ANÁLISIS DE DATOS AGROPECUARIOS

IVÁN RAMÍREZ-MORALES / BERTHA MAZON-OLIVO

Editorial UTMACH

REDES 2017
COLECCIÓN EDITORIAL

Análisis de Datos Agropecuarios

Iván Ramírez-Morales
Bertha Mazon-Olivo

Coordinadores

Primera edición en español, 2018

Este texto ha sido sometido a un proceso de evaluación por pares externos con base en la normativa editorial de la UTMACH

Ediciones UTMACH

Gestión de proyectos editoriales universitarios

302 pag; 22X19cm - (Colección REDES 2017)

Título: Análisis de Datos Agropecuarios. / Iván Ramírez-Morales / Bertha Mazon-Olivo (Coordinadores)

ISBN: 978-9942-24-120-7

Publicación digital

Título del libro: Análisis de Datos Agropecuarios.

ISBN: 978-9942-24-120-7

Comentarios y sugerencias: editorial@utmachala.edu.ec

Diseño de portada: MZ Diseño Editorial

Diagramación: MZ Diseño Editorial

Diseño y comunicación digital: Jorge Maza Córdova, Ms.

Km. 5 1/2 Vía Machala Pasaje

www.utmachala.edu.ec

Machala - Ecuador

Índice

Capítulo I

Ciencia de datos en el sector agropecuario 12

Iván Ramírez-Morales; Bertha Mazon-Olivo ;Alberto Pan

Capítulo II

Obtención de datos en sistemas agropecuarios 45

Salomón Barrezueta Unda; Diego Villaseñor Ortiz

Capítulo III

Internet de las cosas (IoT) 72

Dixys Hernández Rojas; Bertha Mazon-Olivo; Carlos Escudero

Capítulo IV

Matemáticas aplicadas al sector agropecuario 101

Bladimir Serrano; Carlos Loor; Eduardo Tusa

Capítulo V

Estadística básica con datos agropecuarios 127

Irán Rodríguez Delgado; Bill Serrano; Diego Villaseñor Ortiz

Capítulo VI

Estadística predictiva con datos agropecuarios 218

Bill Serrano; Irán Rodríguez Delgado

Capítulo VII

Inteligencia de negocios en el sector agropecuario 246

Bertha Mazon-Olivo; Alberto Pan; Raquel Tinoco-Egas

Capítulo VIII

Inteligencia Artificial aplicada a datos agropecuarios 278

Iván Ramírez-Morales; Eduardo Tusa; Daniel Rivero

Introducción

El análisis de datos es un proceso complejo que trata de encontrar patrones útiles y relaciones entre los datos a fin de obtener información sobre un problema específico y de esta manera tomar decisiones acertadas para su solución.

Las técnicas de análisis de datos que son exploradas en el presente libro son actualmente utilizadas en diversos sectores de la economía. En un inicio, fueron empleadas por las grandes empresas a fin de incrementar sus rendimientos financieros.

El libro se basa en la aplicación de la especialización inteligente, de este modo, gracias al trabajo colaborativo, se combina al sector agropecuario con las tecnologías, matemáticas, estadística y las ciencias computacionales, para la optimización de los procesos productivos.

La idea de descubrir la información oculta en las relaciones entre los datos, incentiva a encontrar aplicaciones para el sector agropecuario, por ejemplo los obtenidos de una producción avícola, o los datos que se generan durante los procesos de fermentación, los parámetros físicos y químicos del suelo, del agua y de las plantas, los datos de sensores, de espectrometría, entre otros.

En la actualidad, este sector se ha mantenido con su producción habitual sin un destacado repunte ni diferenciación, a pesar de existir herramientas científicas que han permitido desarrollar dispositivos tecnológicos y sus aplicaciones.

Este libro ha sido el resultado de la sistematización de las experiencias individuales de un equipo humano con objetivos comunes y una historia académica multidisciplinar, cuyos hallazgos de investigación han sido publicados en revistas científicas y conferencias de alto impacto. El área temática sobre la que se centra este texto es en técnicas de extracción, procesamiento y análisis de datos del ámbito agropecuario, se combinan para entregar al lector una obra de calidad y alto valor científico.

Así, el presente libro está concebido desde diferentes puntos de vista de profesionales agrónomos, informáticos, electrónicos, matemáticos, estadísticos y empresarios. Todos buscan un objetivo en común: "descubrir el conocimiento oculto en los datos que proporcione una ventaja competitiva". Se aborda el ciclo completo del proceso de obtención de conocimiento a partir de datos crudos del sector agropecuario, con la finalidad de apoyar la toma de decisiones. Este ciclo involucra procesos de: selección de los datos (extracción, comunicación, almacenamiento), pre-procesamiento, transformación, aplicación de modelos y/o técnicas de análisis, presentación e interpretación de resultados. El enfoque temático del libro es el siguiente:

Capítulo 1: Ciencia de Datos en el sector Agropecuario.- En este capítulo se aborda una revisión desde los inicios del análisis de datos en el sector agropecuario hasta el progreso actual que se ha dado en esta área del conocimiento que se considera como la nueva revolución en la agricultura y la ganadería de precisión.

Capítulo 2: Obtención de datos en sistemas agropecuarios.- El enfoque del capítulo es la generación de datos crudos en los sistemas agropecuarios, aplicando métodos y técnicas básicas donde se registran información de: número de unidades producidas, cantidad de nutrientes, variables climáticas, muestreo y monitoreo de organismos vivos, entre otros.

Capítulo 3: Internet de las cosas (IoT).- Este capítulo aborda los sistemas de telemetría para obtención de datos y control de dispositivos, aplicando tecnologías como: redes de sensores inalámbricos (dispositivos electrónicos, sensores, actuadores y puertas de enlace), protocolos de comunicación, centros de procesamiento de datos (cloud computing) y aplicaciones IoT para el sector agropecuario.

Capítulo 4: Matemáticas aplicadas al sector agropecuario.- Este capítulo explica los procedimientos para la creación de modelos matemáticos determinísticos que representen procesos asociados al sector agropecuario, como una alternativa de solución en la ingeniería.

Capítulo 5: Estadística básica con datos agropecuarios.- El capítulo se enfoca en los atributos, escalas de medición de las variables, su influencia en la elección del procedimiento estadístico a desarrollar, así como, el papel de las medidas de resumen, estimación puntual y prueba de hipótesis en la investigación científica.

Capítulo 6: Estadística predictiva con datos agropecuarios.- El capítulo considera las principales técnicas de la estadística avanzada aplicada al sector agropecuario, con el propósito de establecer predicciones que permita tomar mejores decisiones.

Capítulo 7: Inteligencia de negocios en el sector agropecuario.- El capítulo comprende la obtención de conocimiento a partir de datos crudos con la finalidad de apoyar la toma de decisiones en empresas del sector agropecuario. Involucra procesos de extracción, transformación y almacenamiento de datos en nuevos almacenes (Data warehouse - Big Data), distribución y análisis de la información con técnicas: multidimensional OLAP y tableros de control (dashboards).

Capítulo 8: Inteligencia Artificial aplicada a datos agropecuarios.- El capítulo trata sobre las principales técnicas de machine learning aplicadas a los datos agropecuarios, entre éstas se destacan: las redes de neuronas artificiales, máquinas de soporte de vectores, vecinos más cercanos, análisis de componentes principales, entre otros.

