

MODELO DE REGRESIÓN LOGÍSTICA



*Eulogio Hurtado Dianderas**
E-mail: eulogio1952@hotmail.com

RESUMEN

La asignatura de Operaciones y Logística constituye un elemento fundamental en la formación académica de un Magíster en Administración ya que proporciona al estudiante las herramientas necesarias para la toma de decisiones en las áreas de abastecimiento, transportes y comunicaciones en toda empresa.

El presente artículo trata sobre el desarrollo y aplicación de un modelo de regresión logística que aporta los conocimientos para una mejor administración de los recursos materiales con la debida calidad, cantidad costos y oportunidad requeridos.

Palabras clave: Modelo, regresión, operaciones, logística.

ABSTRACT

The Logistic subject of Operations and constitutes a fundamental element in the academic formation of Magíster in Administration since it provides to the student the necessary tools for the decision making in the areas of supplyings, I transport and communications in all company.

The present articulate present deals with on the development and application about a model logistic regression that contributes the knowledge for one better administration of the material resources with the due quality, amount required costs and opportunity.

Key words: Model, regression, operations, logistic.

* Doctor en Ingeniería, Magíster en Investigación, MBA ESAN. Director Académico del Doctorado en Ingeniería Industrial de la Universidad Nacional Mayor de San Marcos. Profesor Auxiliar de la Facultad de Ciencias Administrativas.

Las variables ficticias se usan para incorporar en el modelo de regresión lineal posibles variaciones estructurales ocurridas durante el periodo muestral, o efectos socioeconómicos que pudieran distinguir el comportamiento de unos individuos de otros. Existen, sin embargo, muchos casos en donde la variable endógena, a pesar de que es cuantificable, no toma valores en un intervalo infinito sino sólo en un número finito de ellos. A veces esta variable ni siquiera es cuantificable como cuando se pretende caracterizar a una persona por el tipo de transporte que utiliza: 1 si usa bus 2, si usa coche particular, etc. Otros ejemplos son: decisión de tomar parte en el mercado de trabajo, el nivel de estudios de un individuo, el voto a depositar, clase de vivienda a ocupar, si se concede o no un crédito. Por ello se trata de buscar modelos que mejor se ajusten estos casos.

Analicemos el caso en que la variable que se pretende explicar (Y_i), toma dos valores 0 ó 1 como variables explicativas.

MODELO LOGIT

Este modelo surge cuando se usa como F la función de distribución

$$F(z) = \frac{e^z}{1+e^z} \quad -\infty < z < +\infty$$

cuya función de densidad correspondiente es

$$f(z) = \frac{1}{(1+e^z)^2}$$

Usando lo anterior se tiene:

$$P_i = F(x_i' \beta) = \frac{e^{x_i' \beta}}{1+e^{x_i' \beta}}$$

La inversa de la función F se calcula con

$$F^{-1}(w) = \ln \frac{w}{1-w}$$

Bajo los supuestos anteriores:

$$\ln \frac{P_i}{1-P_i} = x_i' \beta$$

Cuando existen observaciones repetidas se puede usar la aproximación lineal de Taylor para obtener:

$$\ln \frac{P_i}{1-P_i} \cong x_i' \beta + \frac{\mu_i}{P_i(1-P_i)}$$

Esta expresión sugiere que se deben calcular las frecuencias muestrales y estimar la regresión. Para ganar eficiencia deberán utilizarse los mínimos cuadrados generalizados. Cuando no existen observaciones repetidas se usa el método de máxima verosimilitud.

El signo de los coeficientes tiene un significado importante. Si son positivos, la variable aumenta la probabilidad del suceso que estamos estudiando y por lo tanto se dice que dicho factor **es de riesgo**. Si el signo es negativo, ocurre lo contrario y en tal caso se llama **factor de protección**.

HIPÓTESIS EN REGRESIÓN LOGÍSTICA SIMPLE

La hipótesis nula que se desea contrastar es si la variable independiente no influye significativamente sobre la variable dependiente. La hipótesis alternativa es: la variable independiente influye significativamente sobre la variable dependiente.

Entre los procedimientos para contrastar las hipótesis anteriores están:

1. **Evaluación estadística del coeficiente**, en donde se usa como estadístico de prueba a

$$\frac{\hat{\beta}_1 - \beta}{ee(\hat{\beta}_1)}$$

el cual corresponde a una variable aleatoria que tiene distribución t con $n - 2$ grados de libertad. (En el caso de la regresión logística múltiple el número de grados de libertad es $n - K - 1$ grados de libertad, en donde K es el número de variables independientes).

2. **Evaluación estadística de WALD**, en donde se usa como estadístico de prueba a

$$Wald = \frac{(\hat{\beta}_1)^2}{(ee(\hat{\beta}_1))^2}$$

Este estadístico se distribuye como una ji-cuadrado con un grado de libertad.

3. **Evaluación del incremento del estadístico** $-2\ln(L_0)$

En este estadístico L_0 indica la razón de verosimilitud. Este valor oscila entre 0 y 1. El procedimiento comienza conociendo el valor de este estadístico cuya distribución es Ji cuadrado.

Para el modelo que sólo contiene la constante el número de grados de libertad es igual al número de casos menos 1; a continuación se

calcula el valor cuando se ha incluido en el modelo a la variable independiente. En este caso el número de grados de libertad es igual al número de casos menos el número de variables independientes menos 1. Se rechaza la hipótesis nula si el incremento es diferente de cero significativamente.

4. Evaluación del estadístico $-2Ln(L_0)$

El valor de L_0 es un número menor que 1. En un modelo perfecto este valor es igual a 1, y por lo tanto su logaritmo es igual a 0. La distribución de este estadístico es Ji cuadrado con $n - 2$ grados de libertad.

La hipótesis a probar es $L_0 = 1$ vs. $L_0 \neq 1$

En este caso no poder rechazar la hipótesis significa que el modelo es significativo.

5. Evaluación del estadístico z^2 (Goddness of fit).

$$z^2 = \sum_{i=1}^n \frac{e_i^2}{p_i(1-p_i)}$$

Donde los valores e_i son los residuos y los p_i son las probabilidades estimadas para el i -ésimo caso. Este estadístico se distribuye como una Ji cuadrado con $n - 2$ grados de libertad.

En este procedimiento, no poder rechazar la hipótesis equivale a decir que el modelo es significativo.

COEFICIENTE DE CORRELACION PARCIAL

La contribución parcial de cada variable es difícil de calcular puesto que su influencia depende del resto de las variables.

Un estadístico que puede usarse es:

$$R = \sqrt{\frac{Wald - 2K}{-2Ln(L_0)}}$$

El valor de R oscila entre 1 y 0. Un valor próximo a 0 indica que la contribución al modelo es pequeña.

RESULTADOS

La aplicación de todas y cada una de las ecuaciones que se derivan del modelo de regresión logística postulado, permite focalizar y puntualizar bajo un enfoque hermenéutico los aspectos más relevantes y aplicativos en la logística moderna.

Por tanto esto constituye una herramienta indispensable para la administración de los recursos en planeamiento y control de abastecimientos, compras, administración de almacenes, gestión de transporte empresarial y de la infraestructura integral de comunicaciones.

BIBLIOGRAFÍA

- CARRANZA, OCTAVIO. (2004). *Logística. Mejores Prácticas en Latinoamérica*. Segunda edición. México, Thompson.
- DURAND, ALFONSO (2006). *Logística y el Comercio Electrónico*. Segunda edición. Editorial Mc Graw Hill.
- LONG, DOUGLAS. *Logística Internacional*. Primera edición. Limusa-Noriega Editores.