

MODELOS DE REGRESSÃO LOGÍSTICA ORDINAL USANDO SOFTWARE LIVRE: UMA APLICAÇÃO À SATISFAÇÃO DO TURISTA EM ACAPULCO, GUERRERO

Data de aceite: 02/05/2023

Octaviano Juárez Romero

docente de la Facultad de Matemáticas-
Universidad Autónoma de Guerrero
ORCID ID 0000-0002-2890-289

Santiago Marquina Benítez

docente de la Facultad de Economía
(Acapulco)-Universidad Autónoma de
Guerrero
ORCID ID 0000-0002-4358-8823

RESUMEN: La modelación matemática en las distintas áreas del conocimiento ha cobrado relevancia favorecida por el incremento del software disponible para su implementación. En particular, la modelación estadística también se ha beneficiado por el incremento de software libre como el paquete estadístico R. El presente trabajo tiene como objetivo la modelación estadística utilizando la metodología de regresión logística ordinal mediante los distintos comandos del paquete estadístico R. Se aplica esta metodología a la actividad turística en Acapulco, Guerrero, en particular a la satisfacción del turista en los servicios de hospedaje y alimentos y bebidas. La base de datos utilizada comprende 465 registros limpios completamente de datos faltantes.

Los modelos seleccionados para cada uno de los servicios satisfacen el supuesto del modelo, satisfacen las pruebas estadísticas de significancia individual y global del modelo. Para la selección de los modelos se utilizan el criterio de la deviance, el AIC y la prueba de bondad de ajuste. Finalmente, los modelos seleccionados como los adecuados pronostican más del 85% de los valores observados de las variables de interés.

PALABRAS CLAVE: Regresión logística ordinal, paquete estadístico R, turismo en Acapulco.

MODELOS DE REGRESIÓN LOGÍSTICA ORDINAL UTILIZANDO SOFTWARE LIVRE: UNA APLICACIÓN A LA SATISFACCIÓN TURÍSTICA EN ACAPULCO, GUERRERO

ABSTRACT: Mathematical modeling in the different areas of knowledge has gained relevance, favored by the increase in software available for its implementation. Statistical modeling has also benefited from the increase in free software such as the statistical package R. The objective of this work is statistical modeling using the ordinal logistic regression methodology through the

different commands of the statistical package R. This methodology to the tourist activity in Acapulco, Guerrero, to the satisfaction of the tourist in the lodging services and food and beverages. The database used comprises 465 clean records completely of missing data. The models selected for each of the services satisfy the assumption of the model, satisfy the statistical tests of individual and global significance of the model. For the selection of the models, the deviance criterion, the AIC and the goodness-of-fit test are used. Finally, the models selected as the adequate ones predict more than 85% of the observed values of the variables of interest.

KEYWORDS: Ordinal logistic regression, statistical package R, tourism in Acapulco

INTRODUCCIÓN

La modelación estadística cobra mayor relevancia en tiempos recientes en las distintas disciplinas como la medicina, la economía, las finanzas, la epidemiología, la ingeniería, entre otras. Los modelos estadísticos lineales comprenden la modelación de la variable de interés tanto cualitativa como cuantitativa. En este último caso, cuando la variable es dicotómica se utiliza la regresión logística, cuando la variable toma más de dos categorías se utiliza la regresión logística multinomial y cuando la variable es ordinal se modela mediante regresión logística ordinal. En el presente trabajo la metodología aplicada a una base de datos de la actividad turística en Acapulco, Guerrero es la regresión logística ordinal.

Un elemento que ha posibilitado que la modelación estadística amplíe su aplicabilidad es la existencia de software adecuado a cada metodología. En este punto se está avanzando en el software que está disponible para todo usuario, como es el paquete estadístico R, el cual utilizamos en este trabajo.

La actividad turística en México y de manera particular en el estado de Guerrero tiene un peso importante en las economías. En el caso del estado de Guerrero, su economía depende en más del 60% de la actividad en sus tres centros turísticos: Acapulco, Ixtapa-Zihuatanejo y Taxco. Dada la importancia de esta actividad durante distintas temporadas vacacionales las autoridades municipales de Acapulco levantaron encuestas para conocer el nivel de satisfacción de los visitantes al Puerto de Acapulco. En el trabajo se utiliza una encuesta con 465 registro limpios de datos faltantes, considerando la satisfacción del turista en dos servicios prestados: el servicio de hospedaje y el servicio de alimentos y bebidas.

