Regresión lineal simple y múltiple: aplicación en la predicción de variables naturales relacionadas con el crecimiento microalgal

Simple and multiple regression: application in the prediction of natural variables related to microalgae growing process

Arys Carrasquilla-Batista[1], Alfonso Chacón-Rodríguez[2], Kattia Núñez-Montero[3], Olman Gómez-Espinoza[4], Johnny Valverde[5], Maritza Guerrero-Barrantes[6]

Fecha de recepción: 31 de marzo de 2016
Fecha de aprobación: 21 de mayo de 2016

Carrasquilla-Batista, A; Chacón-Rodríguez, A; Núñez-Montero, K; Gómez-Espinoza, O; Valverde, J; Guerrero-Barrantes M. Regresión lineal simple y múltiple: aplicación en la predicción de variables naturales relacionadas con el crecimiento microalgal . *Tecnología en Marcha*. Encuentro de Investigación y Extensión 2016. Pág 33-45.

DOI: 10.18845/tm.v29i8.2983

1 Ingeniera Electrónica y Máster en Computación. Ingeniería Mecatrónica. Instituto Tecnológico de Costa Rica. Cartago, Costa Rica. Correo electrónico: acarrasquilla@itcr.ac.cr
2 Ingeniero Electrónico, Máster en Literatura Inglesa y Doctor en Ingeniería. Ingeniería Electrónica. Instituto Tecnológico de Costa Rica. Cartago, Costa Rica. Correo electrónico: alchacon@itcr.ac.cr
3 Ingeniera en Biotecnología. Escuela de Biología. Instituto Tecnológico de Costa Rica. Costa Rica. Correo electrónico: knunez@itcr.ac.cr
4 Ingeniero en Biotecnología. Escuela de Biología. Instituto Tecnológico de Costa Rica. Costa Rica. Correo electrónico: oespinoza@itcr.ac.cr
5 Máster en Nutrición y Salud, Doctor en Ciencias Naturales. Escuela de Química. Instituto Tecnológico de Costa Rica. Costa Rica. Correo electrónico: jovalverde@itcr.ac.cr
6 Bióloga, Máster en Ciencias con énfasis en Ecología. Escuela de Biología. Instituto Tecnológico de Costa Rica. Costa Rica. Correo electrónico: mguerrero@itcr.ac.cr

Palabras clave

Microalgas; regresión simple; regresión múltiple; variables predictivas.

Resumen

En la actualidad, existe una creciente necesidad en diferentes campos de investigación y producción, y en la industria de la agricultura de precisión, de almacenar y procesar datos provenientes de múltiples sensores. Muchas veces estos dispositivos se encuentran ubicados en lugares remotos. El modo usual de recolección de datos implica el uso de un equipo para cada variable de interés, lo cual dificulta y encarece la integración y el procesamiento conjunto. Se considera entonces la posibilidad de incorporar la temática del Internet de las Cosas, con el fin de aprovechar las capacidades computacionales y de procesamiento en la nube para que los investigadores puedan disponer de la información que les permita tomar decisiones oportunas.

La presente investigación se centra en los modelos de regresión lineal, simple y múltiple, con el fin de establecer las bases para modelar la relación entre las variables de temperatura, luz, pH y oxígeno disuelto, y de esta manera poder conocer las factores que afectan l crecimiento del cultivo de microalgas en futuras investigaciones.

Keywords

Microalgae; simple regression; multiple regression; predictive variables.

Abstract

Nowadays, there is a growing need in various research fields and in the industry of precision agriculture to record and process data from multiple sensors, sensors sometimes located in remote areas, miles apart from each other. The usual approach to sensor data recording implies measurement of each variable in separate equipment, making it difficult and expensive to integrate and process jointed data. The possibility of incorporating the theme of Internet of Things (I.oT.) in research is being analyzed to take advantage of the ubiquitous computing capabilities available today.

This article is about simple regression and multiple regression models, which offer the bases to explore the relationship between variables associated to microalgae kinetic growth: temperature, light, pH and dissolved oxygen. Recorded data will provide new approaches to present works; in this way, researchers will perform various data analysis online.

Introducción

En la actualidad, el modo más común de medición de variables ambientales en agricultura e investigación es por medio de múltiples dispositivos electrónicos que trabajan en forma no coordinada; cada dispositivo es utilizado para adquirir diferentes datos de interés, tales como humedad, temperatura, dióxido de carbono, luminosidad y radiación solar. Este acercamiento es muchas veces lento y tedioso porque se requiere usualmente un equipo por cada variable que va a ser medida, y además, se deben instalar y manipular diferentes herramientas de software para leer y procesar la información suministrada, la cual generalmente es integrada manualmente en una base de datos, por un operario que recopila los datos de interés en el campo o en el laboratorio. Por otra parte, estos dispositivos solo guardan los valores máximos y

mínimos de las variables en estudio. En la última década, mucho se ha avanzado en lo referente a la integración de dispositivos y sensores mediante el uso de redes inalámbricas, pero la mayoría de los acercamientos a la solución del problema se han visto seriamente limitados por el corto alcance suministrado y la baja capacidad de procesamiento de los nodos (Yick, Mukherjee & Ghosal, 2008). Esto ha limitado su aplicación a sitios donde los requerimientos de transferencia de datos no son tan altos, y donde existen restricciones de potencia que justifican el uso de plataformas con baja capacidad de procesamiento.