06 Capítulo Estadística predictiva con datos agropecuarios

Bill Serrano; Irán Rodríguez Delgado

Las características de la época actual hacen necesario poseer conocimientos sólidos en el uso de las herramientas estadísticas avanzadas para poder inferir resultados sobre una población desde el análisis de una o varias muestras - con datos históricos y actuales -, que mediante el uso de la estadística inferencial se modelizan procesos, sistemas, respuestas con fines predictivos permitiendo extraer patrones de comportamiento para identificar riesgos, oportunidades y con su ayuda cosechar resultados esperados. En el sector agropecuario estas técnicas se hacen pertinentes al existir la necesidad de mejorar la productividad, donde la estadística inferencial sirve como eje transversal para tomar las decisiones acertadas para dicho propósito.

Bill Serrano: Ingeniero Agrónomo e Ingeniero en Gestión Empresarial, Magister en Administración de Empresas y estudiante doctoral en Análisis Económico y Estrategia Empresarial en la Universidad A Coruña. Fue Gerente de Almacén y Jefe Comercial Zonal en ICESA, Gerente de producto en ICESA y COMPTECO. Actualmente Profesor Titular en la Universidad Técnica de Machala.

Irán Rodríguez Delgado: Ingeniero Agrónomo (1992) Universidad Central de Las Villas, Cuba Magister en Agricultura Sostenible (2009) Universidad de Cienfuegos, Cuba; Investigador Agregado (2009) Instituto de Investigaciones de la Caña de Azúcar, Cuba; Profesor Titular (2015) Universidad Técnica de Machala. Autor de cuatro libros y 17 artículos publicados.

En el presente capítulo se hace exposición relacionada con las correlaciones, el análisis de regresión y la causalidad. Además, se hace una presentación específica relacionada con el análisis de regresión simple, análisis de regresión y el análisis de varianza, análisis de regresión múltiple, análisis de regresión con variables dicotómicas, finalizando con análisis de regresión y ANCOVA. Todas las técnicas descritas se aplicarán a datos agropecuarios agregando varios ejemplos ilustrativos, gráficos y los cálculos correspondientes se obtendrán usando los software estadísticos como stata y SPSS.

Análisis de regresión, correlación y causalidad

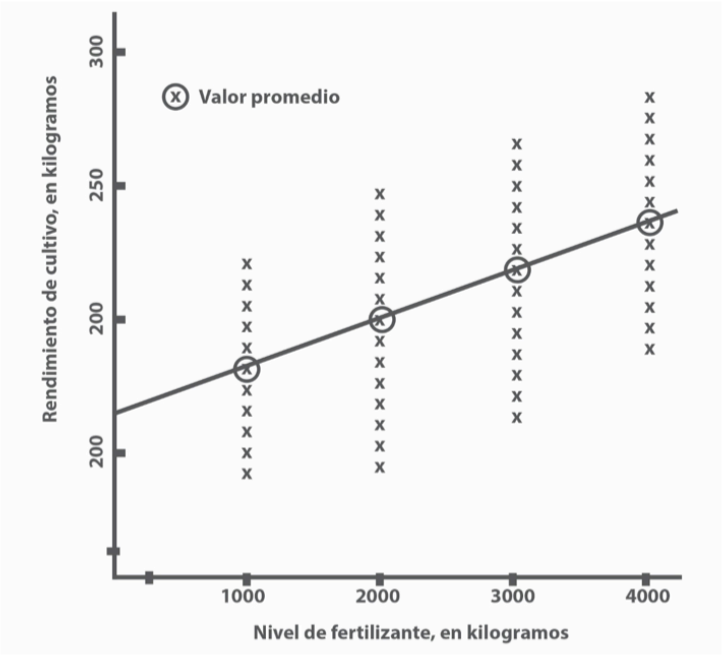
Interpretación de la Regresión

El análisis de regresión abarca el tratamiento de la dependencia de una variable (variable dependiente, variable explicada, predicha, regresada, variable de respuesta, endógena, resultado) respecto de una o varias variables (variables independientes, variables explicativas, predictora, regresora, estímulo, exógena, covariante, variable de control), con el propósito de estimar la media poblacional de la primera en términos de los valores conocidos de las segundas.

Cuando en el tratamiento de la dependencia de una variable se observa una sola variable independiente se denomina análisis de regresión simple y cuando existe más de una variable independiente toma el nombre de análisis de regresión múltiple.

Considere el siguiente ejemplo: A un agrónomo le interesa averiguar las razones de la estabilidad en la distribución del rendimiento de un cultivo dentro de una población. La regresión intenta averiguar cómo cambia el rendimiento promedio del cultivo dado la cantidad de fertilizante aplicado. Es decir, lo que intenta predecir es el rendimiento del cultivo a partir de la cantidad de fertilizante aplicado. Considere el Gráfico 6.1, correspondiente a un diagrama de dispersión. El gráfico muestra la distribución de los rendimientos del cultivo en una población hipotética, correspondientes al conjunto de valores conocidos del fertilizante aplicado. Preste atención que, para cualquier cantidad de fertilizante aplicado, existe un rango (distribución) de rendimientos. Sin embargo, observe también que, a pesar de la variabilidad del rendimiento del cultivo conforme al valor del fertilizante aplicado, el rendimiento promedio del cultivo aumenta, por lo general, en la medida en que lo hace el fertilizante aplicado. Los cruces dentro de los círculos en el gráfico establecen que el rendimiento promedio del cultivo corresponde a una cantidad determinada de fertilizante aplicado. Estos promedios se conectan para obtener la línea recta del gráfico. Esta línea, se conoce como recta de regresión. Dicha recta muestra que el rendimiento del cultivo aumenta conforme crece la cantidad de fertilizante aplicado.

Gráfico 6.1. Diagrama de dispersión



Fuente: Modificado de Gujarati (2006)

En el análisis de regresión lo pertinente es lo que se denomina dependencia estadística entre variables. Esta es propia de variables aleatorias o estocásticas, es decir, variables con distribución de probabilidad. Por ejemplo, siguiendo el ejemplo del rendimiento del cultivo, éste depende del fertilizante aplicado, sin embargo, también lo hace de la lluvia, temperatura, sol, entre otros, y tal dependencia es de naturaleza aleatoria porque las variables explicativas, a pesar de ser importantes, no permiten al agrónomo predecir de forma exacta el rendimiento del cultivo, debido a los errores propios de la medición de estas variables. Por tal motivo, existirá alguna variabilidad intrínseca en la variable explicada.

Regresión y causalidad

La causalidad en forma simple se dice que es el principio o el origen de algo. Este concepto es traído a la práctica para explicar la relación entre una causa y su efecto. En la estadística, este término explica la relación de necesidad de coocurrencia de dos variables.

A pesar de que la regresión establece la relación estadística que pueda existir entre la dependencia de una variable respecto a otras, y por más fuerte que esta sea, no implica que exista causalidad necesariamente. Para determinar la causalidad, es necesario acudir a consideraciones teóricas o a priori. Usando el ejemplo citado del rendimiento del cultivo, no existe motivo estadístico para suponer que el fertilizante depende del rendimiento del cultivo, sin embargo, la lógica indica que la relación es a la inversa, ya que no es posible controlar la cantidad de fertilizante aplicado mediante el rendimiento del cultivo.