DESCRIPCIÓN DE LA METODOLOGÍA

El objetivo del presente trabajo es construir un modelo que explique de la mejor manera la satisfacción del turista en los servicios recibido en su visita al Puerto de Acapulco. En general la modelación pretende obtener el mejor modelo que se ajuste a los datos, aquel que sea el más simple y que explique mejor el fenómeno en estudio. En las últimas décadas el crecimiento que ha tenido la computación así como el software estadístico ha posibilitado

mejorar el trabajo de modelación.

La metodología que se utiliza para la modelación es la regresión logística ordinal. La regresión lineal utiliza como variable de interés una variable cuantitativa. En el caso de una variable cualitativa, la regresión logística utiliza una variable de respuesta binaria y la regresión logística multinomial utiliza una variable con más de dos categorías. Cuando la variable de interés tiene más de dos categorías pero tiene un orden la metodología utilizada es la regresión logística ordinal. Esta metodología es la que se utiliza y describe en el presente trabajo.

La regresión logística ordinal (RLO), parte de considerar que la variable de interés Y asume k categorías y_1, y_2, \dots, y_k y se explica a través de p variables independientes: X_1, X_2, \dots, X_p . Existen en la literatura distintas propuestas de establecer la relación entre la variable de interés Y y las variables independientes como: modelo de la categoría adyacente, modelo de razón continua (Hosmer y Lemeshow, 2000), modelo de riesgo proporcional, el modelo de ventajas proporcionales (McCullagh y Nelder, 1990), entre otros. Este último se utiliza en el presente trabajo.

El modelo de regresión logística ordinal con función de enlace logit se define en (1).

$$\text{logit}\{P[Y \leq y_j | X]\} = \log \left\{ \frac{P[Y \leq y_j | X]}{P[Y > y_j | X]} \right\} = \alpha_j + \beta_1 X_1 + \beta_2 X_2, \dots, \beta_p X_p \quad \text{para } j = 1, 2, \dots, k-1 \quad (1)$$

donde los parámetros del modelo son: $\alpha_1, \alpha_2, \dots, \alpha_{k-1}, \beta_1, \beta_2, \dots, \beta_p$ (Agresti, 2002).

En este modelo se establecen $k-1$ ecuaciones, una para cada categoría y_j donde los coeficientes β_i de las variables independientes, se suponen, constantes. Esta es la diferencia con modelo de regresión logística multinomial, donde se establecen también las $k-1$ ecuaciones pero se permite que los coeficientes β_i de las variables independientes pueden tomar otros valores.

El modelo de regresión logística ordinal (1) es también llamado modelo de las razones proporcionales (McCullagh y Nelder, 1990), ya que las ventajas del evento $Y \leq y_i$ en $X = x_1$ son $\exp[\beta^t(x_1 - x_2)]$ -veces sobre las ventajas en $X=x_2$, esto es dado en (2)

$$\text{logit}\{P[Y \leq y_i | x_1]\} - \text{logit}\{P[Y \leq y_i | x_2]\} = \log \left\{ \frac{P[Y \leq y_i | x_1] / P[Y > y_i | x_1]}{P[Y \leq y_i | x_2] / P[Y > y_i | x_2]} \right\} = \beta^t(x_1 - x_2) \quad (2)$$

Como se observa el lado derecho en (2) es independiente de las categorías de la variable de respuesta.

Estimación de los parámetros

Para efectuar la estimación de los parámetros del modelo especificado en (1) mediante el método de máxima verosimilitud (Hosmer y Lemeshow, 2000), se define primero la probabilidad condicional para la categoría y_j en (3).

$$\phi_j(x) = P[Y = y_j | x] = \frac{e^{g(x)}}{1 + e^{g(x)}} \text{ para } j = 1, 2, \dots, k \quad (3)$$

donde $g(x) = \beta_0 + \beta_1 x_1, \dots, \beta_p x_p$

En la función $\phi_j(x)$ se integran todos los parámetros del modelo $\beta = (\beta_0, \beta_1, \dots, \beta_p)$ definido en (1) incluyendo los interceptos α_j . Se integra $\phi_j(x)$ en (1)

$$\log \left\{ \frac{P[Y \leq y_j | X]}{P[Y > y_j | X]} \right\} = \log \left[\frac{\phi_1(x) + \phi_2(x) + \dots + \phi_j(x)}{\phi_{j+1}(x) + \phi_{j+2}(x) + \dots + \phi_k(x)} \right] = \alpha_j + \beta^t x \text{ para } j = 1, 2, \dots, k - 1$$