Internet de las Cosas (I. de C.)

La evolución de las tecnologías inalámbricas 3G y 4G (asociado con la caída de los costos de estos sistemas de comunicación) y el surgimiento del Internet de las Cosas (IoT., Internet of Things, por sus siglas en inglés) (Evans, 2011), también llamado Internet de Todas las Cosas (Internet of Everything, I. E.), ofrece una solución viable y práctica para la integración de múltiples sensores. Esta integración permite explotar las grandes capacidades de procesamiento y almacenamiento disponibles por medio de Internet, donde se obtiene ventaja de técnicas complejas de procesamiento, usualmente conocidas como "fusión de datos de múltiples sensores", para proveer a los investigadores no solo con un amplio rango de series masivas de datos provenientes de múltiples sitios remotos, sino que además se tiene la posibilidad de aplicar procesos predictivos y de estimación. Las limitaciones más comunes en el uso de este tipo de datos en redes inalámbricas convencionales –donde normalmente se tienen altos requerimientos de almacenamiento y de procesamiento computacional para el desarrollo de algoritmos probabilísticos o estadísticos– son superadas.

Específicamente en esta investigación, se explora sobre los métodos estadísticos de regresión lineal simple y múltiple con el fin de evaluar la posibilidad de su aplicación para la predicción del crecimiento de la microalga *Chlorella sp*. en cultivo. Las variables relacionadas con la dinámica del crecimiento de las microalgas, que se integrarían en un modelo de regresión múltiple pueden ser temperatura, pH, intensidad de luz y oxígeno disuelto.

El presente documento ha sido organizado de la siguiente manera: en la sección introductoria se explica la importancia de la aplicación de los modelos de regresión lineal y múltiple en el estudio del crecimiento de la microalga *Chlorella sp*.; en la segunda sección se describe la regresión lineal, la estructura e hipótesis del modelo y los procedimientos de evaluación; posteriormente, se desarrolla una breve introducción a la terminología utilizada en la regresión múltiple, y se describe la hipótesis y prueba de hipótesis; finalmente, se exponen las conclusiones del estudio realizado.

Microalga *Chlorella sp.*

El calentamiento global es el resultado de la gran cantidad de emisiones de CO_2 y se ha convertido en un tema obligado en todo lo relacionado con el medio ambiente.

Las microalgas son un grupo de organismos unicelulares de crecimiento rápido; una de sus características más importantes es la habilidad para convertir de manera muy eficiente el CO_2 en biomasa. Este proceso de biomitigación representa varias ventajas, tales como una mayor tasa de crecimiento y una mayor fijación de CO_2 en comparación con la obtenida de los bosques, la agricultura y las plantas acuáticas (Borowitzka, 1999). Varios estudios relacionados con la fijación de CO_2 por medio del cultivo microalgal son descritos en la literatura (Sydney, Sturm, Carvalho, Thomaz-Soccol & Larroche, 2010; Wang, Li, Wu & Lan, 2008), así como las ventajas de controlar y mantener las microalgas bajo condiciones ópticas de crecimiento en fotobioreactores, estudiadas por Stewart y Hessami, (2005).

Ya el 23 de junio de 1981 (Fox, 1996), la *Chlorella sp.* fue certificada GRAS (Generally Recognized As Safe, Generalmente Reconocida como Segura) y desde entonces puede ser usada como alimento sin riesgos para la salud.

La *Chlorella sp.* contiene

- 53% de proteína,
- 23% de carbohidratos,
- 9% de grasa,
- 5% minerales y
- 2% clorofila. (Henrikson, 1994)

La clorofila es la que le permite a la *Chlorella sp.* crecer rápidamente, en forma similar a las plantas. El metabolismo principal es realizado por medio de la fotosíntesis (Vonshak, 1997), esto significa que la fuente principal de energía es la luz del sol.

La *Chlorella sp.* también es rica en

- vitamina B, específicamente B12, vital en la formación y regeneración de las células de la sangre, y
- hierro, que puede ser utilizado en el tratamiento y prevención de anemia.