Regresión y correlación

El análisis de regresión y la correlación se vincula de manera estrecha. Por un lado, el propósito principal del análisis de correlación es determinar el grado de asociación lineal entre dos variables, por ejemplo, si se desea conocer la correlación entre la lluvia y el rendimiento de un cultivo; entre la cantidad de fertilizante aplicado y el rendimiento de un cultivo; entre el balanceado utilizado y el crecimiento del camarón en una piscina; entre la salinidad del agua y la producción de tilapia. En cambio, en el análisis de regresión trata de estimar el valor promedio de una variable con base en los valores conocidos de otras. De tal manera, se desee estimar el promedio del rendimiento de una hectárea de camarón desde la cantidad de balanceado utilizado. La correlación al determinar el grado de asocial lineal entre las variables, me permite poner a consideración incluir o no las variables en el modelo de la regresión.

Hay que considerar que en un análisis de regresión existe una asimetría en el trato a las variables dependientes e independientes. También se supone que la variable dependiente tiene una distribución de probabilidad, es decir es estocástica, y se asume que las variables independientes tienen valores fijos o conocidos en muestreos repetitivos – las variables independientes pueden intrínsecamente ser estocásticas, pero para fines del análisis de regresión, se considera como supuesto que sus valores son fijos en el muestreo repetitivo, es decir que la variable explicativa toma los mismos valores en diferentes muestras. De esta forma, en el Gráfico 6.1 se supuso que la cantidad de fertilizante aplicado era fijo en los niveles dados y se obtuvieron rendimientos de los cultivos en esos niveles. En cambio, en el análisis de correlación, entre las variables que intervienen no existe distinción, de tal manera que, la correlación existente entre la cantidad de fertilizante aplicado y el rendimiento de cultivo, es la misma entre el rendimiento del cultivo y la cantidad de fertilizante aplicado.

Para interpretar la correlación, hay que conocer que toma valores en el intervalo de [-1,1], indicando el signo el sentido de la relación entre las variables. Por ejemplo: si el índice de correlación es de 1, se interpreta como una correlación positiva perfecta, es decir, existe una asociación lineal directa perfecta entre las variables: cuando una de ellas aumenta, la otra también lo hace en proporción constante. Es así que:

- Correlación =1, existe una asociación lineal positiva perfecta.

- Correlación mayor a 0 y menor a 1, existe una asociación lineal positiva.

- Correlación =0, no existe asociación lineal entre las variables.

- Correlación menor a 0 y mayor a -1, existe asociación lineal negativa.

- Correlación =-1, existe una asociación lineal negativa perfecta.

Para ilustrar el análisis de correlación, considere los datos del Cuadro 6.1, que detalla el rendimiento de un cultivo y el fertilizante aplicado. El coeficiente de correlación resultante es de 1, es decir, que existe una asociación lineal positiva perfecta. Cuando el fertilizante aplicado por hectárea aumenta, el rendimiento por hectárea también lo hace, sin embargo, que exista correlación entre las variables no quiere decir necesariamente que exista causalidad. La causalidad solo puede aceptarse cuando hay suficientes razones claras. La correlación no implica causalidad.

Cuadro 6.1. Rendimiento de un cultivo y el fertilizante aplicado

Obs	Rendimiento/Kg x Ha	Fertilizante /kg x Ha
1	200	100
2	250	125
3	300	150
4	350	175
5	400	200
6	450	225
7	500	250
8	550	275
9	600	300
10	650	325
11	700	350
12	750	375
13	800	400
14	850	425

Análisis de regresión simple

El análisis de regresión simple también conocido como análisis de regresión con dos variables, tiene como finalidad establecer el valor promedio de la variable dependiente, con base en los valores conocidos de una sola variable independiente. Para entender esto -mediante un ejemplo hipotético- consideremos los datos de la Cuadro 6.2. Estos datos se refieren a 80 piscinas de camarón, donde cada una tiene un tamaño de 1 hectárea, así como el nivel balanceado aplicado (X), en libras y la producción obtenida por corrida (110 días), en libras. Las 80 piscinas se dividen en 8 grupos por la cantidad de balanceado (de 2000 libras a 3200); de igual manera, se reflejan las producciones por corrida de cada piscina de los distintos grupos. Por efecto, existen 8 valores fijos de X y los correspondientes valores de Y (producción por corrida) para cada valor de X.

Cuadro 6.2. Rendimiento de un cultivo y el fertilizante aplicado

X	2000	2200	2400	2600	2800	3000	3200	3400
Y								
Producción por piscina por corrida Y, libras	1800	2000	2100	2300	2350	2420	2590	2790
	1850	2100	2150	2200	2360	2460	2610	2840
	2000	2050	2200	2000	2450	2500	2650	2880
	2050	2130	2000	2100	2500	2580	2800	3100
	1800	2120	1960	2050	2300	2420	2570	2800
	1900	2200	2200	2200	2330	2400	2550	2780
	2020	1980	2090	2340	2220	2340	2400	2640
	1890	1800	2100	2200	2280	2400	2550	2760
	1940	2100	2080	2150	2300	2400	2530	2760
	2000	2010	2120	2300	2230	2320	2470	2690
Total	19250	20490	21000	21840	23320	24240	25720	28040
Media condicional de Y	1925	2049	2100	2184	2332	2424	2572	2804

Se observa un cambio considerable entre la producción por corrida de cada grupo de balanceado. Es decir, que a pesar de la variabilidad de la producción en cada rango de balanceado considerado, en promedio, la producción se incrementa a medida que aumenta las libras de balanceado aplicado. En el Cuadro 6.2 se muestra la media de la producción por corrida que corresponde a cada uno de los 8 niveles de balanceado aplicado. De tal forma, el nivel de balanceado de 2000 libras por hectárea y por corrida le corresponde una media de producción de 1925 libras de camarón, sin embargo, al nivel de 3000, la media correspondiente es de 2424. Existen en total 8 medias, las que son denominadas como valores esperados condicionales, en consideración que dependen de los valores de la variable X (variable condicional).

Se hace necesario diferenciar el valor esperado incondicional y los valores esperados condicionales de la producción por corrida/ha. Si sumamos las 80 producciones que forman la población y la dividimos para 80, obtendremos la cantidad de 2298.75 (183900/80), que es el valor de la media incondicional, del nivel de balanceado aplicado; como no consideramos los niveles de balanceado aplicado es considerada incondicional. Como es de suponer, los distintos valores esperados condicionales de Y del Cuadro 6.2 difieren respecto del valor esperado incondicional de Y (2298.75). Siguiendo con ejemplo, cuando se desea conocer el valor esperado de la producción de una piscina por corrida, la respuesta es de 2298.75 (la media incondicional), sin embargo, si se desea saber cuál es el valor esperado de la producción de una piscina por corrida cuando se aplica 2400 libras de balanceado, la respuesta es 2100 libras (la media condicional). Por tal razón, conocer el nivel de balanceado permite predecir con mayor exactitud el valor medio de la producción. Esta es la esencia de la regresión.

Del anterior ejemplo expuesto, es evidente que cada media condicional está en función de X (lineal). Esta función explica como la media de Y varía con respecto a X. La función adopta una relación lineal, ya que un Biólogo o un productor se puede plantear que la producción de camarón tiene una relación lineal con el balanceado aplicado. Por tanto, inicialmente se puede aproximar que, Y es una función lineal de X, del tipo: $Y = \beta 0 + \beta 1X$, donde $\beta 0$ y $\beta 1$ son parámetros no conocidos pero fijos que se denominan coeficientes de regresión, también se conocen como coeficientes de intersección y de pendiente. Este conjunto de parámetros y variables se conoce como la ecuación de regresión, pero también se la identifica como regresión. En el análisis de regresión el objetivo es estimar los parámetros $\beta 0$ y $\beta 1$ con base en las observaciones de X y Y.

