224 |
| 0,6 | 0,7257 | 0,7291 | 0,7324 | 0,7357 | 0,7389 | 0,7422 | 0,7454 | 0,7486 | 0,7517 | 0,7549 |
| 0,7 | 0,7580 | 0,7611 | 0,7642 | 0,7673 | 0,7704 | 0,7734 | 0,7764 | 0,7794 | 0,7823 | 0,7852 |
| 0,8 | 0,7881 | 0,7910 | 0,7939 | 0,7967 | 0,7995 | 0,8023 | 0,8051 | 0,8078 | 0,8106 | 0,8133 |
| 0,9 | 0,8159 | 0,8186 | 0,8212 | 0,8238 | 0,8264 | 0,8289 | 0,8315 | 0,8340 | 0,8365 | 0,8389 |
| 1,0 | 0,8413 | 0,8438 | 0,8461 | 0,8485 | 0,8508 | 0,8531 | 0,8554 | 0,8577 | 0,8599 | 0,8621 |
| 1,1 | 0,8643 | 0,8665 | 0,8686 | 0,8708 | 0,8729 | 0,8749 | 0,8770 | 0,8790 | 0,8810 | 0,8830 |
| 1,2 | 0,8849 | 0,8869 | 0,8888 | 0,8907 | 0,8925 | 0,8944 | 0,8962 | 0,8980 | 0,8997 | 0,9015 |
| 1,3 | 0,9032 | 0,9049 | 0,9066 | 0,9082 | 0,9099 | 0,9115 | 0,9131 | 0,9147 | 0,9162 | 0,9177 |
| 1,4 | 0,9192 | 0,9207 | 0,9222 | 0,9236 | 0,9251 | 0,9265 | 0,9279 | 0,9292 | 0,9306 | 0,9319 |
| 1,5 | 0,9332 | 0,9345 | 0,9357 | 0,9370 | 0,9382 | 0,9394 | 0,9406 | 0,9418 | 0,9429 | 0,9441 |
| 1,6 | 0,9452 | 0,9463 | 0,9474 | 0,9484 | 0,9495 | 0,9505 | 0,9515 | 0,9525 | 0,9535 | 0,9545 |

Continúa

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 1,7 | 0,9554 | 0,9564 | 0,9573 | 0,9582 | 0,9591 | 0,9599 | 0,9608 | 0,9616 | 0,9625 | 0,9633 |
| 1,8 | 0,9641 | 0,9649 | 0,9656 | 0,9664 | 0,9671 | 0,9678 | 0,9686 | 0,9693 | 0,9699 | 0,9706 |
| 1,9 | 0,9713 | 0,9719 | 0,9726 | 0,9732 | 0,9738 | 0,9744 | 0,9750 | 0,9756 | 0,9761 | 0,9767 |
| 2,0 | 0,9772 | 0,9778 | 0,9783 | 0,9788 | 0,9793 | 0,9798 | 0,9803 | 0,9808 | 0,9812 | 0,9817 |
| 2,1 | 0,9821 | 0,9826 | 0,9830 | 0,9834 | 0,9838 | 0,9842 | 0,9846 | 0,9850 | 0,9854 | 0,9857 |
| 2,2 | 0,9861 | 0,9864 | 0,9868 | 0,9871 | 0,9875 | 0,9878 | 0,9881 | 0,9884 | 0,9887 | 0,9890 |
| 2,3 | 0,9893 | 0,9896 | 0,9898 | 0,9901 | 0,9904 | 0,9906 | 0,9909 | 0,9911 | 0,9913 | 0,9916 |
| 2,4 | 0,9918 | 0,9920 | 0,9922 | 0,9925 | 0,9927 | 0,9929 | 0,9931 | 0,9932 | 0,9934 | 0,9936 |
| 2,5 | 0,9938 | 0,9940 | 0,9941 | 0,9943 | 0,9945 | 0,9946 | 0,9948 | 0,9949 | 0,9951 | 0,9952 |
| 2,6 | 0,9953 | 0,9955 | 0,9956 | 0,9957 | 0,9959 | 0,9960 | 0,9961 | 0,9962 | 0,9963 | 0,9964 |
| 2,7 | 0,9965 | 0,9966 | 0,9967 | 0,9968 | 0,9969 | 0,9970 | 0,9971 | 0,9972 | 0,9973 | 0,9974 |
| 2,8 | 0,9974 | 0,9975 | 0,9976 | 0,9977 | 0,9977 | 0,9978 | 0,9979 | 0,9979 | 0,9980 | 0,9981 |
| 2,9 | 0,9981 | 0,9982 | 0,9982 | 0,9983 | 0,9984 | 0,9984 | 0,9985 | 0,9985 | 0,9986 | 0,9986 |
| 3,0 | 0,9987 | 0,9987 | 0,9988 | 0,9988 | 0,9989 | 0,9989 | 0,9989 | 0,9989 | 0,9990 | 0,9990 |