De acuerdo con Sung, Lee, Park y Choi (1999), la microalga *Chlorella sp.* tiene una tasa casi constante de crecimiento a valores de pH mayores a 4,2, y por lo tanto puede desarrollarse fácilmente en estanques y lagos. Otra ventaja de la *Chlorella sp.* comparada con otras microalgas es la alta tolerancia a las temperaturas y a las concentraciones de CO_2; puede mantener su crecimiento a 42 °C y 40% de CO_2 (Sakai, Sakamoto, Kishimoto, Chihara & Karube, 1995).

En la figura 1 se presenta un diagrama del proceso metabólico de la microalga, en el cual se esquematizan los procesos de biosíntesis de varios productos derivados de las microalgas. A pesar de que el cloroplasto puede funcionar como una fábrica de proteínas e hidrógeno (azul), el núcleo juega un papel fundamental en el control metabólico (rojo punteado). Ambas organelas contienen genomas individuales, lo que la provee de la posibilidad de una incorporación transgénica (línea azul y roja).

El objetivo de la presente investigación es ofrecer una base teórica en el tema de regresión simple y regresión múltiple, con el fin de incorporar en futuras investigaciones algoritmos de predicción del crecimiento de la microalga a partir de la medición de variables ambientales en los medios de cultivo. Los datos serán recolectados por un dispositivo electrónico, el cual tendrá la capacidad de conectarse a Internet; de esta forma los investigadores podrán bajar la base de datos requerida, con el fin de analizar y tomar decisiones y, en un futuro, permitir la integración de los datos en un modelo de predicción del crecimiento, que sea procesado en la nube de Internet.

Regresión lineal simple

Modelo de regresión

Los factores que intervienen en un experimento pueden ser cuantitativos o cualitativos. Un factor cuantitativo es aquel cuyos niveles pueden asociarse con puntos en una escala numérica, como la temperatura, la humedad relativa, la conductividad eléctrica, la presión o el tiempo. Los factores cualitativos son aquellos cuyos niveles no pueden ordenarse por magnitud.

Los operarios, los proveedores, los turnos de trabajo y las máquinas son factores cualitativos, ya que no existe ninguna razón para ordenarlos bajo algún criterio numérico particular (Montgomery, 2008). En el caso especial del cultivo de microalgas, se tienen algunos factores cualitativos como son operario, día de toma de muestras y lixiviado. Las variables relacionadas con el crecimiento son cuantitativas: temperatura, pH, oxígeno disuelto, dióxido de carbono, intensidad de luz y otras. Además, la variable de interés o salida (*Y*) se mide en células por mililitro (cel/ml), por lo tanto también es cuantitativa.

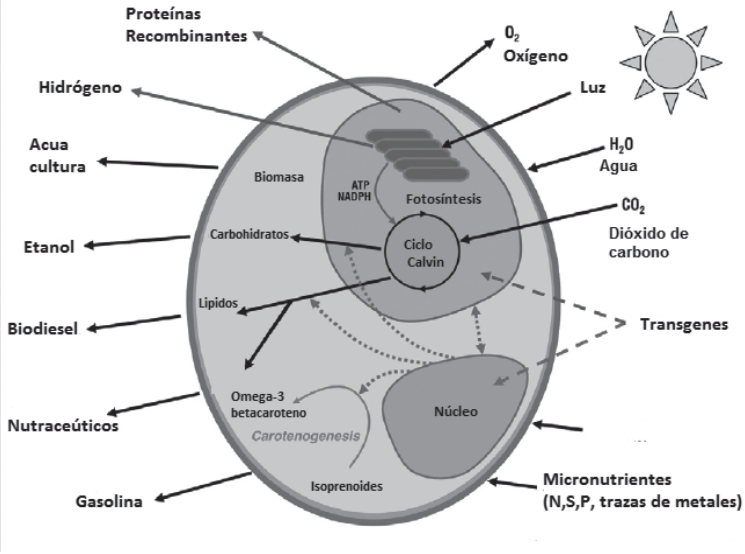

Figura 1. Representación diagramática de un proceso microalgal por medio de sustancias bioactivas. La luz, el agua y el CO_2 se relacionan con el crecimiento de las microalgas, durante el cual se produce oxígeno. Rosenberg et al (2008).

Se conoce como regresión simple el cálculo de la ecuación correspondiente a la línea que mejor describe la relación entre la respuesta y la variable que la explica. Dicha ecuación representa la línea que mejor se ajusta a los puntos en un gráfico de dispersión (ver figura 2).