Método de los mínimos cuadrados

El método de los mínimos cuadrados se ha convertido en uno de los métodos más utilizados para el análisis de regresión por su eficacia y por la calidad de sus propiedades estadísticas, por tal motivo para realizar los análisis de regresión a forma da ejemplificar se utilizará el método de los mínimos cuadrados. En este sentido, los estimadores obtenidos se conocen como estimadores de mínimos cuadrados (MC), los que cumplen las siguientes propiedades:

- Se enuncian en términos de las cantidades observables, es decir de los X y Y que se encuentran en la muestra.
- Cada estimador proporciona un solo valor puntual del parámetro poblacional pertinente.
- Una vez generados los estimadores de MC, se obtiene la línea de la regresión muestral. La regresión obtenida tiene la propiedad de que el valor de la Y estimada (muestra) representa a la Y real (población).

Ejemplo Ilustrativo:

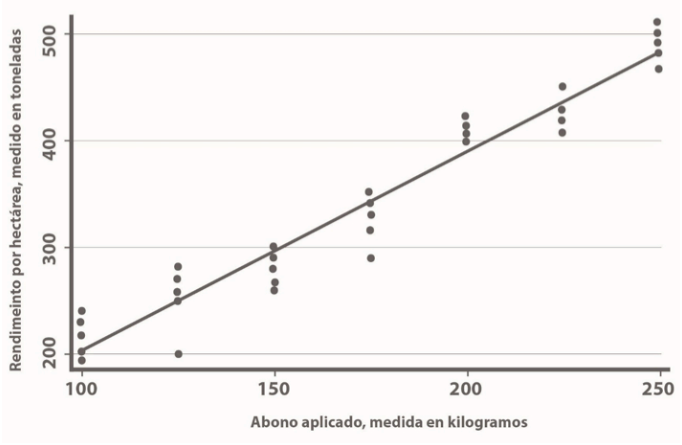
El Cuadro 6.3 proporciona datos sobre el rendimiento de un cultivo por hectárea (medido en tonelada) y el abono aplicado (medido en kilogramos). Mediante el software stata encontramos la ecuación de la regresión:

Y = 17.56429 + 1.856857X (es decir, β0=17.56429 y β1 = 1.856857).

r^2 = 0.9460

La línea de la regresión se muestra en el Gráfico 6.2

Gráfico 6.2. Distribución hipotética de rendimiento por hectárea correspondiente a la cantidad de abono aplicado



El valor de β1 =1.856857, que mide la pendiente, indica que, entre el intervalo muestral de X entre 100 y 250 kg de abono por hectárea, a medida que el valor X aumenta 1 kg, el incremento estimado en el rendimiento promedio por hectárea es 1.856857 toneladas. Es decir, que, por cada kg de abono adicional, en promedio, produce aumento en el rendimiento por hectárea de 1.856857 toneladas.

El valor de β0=17.56429, que pertenece al intercepto, enseña el nivel promedio del rendimiento del cultivo cuando la cantidad de abono aplicado es cero.

El valor de r^2 de 0.946 se interpreta que el nivel de abono aplicado explica el 94.6% de la variación del rendimiento por hectárea.

Antes de finalizar este ejemplo, considere que el modelo ejemplificado es muy sencillo. La teoría y la experiencia indican que, el rendimiento de un cultivo, a parte del abono, existen otros factores importantes en la determinación del rendimiento. Al ingresar más variables X al modelo, esta toma el nombre de regresión múltiple.

Cuadro 6.3. Rendimiento de un cultivo y el fertilizante aplicado

Abono /kg x Ha	Rendimiento/tn x Ha				
100	200	230	195	240	217
125	250	270	200	280	257
150	300	280	260	290	267
175	350	330	290	340	317
200	400	412	408	422	399
225	450	420	430	430	407
250	500	480	510	490	467

Análisis de regresión y análisis de varianza

En el análisis de la regresión al hacer el análisis de varianza (ANOVA) permite descomponer la variabilidad de la variable dependiente en variabilidad explicada por el modelo más la variabilidad no explicada, esto permitirá contrastar si el modelo es significativo o no. Es decir SCT=SCE + SCR, que fragmenta la suma de los cuadrados totales (SCT) en dos componentes: la suma de cuadrados explicada (SCE) y la suma de cuadrados de residuos (SCR). El estudio de estos componentes de SCT se conoce como análisis de varianza desde el punto de vista de la regresión.

Adicional debe considerar que para crear un modelo de regresión lineal es necesario que cumpla con los siguientes supuestos:

Supuesto 1. Linealidad: El modelo de regresión debe ser lineal en los parámetros, no es necesario que lo sea en las variables. Es decir, el modelo de regresión debe ser de la forma $Y = \beta 0 + \beta 1X + u$ (regresión lineal simple); $Y = \beta 0 + \beta 1X + \beta nXn + u$ (regresión lineal múltiple).

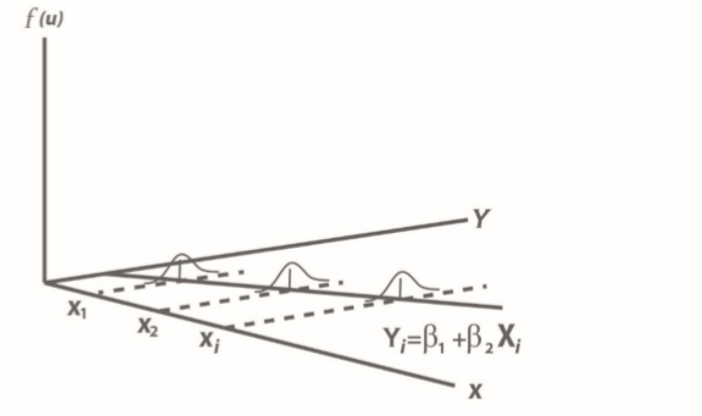
Supuesto 2. Valores fijos de X, o valores de X independientes del término error: Los valores que toma la variable independiente X pueden ser considerados como fijos en el caso de la variable independiente fija, o haber sido muestreada con la variable dependiente Y en el caso de la regresora estocástica. Así mismo que la(s) variable(s) X y el término error son independientes.

Las razones para suponer que los valores de la variable independiente son no estocásticos en ciertas condiciones, es porque en situaciones experimentales asociadas al sector agropecuario se necesite fijar los valores de la variable X. Por ejemplo, un Biólogo puede aplicar distintas cantidades de balanceado en distintas piscinas de camarón para ver el efecto en la producción. De tal manera, puede fijar una cantidad específica de balanceado aplicado en distintas piscinas con el propósito de obtener la producción promedio.

Supuesto 3. El valor de la perturbación (u) tenga una esperanza matemática igual a 0: este supuesto establece que el valor de la media de la perturbación, que depende las variables independientes dadas, es cero. Lo que mantiene el supuesto es que los factores que no se incluyen explícitamente en un modelo dado, y, por consiguiente, pertenecen a u, no afectan el valor de la media de Y. De tal manera, el supuesto 3 manifiesta que X y u no están correlacionadas (ambas ejercen influencia independientes y aditivas en Y). Si fuera lo contrario, no se podría analizar el efecto de cada una (X y u) sobre la variable dependiente, ya que, X aumentaría cuando u aumente y viceversa (correlación positiva) o X se incrementaría cuando u disminuye, y se reduciría cuando u aumenta (correlación negativa). Por tal razón, de forma general el supuesto 3 expresa que no existe error de especificación en el modelo de regresión elegido.