Fuente: ebookbrowse, 2012.

**Tabla 2. Distribución t**




| gl \ α/2 | 0,40 | 0,30 | 0,20 | 0,10 | 0,050 | 0,025 | 0,010 | 0,005 | 0,001 | 0,0005 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0,325 | 0,727 | 1,376 | 3,078 | 6,314 | 12,71 | 31,82 | 63,66 | 318,3 | 636,6 |
| 2 | 0,289 | 0,617 | 1,061 | 1,886 | 2,920 | 4,303 | 6,965 | 9,925 | 22,33 | 31,60 |
| 3 | 0,277 | 0,584 | 0,978 | 1,638 | 2,353 | 3,182 | 4,541 | 5,841 | 10,22 | 12,94 |
| 4 | 0,271 | 0,569 | 0,941 | 1,533 | 2,132 | 2,776 | 3,747 | 4,604 | 7,173 | 8,610 |
| 5 | 0,267 | 0,559 | 0,920 | 1,476 | 2,015 | 2,571 | 3,365 | 4,032 | 5,893 | 6,859 |
| 6 | 0,265 | 0,553 | 0,906 | 1,440 | 1,943 | 2,447 | 3,143 | 3,707 | 5,208 | 5,959 |
| 7 | 0,263 | 0,549 | 0,896 | 1,415 | 1,895 | 2,365 | 2,998 | 3,499 | 4,785 | 5,405 |
| 8 | 0,262 | 0,546 | 0,889 | 1,397 | 1,860 | 2,306 | 2,896 | 3,355 | 4,501 | 5,041 |
| 9 | 0,261 | 0,543 | 0,883 | 1,383 | 1,833 | 2,262 | 2,821 | 3,250 | 4,297 | 4,781 |
| 10 | 0,260 | 0,542 | 0,879 | 1,372 | 1,812 | 2,228 | 2,764 | 3,169 | 4,144 | 4,587 |
| 11 | 0,260 | 0,540 | 0,876 | 1,363 | 1,796 | 2,201 | 2,718 | 3,106 | 4,025 | 4,437 |
| 12 | 0,259 | 0,539 | 0,873 | 1,356 | 1,782 | 2,179 | 2,681 | 3,055 | 3,930 | 4,318 |
| 13 | 0,259 | 0,538 | 0,870 | 1,350 | 1,771 | 2,160 | 2,650 | 3,012 | 3,852 | 4,221 |
| 14 | 0,258 | 0,537 | 0,868 | 1,345 | 1,761 | 2,145 | 2,624 | 2,977 | 3,787 | 4,140 |
| 15 | 0,258 | 0,536 | 0,866 | 1,341 | 1,753 | 2,131 | 2,602 | 2,947 | 3,733 | 4,073 |
| 16 | 0,258 | 0,535 | 0,865 | 1,337 | 1,746 | 2,120 | 2,583 | 2,921 | 3,686 | 4,015 |
| 17 | 0,257 | 0,534 | 0,863 | 1,333 | 1,740 | 2,110 | 2,567 | 2,898 | 3,646 | 3,965 |
| 18 | 0,257 | 0,534 | 0,862 | 1,330 | 1,734 | 2,101 | 2,552 | 2,878 | 3,611 | 3,922 |
| 19 | 0,257 | 0,533 | 0,861 | 1,328 | 1,729 | 2,093 | 2,539 | 2,861 | 3,579 | 3,883 |
| 20 | 0,257 | 0,533 | 0,860 | 1,325 | 1,725 | 2,086 | 2,528 | 2,845 | 3,552 | 3,850 |
| 21 | 0,257 | 0,532 | 0,859 | 1,323 | 1,721 | 2,080 | 2,518 | 2,831 | 3,527 | 3,819 |
| 22 | 0,256 | 0,532 | 0,858 | 1,321 | 1,717 | 2,074 | 2,508 | 2,819 | 3,505 | 3,792 |
| 23 | 0,256 | 0,532 | 0,858 | 1,319 | 1,714 | 2,069 | 2,500 | 2,807 | 3,485 | 3,767 |
| 24 | 0,256 | 0,531 | 0,857 | 1,318 | 1,711 | 2,064 | 2,192 | 2,797 | 3,467 | 3,745 |
| 25 | 0,256 | 0,531 | 0,856 | 1,316 | 1,708 | 2,060 | 2,485 | 2,787 | 3,450 | 3,725 |
| 26 | 0,256 | 0,531 | 0,856 | 1,315 | 1,706 | 2,056 | 2,479 | 2,779 | 3,435 | 3,707 |
| 27 | 0,256 | 0,531 | 0,855 | 1,314 | 1,703 | 2,052 | 2,473 | 2,771 | 3,421 | 3,690 |
| 28 | 0,256 | 0,530 | 0,855 | 1,313 | 1,701 | 2,048 | 2,467 | 2,763 | 3,408 | 3,674 |
| 29 | 0,256 | 0,530 | 0,854 | 1,311 | 1,699 | 2,045 | 2,462 | 2,756 | 3,396 | 3,659 |
| 30 | 0,256 | 0,530 | 0,854 | 1,310 | 1,697 | 2,042 | 2,457 | 2,750 | 3,385 | 3,646 |
| 40 | 0,255 | 0,259 | 0,851 | 1,303 | 1,648 | 2,021 | 2,423 | 2,704 | 3,307 | 3,551 |
| 50 | 0,255 | 0,528 | 0,849 | 1,298 | 1,676 | 2,009 | 2,403 | 2,678 | 3,262 | 3,495 |
| 60 | 0,254 | 0,527 | 0,848 | 1,296 | 1,671 | 2,000 | 2,390 | 2,660 | 3,232 | 3,460 |
| 80 | 0,254 | 0,527 | 0,846 | 1,292 | 1,664 | 1,990 | 2,374 | 2,639 | 3,195 | 3,415 |
| 100 | 0,254 | 0,256 | 0,845 | 1,290 | 1,660 | 1,984 | 2,365 | 2,626 | 3,174 | 3,389 |
| 200 | 0,254 | 0,525 | 0,843 | 1,286 | 1,653 | 1,972 | 2,345 | 2,601 | 3,131 | 3,339 |
| 500 | 0,253 | 0,525 | 0,842 | 1,283 | 1,648 | 1,965 | 2,334 | 2,586 | 3,106 | 3,310 |
| ∞ | 0,253 | 0,524 | 0,842 | 1,282 | 1,645 | 1,960 | 2,326 | 2,576 | 3,090 | 3,291 |