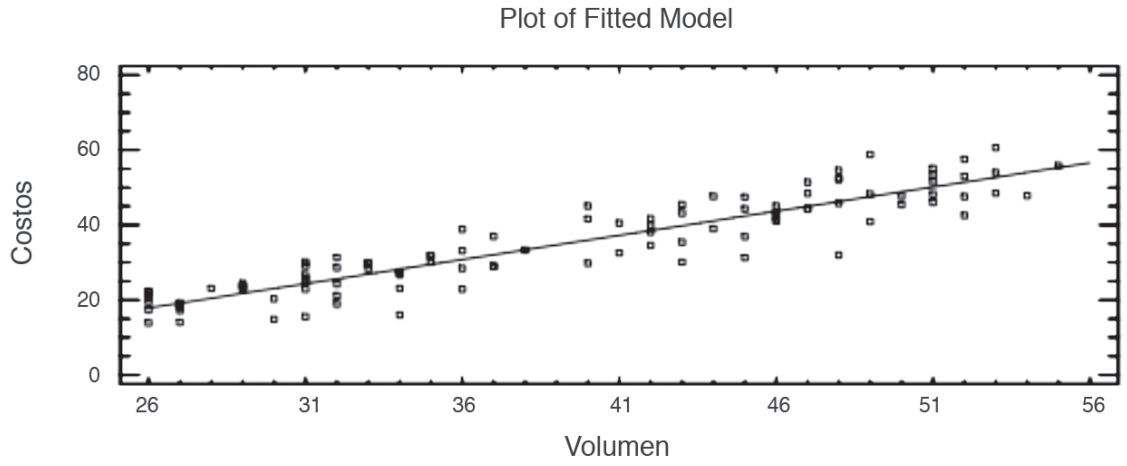

Figura 2. Representación de una ecuación de regresión simple para la relación entre costos y volumen.

Se debe recordar que la respuesta es la variable que se mide. Si se asocia con el crecimiento de microalgas será células/ml y la misma se establece como dependiente. El factor que influye en esta respuesta puede ser una cualquiera de las variables cuantitativas mencionadas, las cuales son denominadas independientes.

En la regresión simple se tiene una única variable predictora. Algunas veces se tiene interés en dos o más variables regresoras o predictoras En esos casos, se debe recurrir al uso de regresión múltiple. A partir de la regresión lineal es posible hacer predicciones sobre la respuesta con base en valores de la variable predictora.

La ecuación para una línea recta es $y = b_0 + b_1 x$ donde

- y es la variable respuesta;
- x es la variable predictora
- b_0 es la intersección, determina el valor de *y* cuando *x* es cero, y
- b_1 es la pendiente, determina la cantidad en la que cambia *y* cuando *x* se incrementa en una unidad.

Las distancias entre los puntos y la línea de regresión se llaman residuos. Ellos representan la porción de la respuesta que no es explicada por la ecuación de regresión; es decir que la diferencia entre el valor observado y el valor aproximado es el residuo.

En cualquier análisis de regresión se observará que algunos puntos están más cerca de la línea y otros mucho más lejos de ella. Entre más cerca se encuentren los puntos a la línea, mejor será el ajuste entre la línea de regresión y el dato. Los residuos permiten verificar la ecuación con el fin de comprobar cuan bien se ajusta la línea a los datos.

En resumen, se puede establecer que la pendiente de una ecuación de regresión indica el efecto de la variable predictora sobre la variable respuesta. En la regresión se utilizan los llamados mínimos cuadrados, también conocidos como mínimos cuadrados de regresión, los cuales determinan la línea que minimiza la suma de las distancias verticales cuadradas, desde los puntos hacia la recta (ver figura 3).

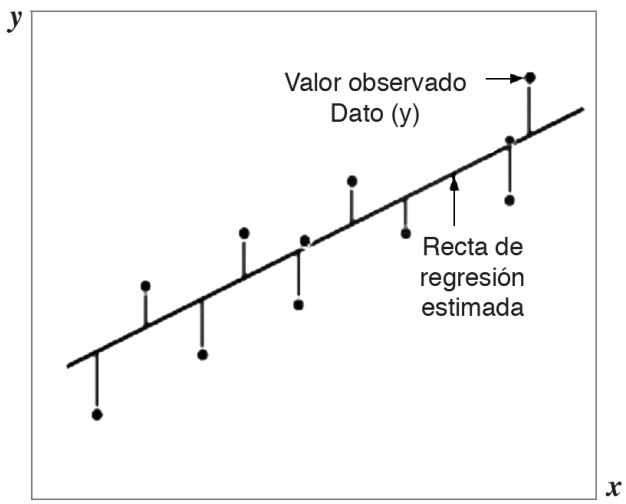

Figura 3. Representación de los mínimos cuadrados.

Hipótesis del modelo de regresión líneal simple

Linealidad: La relación existente entre *X* e *Y* es lineal, f(x) = β0 + β1x

Homogeneidad: El valor promedio del error es cero, E[ui] = 0

Homocedasticidad: La varianza de los errores es constante, Var(ui) = σ 2

Independencia: Las observaciones son independientes, E[ui uj] = 0

Normalidad: Los errores siguen una distribución normal, ui ~ N(0, σ)

Prueba de hipótesis y R^2

¿Cómo podemos comprobar si la línea de regresión es significativa? Para esto, se establecen las siguientes relaciones:

- Hipótesis nula H_0, $\beta_1 = 0$
- Hipótesis alternativa H_1, $\beta_1 \neq 0$
- P es la probabilidad, está dada entre 0 y 1.