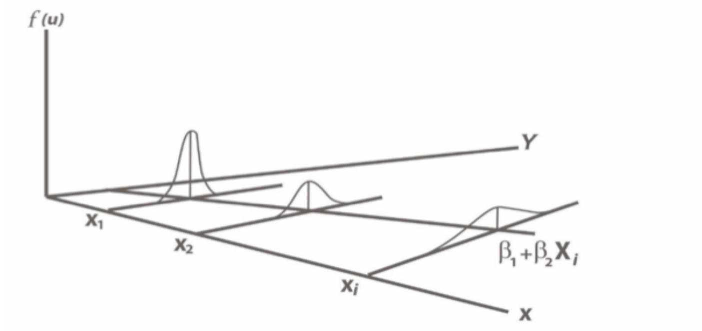
Supuesto 4. Homocedasticidad: La varianza de las perturbaciones es la misma sin importar el valor de X. Es decir, la variación alrededor de la línea de la regresión formada por elpromedio entre X y Y, es la misma para todos los valores de X (en el Gráfico 6.3 se aprecia la situación), lo contrario se conoce como heteroscedasticidad (ver Gráfico 6.4).

Gráfico 6.3. Homocedasticidad.



Fuente: Gujarati (2006)

Gráfico 6.4.Heterocedasticidad

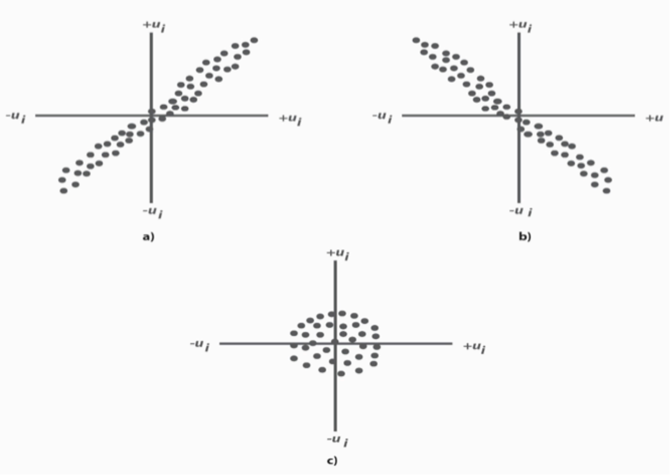


Fuente: Gujarati (2006)

Para entenderlo claramente, supongamos que Y es la producción por hectárea de un cultivo y X el fertilizante aplicado. Los gráficos 6.3 y 6.4 reflejan que aumenta la producción a medida que la cantidad de fertilizante aplicado también aumenta. Sin embargo, en el gráfico 6.3 la varianza de la producción permanece igual para todos los niveles de fertilizante aplicado, mientras que en el 6.4 aumenta la variabilidad en la producción a medida que va aumentando la dosis del fertilizante.

Supuesto 5. Incorrelación: La correlación entre las perturbaciones es igual a cero, es decir, dado X, las desviaciones de dos valores cualesquiera de Y en relación a su valor promedio muestran patrones como en el gráfico 6.5, donde se nota que las u están correlacionadas, pues a una u positiva la sigue una u positiva, y lo que requiere el supuesto 4 es que las correlaciones estén ausentes.

Gráfico 6.5. Patrones de correlación entre las perturbaciones a) correlación serial positiva; b) correlación serial negativa; c) correlación cero.



Supuesto 6. No colinealidad (para regresiones múltiples): Es decir la inexistencia de colinealidad perfecta – si una de las variables independientes tiene una relación lineal con otras variables independientes – y colinealidad parcial – si entre las variables independientes existen altas correlaciones -. Al existir colinealidad los predictores se encontrarían en combinación lineal, y la influencia de cada uno de ellos en la variable dependiente no puede distinguirse al quedar solapados unos con otros.

Supuesto 7. La naturaleza de la variables X: Los valores de la variable X no deben ser atípicos, es decir, valores muy dispersos en relación con el resto de las observaciones, con el fin de que los resultados de las regresiones estén subyugadas por tales valores atípicos. Así mismo, todos los valores de X en una muestra específica no deben ser iguales. Si los valores de X son idénticos se imposibilita la estimación de los Beta, la variación que existe tanto en los valores de X como de Y es necesario para poder utilizar la regresión como herramienta.

Informe de Resultados del análisis de regresión

Existen distintas formas de dar a conocer los resultados de un análisis de regresión; sin embargo, en este texto utilizaremos el que creemos es el más conveniente por su facilidad de análisis, como ilustración utilizamos el ejemplo del Cuadro 6.2 producción por piscina por corrida en libras y balanceado aplicado en libras:

Y = 709.6071 + 0.5885714X1

ee = (72.12439) (0.0263361)

t = (9.84) (22.35)

p = (0.001)(0.001)

F = 499.45

p = 0.001

r2 = 0.8649

Los primeros valores representan a la ecuación de la regresión; el primer conjunto de paréntesis son los errores estándar de los beta de la regresión; los valores del segundo conjunto son los estadísticos calculados (valores de t estimados), es decir, t (22.35) = 0.5885714 / 0.0263361; y los valores del tercer grupo son valores de p estimados. De tal manera, la probabilidad de obtener un valor de t igual o mayor de 22.35 es de 0.001. o prácticamente 0. En otras palabras, se rechaza la hipótesis nula que establece que el verdadero valor poblacional de cada coeficiente de regresión individual es cero.

Mientras menor sea el valor de p, más significativo es el modelo, – se considera significativo cuando es el valor de p es menor o igual a 0.05, sin embargo, es preferible dejar que el investigador decida si debe rechazar la hipótesis nula con el valor de p dado – es decir, que cuanto menor sea el valor de p, menor será la probabilidad de cometer un error si se rechaza la hipótesis nula. Por ejemplo, si un análisis del rendimiento de un cultivo, el valor de p de un estadístico de prueba resulta ser 0.13 o 13% y el investigador establecer rechazar la hipótesis en este nivel de significancia, que así sea, está asumiendo el riesgo de equivocarse el 13% de las veces si se rechaza la hipótesis nula verdadera.

Siguiendo con el ejemplo de la regresión, se presenta el valor de significancia exacto de cada valor de t estimado. Así, en relación a la hipótesis nula de que el verdadero valor poblacional de cada coeficiente individual es cero (que la cantidad de balanceado aplicado en las piscinas de camarón no produce ningún efecto en el nivel de producción), la probabilidad de que sea cierto es prácticamente cero. Por lo tanto, se rechaza la hipótesis nula y se acepta la alternativa de que el verdadero valor de la pendiente poblacional es diferente de cero (que la cantidad de balanceado aplicado en las piscinas de camarón si produce efecto en el nivel de producción). Como el signo del $\beta 1$ es positivo el efecto de la variable independiente sobre la dependiente es directa, para los casos donde el signo es negativo, se interpreta como una relación inversa.