**Tabla 3. Tabla de distribución F de Fisher con probabilidad de 0,05**

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 12 | 15 | 20 | 24 | 30 | 40 | 50 | 60 | 70 | 80 | 100 | 120 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 161.45 | 199.50 | 215.71 | 224.58 | 230.16 | 233.99 | 236.77 | 238.88 | 240.54 | 241.88 | 243.90 | 245.95 | 248.02 | 249.05 | 250.10 | 251.14 | 251.77 | 252.20 | 252.50 | 252.72 | 253.04 | 253.25 |
| 2 | 18.513 | 19.000 | 19.164 | 19.247 | 19.296 | 19.329 | 19.353 | 19.371 | 19.385 | 19.396 | 19.412 | 19.429 | 19.446 | 19.454 | 19.463 | 19.471 | 19.476 | 19.48 | 19.48 | 19.48 | 19.49 | 19.49 |
| 3 | 10.128 | 9.552 | 9.277 | 9.117 | 9.013 | 8.941 | 8.887 | 8.845 | 8.812 | 8.785 | 8.745 | 8.703 | 8.660 | 8.638 | 8.617 | 8.594 | 8.581 | 8.572 | 8.566 | 8.561 | 8.554 | 8.549 |
| 4 | 7.709 | 6.944 | 6.591 | 6.388 | 6.256 | 6.163 | 6.094 | 6.041 | 5.999 | 5.964 | 5.912 | 5.858 | 5.803 | 5.774 | 5.746 | 5.717 | 5.699 | 5.688 | 5.679 | 5.673 | 5.664 | 5.658 |
| 5 | 6.608 | 5.786 | 5.409 | 5.192 | 5.050 | 4.950 | 4.876 | 4.818 | 4.772 | 4.735 | 4.678 | 4.619 | 4.558 | 4.527 | 4.496 | 4.464 | 4.444 | 4.431 | 4.422 | 4.415 | 4.405 | 4.398 |
| 6 | 5.987 | 5.143 | 4.757 | 4.534 | 4.387 | 4.284 | 4.207 | 4.147 | 4.099 | 4.060 | 4.000 | 3.938 | 3.874 | 3.841 | 3.808 | 3.774 | 3.754 | 3.74 | 3.73 | 3.722 | 3.712 | 3.705 |
| 7 | 5.591 | 4.737 | 4.347 | 4.120 | 3.972 | 3.866 | 3.787 | 3.726 | 3.677 | 3.637 | 3.575 | 3.511 | 3.445 | 3.410 | 3.376 | 3.340 | 3.319 | 3.304 | 3.294 | 3.286 | 3.275 | 3.267 |
| 8 | 5.318 | 4.459 | 4.066 | 3.838 | 3.688 | 3.581 | 3.500 | 3.438 | 3.388 | 3.347 | 3.284 | 3.218 | 3.150 | 3.115 | 3.079 | 3.043 | 3.020 | 3.005 | 2.994 | 2.986 | 2.975 | 2.967 |
| 9 | 5.117 | 4.256 | 3.863 | 3.633 | 3.482 | 3.374 | 3.293 | 3.230 | 3.179 | 3.137 | 3.073 | 3.006 | 2.936 | 2.900 | 2.864 | 2.826 | 2.803 | 2.787 | 2.776 | 2.768 | 2.756 | 2.748 |