Se comprueba si el valor verdadero de la pendiente, β_1, es igual a 0. Si la línea es totalmente horizontal, la pendiente es cero y no existe ninguna relación lineal entre las variables. Pero, si la línea no es horizontal, la pendiente no es cero, y puede ser que exista una relación entre las variables.

Si el valor de $P > \alpha$, no se rechaza H_0 y si $P \leq \alpha$, se rechaza H_0. Esto implica que se rechaza la hipótesis nula de que la pendiente es igual a cero. Un valor común para α es 0,05.

Después que se determina que la relación entre variables es estadísticamente significativa, se puede establecer si la respuesta se explica por la variable de regresión, es decir, si la variable predictora explica la mayor parte de las variaciones en la respuesta. En este caso, los puntos en el gráfico de dispersión están ubicados cerca de la línea y los residuos son pequeños.

Para medir cuan alejado está un resultado de la respuesta o cuánta variabilidad existe en la respuesta, de acuerdo a la variable explicatoria, se utiliza R^2. R^2 es un valor entre cero y uno que usualmente es expresado como un porcentaje para hacerlo más fácil de interpretar. Como porcentaje R^2 tiene su valor entre 0% y 100%.

Para determinar si una variable predictora está relacionada con la variable respuesta, se puede llevar a cabo una prueba de hipótesis de la pendiente. La hipótesis nula establece que la pendiente es cero y la hipótesis alternativa especifica que la pendiente no es igual a cero.

Si el valor P de la prueba es menor que α, se rechaza la hipótesis nula y se concluye que la variable predictora está significativamente relacionada con la variable respuesta.

El valor de R^2 establece numéricamente si la línea de regresión se alinea o encaja con los puntos e indica cuánta variación existe en la variable respuesta y si esta se explica por la variable de regresión.

En la regresión se deben cumplir los tres rubros siguientes:

1. Los errores son aleatorios e independientes (*Minitab residual plots*).
2. Los errores tienen una distribución normal (*Minitab normal probability plot*).
3. Los errores tienen varianza constante a lo largo de todos los valores de x (*Minitab plot the residuals vs fitted values*).

Tipos de factores en el modelo de regresión

Tomando en consideración el diseño inicial y el análisis de un experimento, los factores involucrados se tratan de manera idéntica. El investigador está interesado en determinar diferencias, en caso de haberlas, entre los niveles de los factores

En algunos modelos puede encontrarse que la diferencia en las respuestas asociadas a los niveles de un factor no es la misma que en las respuestas asociadas a otro u otros factores. Cuando esto ocurre, existe una interacción entre los factores; si además algunos factores del diseño son cuantitativos, entonces una representación en un modelo de regresión del experimento factorial de un factor, podría escribirse como sigue (Gutiérrez & De la Vara, 2012):

Si la recta de regresión es

$$y = \beta_0 + \beta_1 x$$

Cada valor y_i observado para un x_1 puede considerarse como el valor esperado de *Y* dado x_1 más un error (expresado como ε_i):

$$y_i = \beta_0 + \beta_1 x_1 + \varepsilon_i$$

Donde los ε_i se suponen errores aleatorios con distribución normal, media cero y varianza σ^2; β_0 y β_1 son constantes desconocidas (parámetros del modelo de regresión). Al enfoque general para ajustar modelos empíricos se le llama análisis de regresión, el cual toma en consideración la respuesta a un evento determinado. El análisis de regresión es una técnica estadística para investigar la relación funcional entre dos o más variables, mediante ajustes en un modelo matemático. La regresión lineal simple, como ya se dijo, utiliza una sola variable de regresión y el caso más sencillo es el modelo de línea recta. Supóngase que se tiene un conjunto de pares de observaciones (x_1, y_i), se busca encontrar una recta que describa de la mejor manera cada uno de esos pares observados.

Se considera que la variable *X* es la variable independiente o regresiva y se mide sin error, mientras que *Y* es la variable respuesta para cada valor específico xi de *X*; además, *Y* es una variable aleatoria con alguna función de densidad para cada nivel de *X*. En las figuras 4 y 5, se muestra esta relación.

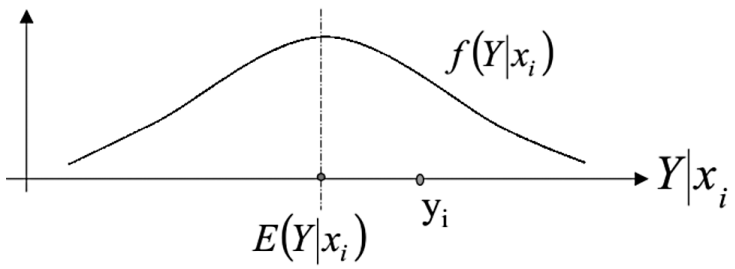

Figura 4. Densidad de la variable respuesta Y.