Análisis de regresión múltiple

El análisis de regresión simple, estudiando anteriormente, suele ser no tan útil en la práctica. Es el caso del cuadro 6.2, donde se estableció que solo el nivel de balanceado X se relaciona con la producción del camarón por hectárea Y. Definitivamente rara vez es tan simple, pues, algunas otras variables afectan la producción. Para citar otro ejemplo, es altamente probable que la producción de una bananera no solo dependa de la fertilización aplicada sino también de otros insumos y recursos, tales como horas sol, agua, etc. Por tal motivo, es necesario ampliar el modelo de regresión simple y considerar modelos con más de dos variables. Al momento de adherir más variables toma el nombre de análisis de regresión múltiple, donde interviene una variable dependiente y al menos dos variables independientes.

La ecuación de la regresión - para el caso de dos variables independientes – se la representa de la siguiente forma:

$Y = \beta 0 + \beta 1X1 + \beta 2X2 + u$

El $\beta 0$ es el término del intercepto, este se conoce como el valor promedio sobre Y de todas las variables no incluidas en el modelo, aunque para mejor interpretación se conocen como el valor promedio de Y cuando las demás variables independientes se igualan a cero. Los coeficientes $\beta 1$ y $\beta 2$ son conocidos como coeficientes de regresión parcial o pendientes. El coeficiente $\beta 1$ estima el cambio en el valor de la media de Y por unidad de cambio en X1, manteniendo X2 constante. Es decir, es el efecto que tiene X1 por cada cambio unitario sobre el valor medio de Y, neto del efecto que X2 ejerza en la media Y. De la misma manera $\beta 2$ estima el cambio en el valor de la media de Y por unidad de cambio en X2, manteniendo X1 constante. Es decir, es el efecto que tiene una unidad de cambio en X2 sobre el valor medio de Y, neto del efecto que X1 ejerza en la media Y.

El modelo de regresión lineal múltiple debe ser lineal en los parámetros, no es necesario que lo sea en las variables. Es decir, el modelo de regresión debe ser de la forma $Y = \beta 0 + \beta 1X + \beta nXn + u$.

Como ilustración utilizamos el ejemplo del Cuadro6.4 que muestra el rendimiento de un cultivo en kg/ha como variable dependiente y al abono aplicado y los mm de agua recibidos como variables independientes:

Cuadro 6.4. Rendimiento por hectárea considerando el abono aplicado y los milímetros de agua recibidos.

Rendimiento/kg x Ha	Riego (mm)	Abono / kg x Ha	Rendimiento/kg x Ha	Riego (mm)	Abono / kg x Ha
200	1500	100	195	1640	100
250	1400	125	200	1540	125
300	1500	150	260	1640	150
350	2300	175	290	2440	175

Rendi-miento/kg x Ha	Riego (mm)	Abono / kg x Ha	Rendi-miento/kg x Ha	Riego (mm)	Abono / kg x Ha
400	2300	200	408	2440	200
450	2300	225	430	2440	225
500	2600	250	510	2740	250
230	2000	100	240	1900	100
270	1900	125	280	1800	125
280	2000	150	290	1900	150
330	3200	175	340	3100	175
412	3200	200	422	3100	200
420	3200	225	430	3100	225
480	3500	250	490	3400	250
217	1500	100	317	3100	317
257	1400	125	399	2900	399
267	1500	150	407	3100	407

Así, con el paquete estadístico stata obtenemos la siguiente regresión:

Y =	83.66359	+ 0.0775272X2	+ 0.3918139X3
ee=	(35.84106)	(0.0210883)	(0.1868419)
t =	(2.33)	(3.68)	(2.10)
p =	(0.026)	(0.044)	(0.001)

F = 29.74

p = 0.001

r^2 = 0.6574

Interpretemos los coeficientes de regresión: 0.0775272 es el coeficiente de regresión parcial del riego, este indica que, si se mantiene constante la influencia del abono aplicado, a medida que los mm de agua aplicado en la planta (riego) se incrementa, por ejemplo, en 1 mm en promedio, el rendimiento del cultivo aumenta en 0.0775272 kg/ha. Interpretando esto desde el punto de vista agronómico, si el productor incrementa en el riego 100 mm de agua, en promedio, el rendimiento del cultivo por ha aumenta en 7.75272 kilogramos. El coeficiente 0.3918139 establece que, si la influencia del riego se mantiene constante, el rendimiento del cultivo aumentará, en promedio 0.3918139kg/ha, si el abono aplicado subiera 1kg/ha. El valor del intercepto de 83.66359, interpretado de forma mecánica, expresa que, si los valores del riego y el abono fuesen cero, el rendimiento del cultivo en promedio sería de 83.66359 kg/Ha. El valor de R^2 es de 0.6574, lo que significa que el 65.74% de la variación en el rendimiento del cultivo por hectárea se explica mediante el abono y el nivel de riego aplicado. Sin embargo, es necesario aclarar, que a pesar de que el R^2 es una medida de bondad de ajuste del modelo, que permite determinar que tan bueno es el modelo para predecir, este desempeña un papel modesto en el análisis de regresión, ya que en un análisis de regresión el objetivo no es obtener una R^2, sino obtener estimadores confiables de los verdaderos coeficientes de regresión poblacional que permita hacer inferencia estadística sobre ellos. Por lo expresado, la pertinente es establecer correctamente las variables explicativas para la variable dependiente y por su significancia estadística.

Si se desea conocer qué pasaría con el rendimiento del cultivo si el abono y el nivel de riego se incrementaran simultáneamente, solo se tendría que multiplicar ambos coeficientes por los cambios generados y sumar los resultados. Como ejemplo, supongamos que el producto se pregunta ¿Cuál sería el efecto simultáneo en una unidad de cambio en el abono (1kg) y una de riego (1 mm) en el rendimiento del cultivo?, esto da:

$$0.0775272(1) \quad + \quad 0.3918139(1) \quad = \quad 0.4693411$$

Es decir, como resultado de este cambio simultaneo en el abono y en el riego, el rendimiento del cultivo por ha aumentará 0.4693411 kg.

Análisis de regresión con variables dicotómicas

Los modelos de regresión analizados anteriormente fueron en esencia con variables de tipo escala de razón, sin embargo, debemos considerar que los modelos de regresión también trabajan con otro tipo de variables (escala ordinal, escala de intervalo y escala nominal). En este apartado se consideran modelos que no solo tengan variables en escala de razón, sino también variables en escala nominal. Estas últimas también son denominadas variables categóricas, variables cualitativas o variables indicadoras.

En el análisis de regresión la variable dependiente frecuentemente no solo es influenciada por variables en escala de razón (por ejemplo: producción, cantidad de fertilizante aplicado) sino también por variables de naturaleza cualitativa (por ejemplo: variedad, tipo de suelo). Como ejemplo, con los demás factores constantes, se ha visto que la producción de un cultivo es superior en suelos francos que en suelos arenosos. Por tal motivo, las variables cualitativas, como el tipo de suelo y la variedad, si influyen en la variable dependiente deben ser incluidas en el modelo como parte de las independientes o explicativas.

Como la característica de estas variables es la de indicar la presencia o ausencia de un atributo como suelo franco o no, variedad Cavendish o no, son variables en escala nominal principalmente. Si la variable tiene más de dos categorías, conviene dicotomizarla mediante la introducción de variables dummy que toman valores de 0 y 1, estos atributos se pueden cuantificar, donde 1 indica la presencia del atributo y 0 su ausencia. Por ejemplo, 1 puede indicar que el suelo de tipo arcilloso y 0 de un suelo de distinto al arcilloso; 0 1 puede indicar que el banano es de la variedad Cavendish y 0 que es de otra variedad. Las variables que toman los valores de 0 y 1 se denominan variables dicotómicas. Dicotomizar

Las variables dicotómicas pueden tratarse de igual forma en los modelos de regresión como las variables cuantitativas. Incluso, un modelo de regresión puede estar conformado únicamente por variables independientes dicotómicas. Los modelos con la composición mencionada se conocen como modelos de análisis de varianza (ANOVA), descrito en el capítulo 5 como ANOVA de un factor.