| 10 | 4.965 | 4.103 | 3.708 | 3.478 | 3.326 | 3.217 | 3.135 | 3.072 | 3.020 | 2.978 | 2.913 | 2.845 | 2.774 | 2.737 | 2.700 | 2.661 | 2.637 | 2.621 | 2.609 | 2.601 | 2.588 | 2.580 |
| 11 | 4.844 | 3.982 | 3.587 | 3.357 | 3.204 | 3.095 | 3.012 | 2.948 | 2.896 | 2.854 | 2.788 | 2.719 | 2.646 | 2.609 | 2.570 | 2.531 | 2.507 | 2.490 | 2.478 | 2.469 | 2.457 | 2.448 |
| 12 | 4.747 | 3.885 | 3.490 | 3.259 | 3.106 | 2.996 | 2.913 | 2.849 | 2.796 | 2.753 | 2.687 | 2.617 | 2.544 | 2.505 | 2.466 | 2.426 | 2.401 | 2.384 | 2.372 | 2.363 | 2.350 | 2.341 |
| 13 | 4.667 | 3.806 | 3.411 | 3.179 | 3.025 | 2.915 | 2.832 | 2.767 | 2.714 | 2.671 | 2.604 | 2.533 | 2.459 | 2.420 | 2.380 | 2.339 | 2.314 | 2.297 | 2.284 | 2.275 | 2.261 | 2.252 |
| 14 | 4.600 | 3.739 | 3.344 | 3.112 | 2.958 | 2.848 | 2.764 | 2.699 | 2.646 | 2.602 | 2.534 | 2.463 | 2.388 | 2.349 | 2.308 | 2.266 | 2.241 | 2.223 | 2.210 | 2.201 | 2.187 | 2.178 |
| 15 | 4.543 | 3.682 | 3.287 | 3.056 | 2.901 | 2.790 | 2.707 | 2.641 | 2.588 | 2.544 | 2.475 | 2.403 | 2.328 | 2.288 | 2.247 | 2.204 | 2.178 | 2.160 | 2.147 | 2.137 | 2.123 | 2.114 |
| 16 | 4.494 | 3.634 | 3.239 | 3.007 | 2.852 | 2.741 | 2.657 | 2.591 | 2.538 | 2.494 | 2.425 | 2.352 | 2.276 | 2.235 | 2.194 | 2.151 | 2.124 | 2.106 | 2.093 | 2.083 | 2.068 | 2.059 |
| 17 | 4.451 | 3.592 | 3.197 | 2.965 | 2.810 | 2.699 | 2.614 | 2.548 | 2.494 | 2.450 | 2.381 | 2.308 | 2.230 | 2.190 | 2.148 | 2.104 | 2.077 | 2.058 | 2.045 | 2.035 | 2.020 | 2.011 |
| 18 | 4.414 | 3.555 | 3.160 | 2.928 | 2.773 | 2.661 | 2.577 | 2.510 | 2.456 | 2.412 | 2.342 | 2.269 | 2.191 | 2.150 | 2.107 | 2.063 | 2.035 | 2.017 | 2.003 | 1.993 | 1.978 | 1.968 |
| 19 | 4.381 | 3.522 | 3.127 | 2.895 | 2.740 | 2.628 | 2.544 | 2.477 | 2.423 | 2.378 | 2.308 | 2.234 | 2.155 | 2.114 | 2.071 | 2.026 | 1.999 | 1.980 | 1.966 | 1.955 | 1.940 | 1.930 |