Puede resultar útil para el investigador ajustar una curva de respuesta a los niveles de un factor cuantitativo para contar así con una ecuación que relacione la respuesta con el factor. Esta ecuación podría utilizarse para hacer interpolaciones, es decir, para predecir la respuesta en niveles intermedios de los factores, respecto de los que se utilizaron realmente en el experimento.

Cuando al menos dos de los factores son cuantitativos, puede ajustarse una superficie de respuesta para predecir *Y* con varias combinaciones de los factores del diseño. En general, se usan métodos de regresión lineal para ajustar estos modelos a los datos experimentales.

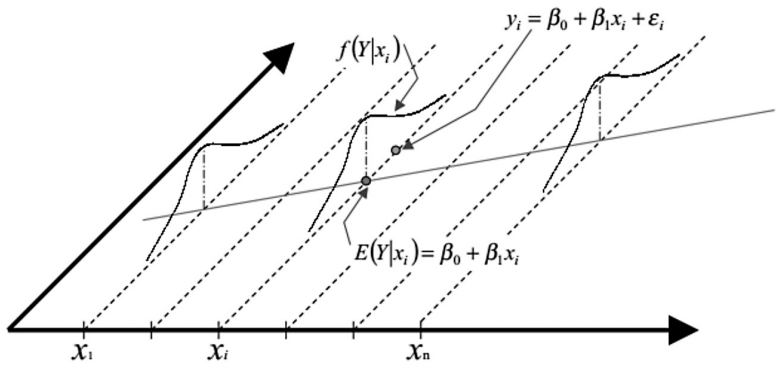

Figura 5. Valor esperado de Y dado x_1.

Ajuste de curvas y superficies de respuesta

En un diseño factorial 2^k, es posible expresar los resultados del experimento en términos de un modelo de regresión. Puesto que 2^k es tan solo un diseño factorial, puede utilizarse un modelo de los efectos o de las medias, pero el enfoque del modelo de regresión es mucho más natural e intuitivo.

En el sistema de los diseños 3^k, cuando los factores son cuantitativos, es común denotar los niveles bajo, intermedio y alto con -1, 0 y +1, respectivamente. Con esto se facilita el ajuste de un modelo de regresión que relaciona la respuesta con los niveles de los factores.

Regresión múltiple

Modelos de regresión múltiple

El modelo de regresión múltiple es la extensión del modelo de regresión simple a k variables explicativas. La estructura del modelo de regresión múltiple es la siguiente:

$$y = f(x_1, \ldots., x_k) + \epsilon$$

Donde

- y es la variable explicada, dependiente o respuesta.
- $x_1, \ldots., x_k$ son las variables explicativas, regresores o variables independientes.
- $y = f(x_1, \ldots., x_k)$ es la parte determinista del modelo.
- $\in$ representa el error aleatorio. Contiene el efecto sobre y de todas las variables distintas de $x_1, \ldots., x_k$.

El modelo de regresión lineal múltiple tiene la forma:

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \epsilon$$

El modelo de regresión lineal múltiple se utiliza cuando

1. La variable dependiente, Y , depende linealmente de cada una de las variables explicativas, $x_1, \ldots, x_k$.
2. Un regresor no basta para explicar suficientemente la variabilidad de y.

Modelo de regresión múltiple con dos regresores

En el caso particular en que haya dos regresores, k = 2 , el modelo tendría la forma:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$$

Gráficamente, el modelo de regresión lineal con dos regresores supone calcular la ecuación de un plano que describe la relación de y con x_1 y x_2 . En la figura 6, se muestra la relación entre y, x_1 y x_2.

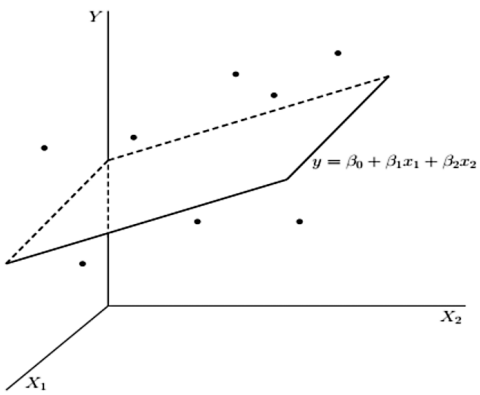

Figura 6. Modelo de regresión lineal con dos regresores.

La estimación por mínimos cuadrados de los parámetros del modelo consiste en calcular la ecuación del plano que haga mínimo el valor de $\sum e_i^2$ con

$$e_i = y_i - \hat{y}_i$$

En la figura 7 se muestra el plano con la estimación por mínimos cuadrados.