Para ilustrar el análisis de varianza con variables dummy se usarán los datos que se presentan en el cuadro 6.5, éstos se refieren a la producción del cultivo de arroz (en toneladas) de 60 productores en distintas provincias del Ecuador – Las provincias es una variable cualitativa nominal con k=3 categorías, por tal motivo, para incluirla en el modelo de regresión ésta debe dicotomizarse; es decir, crear k-1=2 variables dummy (D1 y D2) –. Las provincias son: 1) El Oro (19 productores en total); 2) Los Ríos (25 productores en total), y 3) Guayas (16 productores en total).

Se desea conocer si la producción promedio de los productores de arroz difiere en relación a las 3 provincias del Ecuador. Si se toma el promedio aritmético simple de los productores de las tres provincias, obtenemos los siguientes promedios para las tres provincias: 4.57 (El Oro), 3.64 (Los ríos) y 4.18 (Guayas). Esos números difieren entre sí, pero, ¿son estadísticamente diferentes? Para compara dos o más valores medios por lo general se usa la técnica conocida como análisis de varianza, pero se logra lo mismo por medio de una regresión. En relación a lo mencionado, considere el siguiente modelo para continuar con el ejemplo:

$$Y = \beta 0 + \beta 1 D1 + \beta 2 D2 + u$$

Donde:

Y = producción (promedio) de los productores de arroz

D1 = 1 si el productor pertenece a la provincia de El Oro

= 0 para otra provincia del País

D2 = 1 si el productor pertenece a la provincia de El Guayas

= 0 para otra provincia del País

Observe que el modelo es como los otros modelos presentados en este capítulo, a diferencia que, en vez de tener variables independientes cuantitativas, se tienen solo variables dicotómicas o cualitativas.

Cuadro 6.5. Producción hipotética promedio del cultivo de arroz en las provincias de El Oro, Guayas y Los Ríos.

Productor	Producción	D2	D3	Productor	Producción	D2	D3
1	3.8	0	0	31	4.5	0	1
2	3.6	0	0	32	4.3	0	1
3	3.9	0	0	33	4	0	1
4	3.2	0	0	34	4	0	1
5	3.6	0	0	35	3.9	0	1
6	3.7	0	0	36	4.4	0	1
7	3.7	0	0	37	4.3	0	1
8	3.7	0	0	38	4.2	0	1
9	4	0	0	39	4.4	0	1
10	3.6	0	0	40	4.3	0	1
11	3.5	0	0	41	3.8	0	1
12	3.9	0	0	42	4	1	0
13	3.5	0	0	43	4.4	1	0
14	3.3	0	0	44	4.7	1	0
15	3.4	0	0	45	4.8	1	0
16	3.7	0	0	46	4.7	1	0
17	3.8	0	0	47	4.9	1	0
18	3.8	0	0	48	4.5	1	0
19	3.9	0	0	49	4.7	1	0
20	3.7	0	0	50	4.6	1	0
21	3.4	0	0	51	4.7	1	0
22	3.5	0	0	52	4.8	1	0
23	3.6	0	0	53	4.3	1	0
24	3.6	0	0	54	4.4	1	0
25	3.5	0	0	55	4.6	1	0

Productor	Producción	D2	D3	Productor	Producción	D2	D3
26	4.2	0	1	56	4.5	1	0
27	3.9	0	1	57	4.4	1	0
28	4	0	1	58	4.8	1	0
29	4.4	0	1	59	4.4	1	0
30	4.3	0	1	60	4.7	1	0

El modelo expresa, que la producción promedio de los productores de Los Ríos está dado por el intercepto β0, en la regresión múltiple; los coeficientes β1 y β2 indican la producción promedio de los productores de El Oro, así como los del Guayas, difieren respecto de la producción promedio de los productores de Los Ríos. Sin embargo, ¿cómo determinar si las diferencias encontradas son significativas estadísticamente? Con los datos del Cuadro 6.5 se obtienen los siguientes resultados.

$$
\begin{array}{llll}
Y = & 3.636+ & 0.9376842D1 & + 0.54525D2 \\
ee = & (0.0420427) & (0.0639794) & (0.0673012) \\
t = & (86.48) & (14.66) & (8.10) \\
p = & (0.001) & (0.001) & (0.001) \\
F = & 110.01 & & \\
p = & 0.001 & & \\
r^2 = & 0.7942 & &
\end{array}
$$

Como reflejan los resultados de la regresión, la producción promedio de los productores de Los Ríos es de 3.64 toneladas, el de los productores de la provincia de El Oro es mayor al de los Ríos por cerca de 0.9377 toneladas, así como los de Guayaquil, es mayor por alrededor de 0.545 toneladas. Las producciones medias reales de las últimas provincias se obtienen sumando estos promedios diferenciales a la producción promedio de los productores de los Ríos. Al realizar esto, tendremos que la producción promedio de las últimas dos provincias son próximos a 4.57 y 4.18 respectivamente.

Pero, para saber si estas producciones promedio son estadísticamente diferentes de la producción promedio de Los Ríos, que es la provincia con la que se compara, hay que averiguar si cada coeficiente β1 y β2 son significativos. Como se observa en la regresión, el coeficiente estimado de la pendiente para la provincia de El Oro es estadísticamente significativo, porque el valor de p es menor a 0.05; también el del Guayas es estadísticamente significativo, ya que el valor de p también es menor a 0.05. Por tal manera, se concluye, que estadísticamente, la producción de arroz de los productores de las provincias de Los Ríos, El Oro y Guayas son distintas entre sí.

Debe considerar que las variables dicotómicas solo establecen las diferencias, si es que existen, pero no explica las razones por las que se generan. A lo mejor, el nivel de fertilizante, la nutrición, el tipo de suelo, los mm de precipitación puedan ejercen influencia en los resultados.

Es necesario aclarar que: 1) la categoría a la que no se la asigna variable dicotómica es conocida como categoría omitida, de referencia, de comparación, de control u categoría base; 2) el valor del intercepto β0 representa el valor medio de la variable de referencia, en el ejemplo trabajado dicha categoría es la correspondiente a la provincia de Los Ríos; 3) los coeficientes asociados a las variables dicotómicas se conocen como coeficientes de intercepto diferencial, indicando la medida en que la categoría que toma el valor de 1 difiere del coeficiente del intercepto correspondiente a la categoría de referencia o comparación y 4) queda a criterio del analista o investigador establecer la categoría de comparación.

Modelo de regresión y ANCOVA

Los modelos de regresión que contienen como variables independientes una mezcla de variables cualitativas y cuantitativas se denominan modelos de análisis de covarianza (ANCOVA). Para ilustrar el modelo utilizaremos los datos del Cuadro 6.6.

Cuadro 6.6.Rendimiento promedio de un cultivo considerando el riego, la cantidad de abono aplicado y el tipo de suelo.