Continúa

| | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 20 | 4.351 | 3.493 | 3.098 | 2.866 | 2.711 | 2.599 | 2.514 | 2.447 | 2.393 | 2.348 | 2.278 | 2.203 | 2.124 | 2.082 | 2.039 | 1.994 | 1.966 | 1.946 | 1.932 | 1.922 | 1.907 | 1.896 |
| 21 | 4.325 | 3.467 | 3.072 | 2.840 | 2.685 | 2.573 | 2.488 | 2.420 | 2.366 | 2.321 | 2.250 | 2.176 | 2.096 | 2.054 | 2.010 | 1.965 | 1.936 | 1.916 | 1.902 | 1.891 | 1.876 | 1.866 |
| 22 | 4.301 | 3.443 | 3.049 | 2.817 | 2.661 | 2.549 | 2.464 | 2.397 | 2.342 | 2.297 | 2.226 | 2.151 | 2.071 | 2.028 | 1.984 | 1.938 | 1.909 | 1.889 | 1.875 | 1.864 | 1.849 | 1.838 |
| 23 | 4.279 | 3.422 | 3.028 | 2.796 | 2.640 | 2.528 | 2.442 | 2.375 | 2.320 | 2.275 | 2.204 | 2.128 | 2.048 | 2.005 | 1.961 | 1.914 | 1.885 | 1.865 | 1.850 | 1.839 | 1.823 | 1.813 |
| 24 | 4.260 | 3.403 | 3.009 | 2.776 | 2.621 | 2.508 | 2.423 | 2.355 | 2.300 | 2.255 | 2.183 | 2.108 | 2.027 | 1.984 | 1.939 | 1.892 | 1.863 | 1.842 | 1.828 | 1.816 | 1.800 | 1.790 |
| 25 | 4.242 | 3.385 | 2.991 | 2.759 | 2.603 | 2.490 | 2.405 | 2.337 | 2.282 | 2.236 | 2.165 | 2.089 | 2.007 | 1.964 | 1.919 | 1.872 | 1.842 | 1.822 | 1.807 | 1.796 | 1.779 | 1.768 |
| 26 | 4.225 | 3.369 | 2.975 | 2.743 | 2.587 | 2.474 | 2.388 | 2.321 | 2.265 | 2.220 | 2.148 | 2.072 | 1.990 | 1.946 | 1.901 | 1.853 | 1.823 | 1.803 | 1.788 | 1.776 | 1.76 | 1.749 |
| 27 | 4.210 | 3.354 | 2.960 | 2.728 | 2.572 | 2.459 | 2.373 | 2.305 | 2.250 | 2.204 | 2.132 | 2.056 | 1.974 | 1.930 | 1.884 | 1.836 | 1.806 | 1.785 | 1.770 | 1.758 | 1.742 | 1.731 |
| 28 | 4.196 | 3.340 | 2.947 | 2.714 | 2.558 | 2.445 | 2.359 | 2.291 | 2.236 | 2.190 | 2.118 | 2.041 | 1.959 | 1.915 | 1.869 | 1.820 | 1.790 | 1.769 | 1.754 | 1.742 | 1.725 | 1.714 |
| 29 | 4.183 | 3.328 | 2.934 | 2.701 | 2.545 | 2.432 | 2.346 | 2.278 | 2.223 | 2.177 | 2.104 | 2.027 | 1.945 | 1.901 | 1.854 | 1.806 | 1.775 | 1.754 | 1.738 | 1.726 | 1.71 | 1.698 |
| 30 | 4.171 | 3.316 | 2.922 | 2.690 | 2.534 | 2.421 | 2.334 | 2.266 | 2.211 | 2.165 | 2.092 | 2.015 | 1.932 | 1.887 | 1.841 | 1.792 | 1.761 | 1.740 | 1.724 | 1.712 | 1.695 | 1.683 |