Hipótesis del modelo de regresión múltiple

Generalizando, al ajustar un modelo de regresión lineal múltiple, se supondrá que se verifican las siguientes hipótesis:

1. Fijados los valores $x_{1i}, \ldots, x_{ki}$ de las variables $X_1,\ldots,X_k$, se tiene que

$$y_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ki} + e_i$$

2. Cada error $e_i \approx N(0, \sigma^2)$.
3. Cualquier par de errores e_i y e_j son independientes.
4. Las variables explicativas son, algebraicamente, linealmente independientes.
5. El número de datos es mayor o igual que k + 2.

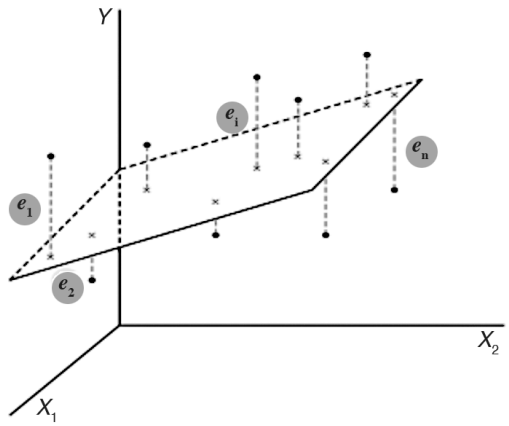

Figura 7. Plano para la estimación por mínimos cuadrados en un modelo con dos regresores (x_1 y x_2).

Se llega a las siguientes observaciones:

1. Las tres primeras hipótesis del modelo se justifican igual que en regresión simple.
2. La condición de independencia lineal algebraica de los regresores tiene por objeto ajustar la dimensión del problema, ya que si no se cumpliese se pueden eliminar regresores del modelo.
3. El número de datos debe ser mayor o igual que k + 2 para poder estimar todos los parámetros del modelo.

Parámetros del modelo de regresión múltiple

1. El parámetro βˆi, en regresión múltiple, representa el efecto del aumento de una unidad del regresor x_{ki} sobre la respuesta, *Y*, cuando el resto de los regresores permanecen constantes.2. Si los regresores no se interrelacionan, ρij = 0, para todo i, j, los estimadores de los coeficientes de regresión en el modelo múltiple y en los distintos modelos simples coinciden.

Se puede demostrar que

1. βˆi sigue una distribución normal, para todo i = 0, . . . , k.
2. Para todo βˆi, con i = 0, 1, . . . , k, se cumple que E (βˆi) = βi. Es decir βˆi es un estimador centrado de βi, para todo i.
3. La matriz de varianzas y covarianzas de βˆ 0, . . . , βˆk viene dada por la expresión:

   COV (βˆ) = σ 2(X 0X).

Prueba de hipótesis en el modelo de regresión múltiple

En la prueba de hipótesis en el modelo de regresión múltiple, se establecen comparaciones de hipótesis nula y alternativa:

- Hipótesis nula H_0: $\beta_1 = \beta_2 = \cdots = \beta_k = 0$
- Hipótesis alternativa H_1: Existe algún $\beta_i \; con \; i = 1, \ldots, k \; tal \; que \; \beta_i \neq 0$

La aceptación de la hipótesis nula de la Prueba de Hipótesis, representada por la letra F.

$$H_0 : \quad \beta_1 = \beta_2 = \cdots = \beta_k = 0,$$

Puede ser debida a

- Independencia de la variable explicada (respuesta) frente a todas las variables predictoras /regresoras.
- Dependencia no lineal de la variable respuesta respecto de alguna variable predictora.

El rechazo de la hipótesis nula significa que la variable explicada depende linealmente de alguno de los regresores.

- Para determinar cuál o cuáles de las variables independientes/predictivas explican significativamente a la variable dependiente, es necesario tomar en consideración las comparaciones individuales. En el cuadro 1 se detallan los casos, el resultado de las pruebas de hipótesis y las comparaciones individuales.

Cuadro 1. Prueba de hipótesis y comparaciones individuales en regresión múltiple

Caso	Prueba de hipótesis	Comparaciones individuales	Interpretación
1	Significativo	Todos significativos	Todas las variables predictoras influyen significativamente en la variable respuesta.
2	Significativo	Alguno significativo	Los predictores no significativos deben ser eliminados del modelo o transformados si se intuye relación de dependencia no lineal.
3	Significativo	Ninguno significativo	Problema de multicolinealidad: existe fuerte correlación entre variables explicativas del modelo.
4	No significativo	Todos significativos	Existen casos particulares de multicolinealidad
5	No significativo	Alguno significativo	Existen casos particulares de multicolinealidad
6	No significativo	Ninguno significativo	No se detecta relación de dependencia lineal entre la variable explicada y los regresores.

El tratamiento de la multicolinealidad consiste en eliminar regresores del modelo con alta correlación, y con esto disminuye el número de parámetros que hay que determinar.