Rendi-miento/ kg x Ha	Riego (mm)	Abono /kg x Ha	D3	D4	Rendi-miento/ kg x Ha	Riego (mm)	Abono / kg x Ha	D3	D4
200	1500	100	0	0	195	1640	100	0	0
250	1400	125	0	0	200	1540	125	1	0
300	1500	150	0	0	260	1640	150	1	0
350	2300	175	0	0	290	2440	175	0	0
400	2300	200	0	0	408	2440	200	1	0
450	2300	225	1	0	430	2440	225	1	0
500	2600	250	1	0	510	2740	250	0	1
230	2000	100	0	0	240	1900	100	0	0
270	1900	125	0	0	280	1800	125	0	0
280	2000	150	0	1	290	1900	150	0	0
330	3200	175	0	0	340	3100	175	0	0
412	3200	200	1	0	422	3100	200	1	0
420	3200	225	1	0	430	3100	225	0	1
480	3500	250	1	0	490	3400	250	1	0
217	1500	100	0	0	317	3100	317	0	1
257	1400	125	0	0	399	2900	399	0	1
267	1500	150	0	0	407	3100	407	0	1

Con los datos en mención (Cuadro 6.6) se desarrolla el siguiente modelo:

Y = β0 + β1X1 + β2X2 + β3D3 + β4D4

Donde:

Y = rendimiento (promedio) de un cultivo

D3 = 1 si el suelo es franco

= 0 para otro tipo de suelo

D4 = 1 si el suelo es franco arcilloso

= 0 para otro tipo de suelo

X1 = mm de riego

X2 = Abono aplicado

Tenga presente que los datos proporcionados en el Cuadro 6.6 considera al suelo franco arenoso como la categoría de comparación. También, note que, además de las dos variables independientes o regresoras cualitativas, se tienen dos variables cuantitativas, X1 y X2, que en los modelos de ANCOVA se denominan covariante. Ya que, estos modelos representan una extensión de los modelos ANOVA en el sentido de que establecen un método para controlar estadísticamente los efectos de las variables independientes cuantitativas (nombradas variables concominantes o covariables).

De los datos del Cuadro 6.6, los resultados obtenidos del modelo (4) son los siguientes:

Y = 92.73969 + 0.60828X1 + 0.470243X2 + 53.1269D3 - 10.97473D4

ee = (35.6568) (0.203315) (0.223505) (24.84383) (37.64764)

t = (2.60) (2.99) (2.10) (2.14) (-0.29)

p = (0.014) (0.006) (0.044) (0.041) (0.773)

F = 19.48

p = 0.001

r2 = 0.7287

Como los resultados reflejan, cuando todo lo demás se mantiene constante: mientras el riego aumenta 1mm, en promedio, el rendimiento por hectárea se incrementa alrededor de 0.60 kilogramos. Así mismo, mientras el abono aplicado aumenta en 1 kg/ha, el rendimiento se incrementa más o menos en 0.47 kg/ha. Si controlamos los mm de riego y el abono, se observa que el coeficiente del intercepto diferencial es significativo para el suelo franco, pero no lo es para el suelo franco arcilloso.

Referencia bibliográfica

Berndt, Ernst R. (1991). The Practice of Econometrics, Classic and Contemporary, Addison-Wesley.

Cameron, A. Colin y Pravin K. Trivedi. (2005). Microeconomics: Methods and Applications, Cambridge University Press, Nueva York.

Depool, R., & Monasterio, D. (2013). *Probabilidad y estadística. Aplicaciones a la ingeniería.* Barquisimeto, Venezuela: Universidad Nacional Experimental Politécnica Antonio José de Sucre (Unexpo). Obtenido de http://www.bqto.unexpo.edu.ve/avisos/PROBABILIDADYESTADISTICA(2-7-13).pdf

Garriga, A. J., Lubin, P., Merino, J. M., Padilla, M., Recio, P., & Suárez, J. C. (2010). *Introducción al análisis de datos.* Madrid, España: Universidad Nacional de Educación a Distancia (UNED).

Goldberger, Arthur S. (1998). Introductory Econometrics, Harvard University Press.

Goldberger, A. S. (1968). Topics in Regression Analysis, Macmillan, Nueva York.

Greene, William H. (2000). Econometric Analysis, 4a. ed., Prentice Hall, Englewood Cliffs.

Gujarati, Damodar N. (2006). Essentials of Econometrics, 3a. ed., McGraw-Hill, Nueva York.

Gujarati y Porter. (2010). *Econometría,*5a. ed., McGraw-Hill, Nueva York.

Hayashi, Fumio. (2000). Econometrics, Princeton University Press, Princeton, N. J.

Hernández, R., Fernández, C., & Baptista, P. (2014). *Metodología de la investigación* (Sexta ed.). D.F. México: McGraw-Hill.

IBM Corp. (2016). *SPSS Statistics versión 24.0.0.0* de prueba para Windows. Barcelona: International Business Machines Corp.

Johnston, J. (1984). Econometric Methods, 3a. ed., McGraw-Hill, Nueva York.

Kennedy, Peter. (1998). A Guide to Econometrics, 4a. ed., MIT Press, Cambridge, Mass.

Klein, Lawrence R. (1962). An Introduction to Econometrics, Prentice Hall, Englewood Cliffs, N.J.

Lind, D. A., Marchal, W. G., & Wathen, S. A. (2015). *Estadística aplicada a los negocios y la economía*. México, D. F.: McGraw-Hill Internacional.

Patterson, Kerry. (2000). An Introduction to Applied Econometrics, A Time Series Approach, St. Martin's Press, Nueva York.

Peracchi, Franco. (2001). Econometrics, John Wiley & Sons, Nueva York.

Sáez, A. J. (2012). *Apuntes de estadística para ingenieros*. España: Dpto de Estadística e Investigación Operativa. Universidad de Jaén. Obtenido de http://www4.ujaen.es/~ajsaez/recursos/EstadisticaIngenieros.pdf

Salcedo, A. (2013). *Estadística en el investigación*. Caracas: ISBN: 978-980-00-2743-1.

Walters, A. A. (1968). An Introduction to Econometrics, Macmillan, Londres.

Wooldridge, Jeffrey M.. (2000). Introductory Econometrics, 3a. ed., South-Western College Publishing.

Análisis de Datos Agropecuarios
Edición digital 2017- 2018.
www.utmachala.edu.ec

Redes

Redes es la materialización del diálogo académico y propositivo entre investigadores de la UTMACH y de otras universidades iberoamericanas, que busca ofrecer respuestas glocalizadas a los requerimientos sociales y científicos. Los diversos textos de esta colección, tienen un espíritu crítico, constructivo y colaborativo. Ellos plasman alternativas novedosas para resignificar la pertinencia de nuestra investigación. Desde las ciencias experimentales hasta las artes y humanidades, Redes sintetiza policromías conceptuales que nos recuerdan, de forma empeñosa, la complejidad de los objetos construidos y la creatividad de sus autores para tratar temas de acalorada actualidad y de demanda creciente; por ello, cada interrogante y respuesta que se encierra en estas líneas, forman una trama que, sin lugar a dudas, inervará su sistema cognitivo, convirtiéndolo en un nodo de esta urdimbre de saberes.

UNIVERSIDAD TÉCNICA DE MACHALA

UNIVERSIDADE DA CORUÑA

UNIVERSIDAD TÉCNICA DE MACHALA
Editorial UTMACH
Km. 5 1/2 Vía Machala Pasaje
www.investigacion.utmachala.edu.ec / www.utmachala.edu.ec

ISBN: 978-9942-24-120-7
9 789942 241207