| 35 | 4.121 | 3.267 | 2.874 | 2.641 | 2.485 | 2.372 | 2.285 | 2.217 | 2.161 | 2.114 | 2.041 | 1.963 | 1.878 | 1.833 | 1.786 | 1.735 | 1.703 | 1.681 | 1.665 | 1.652 | 1.635 | 1.623 |
| 40 | 4.085 | 3.232 | 2.839 | 2.606 | 2.449 | 2.336 | 2.249 | 2.180 | 2.124 | 2.077 | 2.003 | 1.924 | 1.839 | 1.793 | 1.744 | 1.693 | 1.660 | 1.637 | 1.621 | 1.608 | 1.589 | 1.577 |
| 45 | 4.057 | 3.204 | 2.812 | 2.579 | 2.422 | 2.308 | 2.221 | 2.152 | 2.096 | 2.049 | 1.974 | 1.895 | 1.808 | 1.762 | 1.713 | 1.660 | 1.626 | 1.603 | 1.586 | 1.573 | 1.554 | 1.541 |
| 50 | 4.034 | 3.183 | 2.790 | 2.557 | 2.400 | 2.286 | 2.199 | 2.130 | 2.073 | 2.026 | 1.952 | 1.871 | 1.784 | 1.737 | 1.687 | 1.634 | 1.599 | 1.576 | 1.558 | 1.544 | 1.525 | 1.511 |
| 60 | 4.001 | 3.150 | 2.758 | 2.525 | 2.368 | 2.254 | 2.167 | 2.097 | 2.040 | 1.993 | 1.917 | 1.836 | 1.748 | 1.700 | 1.649 | 1.594 | 1.559 | 1.534 | 1.516 | 1.502 | 1.481 | 1.467 |
| 70 | 3.978 | 3.128 | 2.736 | 2.503 | 2.346 | 2.231 | 2.143 | 2.074 | 2.017 | 1.969 | 1.893 | 1.812 | 1.722 | 1.674 | 1.622 | 1.566 | 1.530 | 1.505 | 1.486 | 1.471 | 1.45 | 1.435 |
| 80 | 3.960 | 3.111 | 2.719 | 2.486 | 2.329 | 2.214 | 2.126 | 2.056 | 1.999 | 1.951 | 1.875 | 1.793 | 1.703 | 1.654 | 1.602 | 1.545 | 1.508 | 1.482 | 1.463 | 1.448 | 1.426 | 1.411 |
| 90 | 3.947 | 3.098 | 2.706 | 2.473 | 2.316 | 2.201 | 2.113 | 2.043 | 1.986 | 1.938 | 1.861 | 1.779 | 1.688 | 1.639 | 1.586 | 1.528 | 1.491 | 1.465 | 1.445 | 1.429 | 1.407 | 1.391 |
| 100 | 3.936 | 3.087 | 2.696 | 2.463 | 2.305 | 2.191 | 2.103 | 2.032 | 1.975 | 1.927 | 1.850 | 1.768 | 1.676 | 1.627 | 1.573 | 1.515 | 1.477 | 1.450 | 1.430 | 1.415 | 1.392 | 1.376 |
| 120 | 3.920 | 3.072 | 2.680 | 2.447 | 2.290 | 2.175 | 2.087 | 2.016 | 1.959 | 1.910 | 1.834 | 1.750 | 1.659 | 1.608 | 1.554 | 1.495 | 1.457 | 1.429 | 1.408 | 1.392 | 1.369 | 1.352 |

Fuente: López, 2010.

Continúa

**Universidad del Rosario**
Facultad de Administración