Diagnóstico y validación del modelo de regresión múltiple

Al igual que en el modelo de regresión simple, antes de utilizar el modelo de regresión múltiple, es necesario verificar las hipótesis básicas del modelo y esto se realiza por medio del análisis de los residuos.

Se pueden considerar las siguientes particularidades:

1. La normalidad del error.
2. Las hipótesis de linealidad, homocedasticidad (la varianza del error es constante) e independencia.
3. La conveniencia de introducir una nueva variable.

Finalizado el proceso de definición y validación del modelo de regresión múltiple, se puede utilizar para hacer predicciones. Es posible considerar

- $\hat{y}\,(x_{1i}\,,\ldots,\,x_{ki})$ para predecir el valor de $\mathrm{E}(\mathrm{Y}|\mathrm{X}_1 = x_{1i},\ldots,x_k = x_{ki})$

- $\hat{y}\,(x_{1i}\ ,\ldots,\ x_{ki})$ para predecir el valor de un elemento o individuo de la variable $(\mathrm{Y}|\mathrm{X}_1 = x_{1i},\ldots,x_k = x_{ki})$.

Conclusiones

Los modelos de regresión simple y múltiple presentan las características ideales para el tratamiento de variables cuantitativas que responden según las variables predictoras oregresoras dentro del fenómeno estudiado.

En el proceso de cultivo de microalgas, existen variables que afectan la dinámica del crecimiento, principalmente la temperatura, el pH y la intensidad de luz. Las microalgas se alimentan de CO_2 y lo convierten en biomasa, en ese proceso liberan oxígeno.

Se sugiere el uso de la regresión simple para modelar la relación entre cada variable predictiva, de forma independiente, y la variable respuesta (crecimiento celular de la microalga) y posteriormente plantear un modelado de regresión múltiple.

El Internet de las Cosas puede ser aplicado en el estudio estadístico del crecimiento de la microalga *Chlorella sp.*, en conjunto con algoritmos de predicción de alta complejidad que requieran procesamiento estadístico, como lo es el modelo de regresión múltiple. El hecho de poder acceder a la nube de Internet facilita el manejo de los datos para los investigadores y especialistas que requieran información del estado del cultivo.

Bibliografía

Borowitzka, M.A. (1999). Commercial production of microalgae: ponds, tanks, tubes and fermenters. *Journal of Biotechnology,* 70, 313– 321.

Evans, D., (2011). Internet of Things - How the Next Evolution of the Internet Is Changing Everything. White paper Cisco Internet Business Solutions Group (IBSG).

Fox, R. (1996) Spirulina production and potential, France: Edisud, ISBN 2-84744-883-X.

Gutiérrez, H., De la Vara, R., (2012) Análisis y diseño de experimentos, 3ra edición, McGraw –Hill, México.

Henrikson, R. (1994) Microalga Spirulina: Superalimento del futuro, Barcelona: Ediciones S.A. Urano, ISBN: 84-7953-047-2.

Montgomery, D., (2008) Diseño y análisis de experimentos, 2da edición, Limusa-Wiley, México

Rosenberg, J., Oyler, G., Wilkinson, J. and Betenbaugh, M. (2008) A green light for engineered algae: Redirecting metabolism to fuel a biotechnology revolution. Current Opinion in Biotechnology 19: 430–436.

Sakai, N., Sakamoto, Y., Kishimoto, N., Chihara, M. and Karube, I. (1995) *Chlorella* strains from hot springs tolerant to high temperature and high CO_2. Energy Conversion and Management 16: 693–696.

Stewart, C. and M.A. Hessami (2005). A study of methods of carbon dioxide capture and sequestration—the sustainability of a photosynthetic bioreactor approach. *Energy Conversion and Management,* 46, 403–420.

Sung, K., Lee, J., Park, S. and Choi, M. (1999) CO_2 fixation by *Chlorella* sp KR-1 and its cultural characteristics. Bioresource Technology 68: 269–73.

Sydney, E., Sturm, W., de Carvalho, J., Thomaz-Soccol, V., Larroche, C. (2010) Potential carbon dioxide fixation by industrially important microalgae Published by Elsevier Ltd. doi:10.1016/j.biortech.2010.02.088s.

Vonshak A. (1997), Spirulina platensis (Arthrospira): Physiology, Cell Biology and Biotechnology. Taylor & Francis, London. ISBN 0-7484-0674-3.

Wang, B., Li, Y., Wu, N., Lan, C.Q. (2008). CO_2 mitigation using microalgae. Applications in .Microbiology. Biotechnology. 79, 707–718.

Yick, J., Mukherjee, B., Ghosal, D., (2008). Wireless sensor network survey. Department of Computer Science, University of California, Davis, USA. Science Direct. Available online 14 April 2008 : doi:10.1016/j.comnet.2008.04.002