Para la estimación de los parámetros se supone una muestra de tamaño n , con las observaciones (y_i, x_i) para $i=1, 2, \dots, n$, además, se define una variable de respuesta z $(k+1)$ dimensional fijando $z_j=1$ si $Y=y_j$ en otro caso $z_j=0$ para $j=1, 2, \dots, k$. Con estos elementos se establece la función de verosimilitud en (4).

$$L(\beta) = \prod_{i=1}^n [\phi_1(x_i)^{z_{1i}} \times \phi_2(x_i)^{z_{2i}} \times \dots \times \phi_k(x_i)^{z_{ki}}] \quad (4)$$

Con lo cual se obtiene el logaritmo de la verosimilitud en (5).

$$L(\beta) = \sum_{i=1}^n [z_{1i} \log \phi_1(x_i) + z_{2i} \log \phi_2(x_i) + \dots + z_{ki} \log \phi_k(x_i)] \quad (5)$$

Para obtener los estimadores máximo-verosímiles de los parámetros del modelo, se procede de la forma usual derivando parcialmente (5) con respecto a cada uno de los parámetros formando $(k+p)$ ecuaciones igualadas a cero. La solución del sistema de ecuaciones será los valores de los estimadores $\hat{\beta}$.

Evaluación estadística de los modelos

Supuesto de las rectas paralelas

Para efectuar la prueba de las rectas paralelas, esto es, para verificar el supuesto del modelo se comparan los resultados del ajuste del modelo de regresión logística ordinal con el modelo de regresión logística multinomial. Esta es la prueba de la razón de verosimilitud, esto es,

$$G = -2(\text{LogVerosimilitud}(M1) - \text{LogVerosimilitud}(M2))$$

donde $M1$ es el modelo de regresión logística ordinal y $M2$ es el modelo de regresión logística multinomial. El estadístico G sigue una distribución Chi-cuadrada con los grados de libertad igual a la diferencia entre el número de parámetros estimados en $M2$ y $M1$. La hipótesis nula asegura que los modelos son iguales, es decir, que los coeficientes estimados de las variables independientes son iguales en los distintos modelos de las categorías, esta hipótesis se rechaza para un valor suficientemente grande de G .

Significancia de los coeficientes de las variables independientes.

Para llevar a cabo la prueba de hipótesis de la significancia de los coeficientes β_i , se utiliza el estadístico de Wald, dado en (6).

$$Z_{wala} = \frac{\hat{\beta}_i}{se(\hat{\beta}_i)} \text{ para } i = 1, 2, \dots, p \quad (6)$$

Donde $\hat{\beta}_i$ es el estimador del coeficiente de la variable X_i y $se(\hat{\beta}_i)$ su error estándar. El estadístico Z_{Wald} sigue una distribución normal estándar, bajo la hipótesis nula. Cuando se rechaza la hipótesis nula, se puede asegurar que la variable X_i tiene un efecto sobre las probabilidades de ocurrencia de las categorías de la variable de interés (Heredia et al. 2012).

Bondad de ajuste del modelo

Para efectuar la prueba de bondad de ajuste el paquete estadístico R proporciona el valor del estadístico denominado deviance residual. En la hipótesis nula se establece que el modelo es el adecuado, esto es, que los valores predichos bajo el modelo ajustado y los valores observados de la variable de interés son similares. Bajo la hipótesis nula la deviance residual se comporta asintóticamente como una Chi-cuadrada con los grados de libertad igual al número de casos menos el número de parámetros estimados. Se rechaza la hipótesis nula si la Chi-cuadrada calculada es mayor que la Chi-cuadrada teórica a un nivel de significancia α (Hosmer y Lemeshow, 2000; McCullagh y Nelder, 1990).

Selección de modelos

Criterio de AIC (Akaike Information Criterion) propuesto por Akaike (1973) para seleccionar modelos se define en términos de un criterio de información adecuado, esto es, un mecanismo que da a cada modelo candidato un puntaje.

Para un modelo M se define el AIC como

$$AIC(M) = -2 \log(\text{Máxima verosimilitud}) + 2 \dim(M)$$

Donde $\dim(M)$ es la dimensión del vector de parámetros asociado al modelo M. Por lo cual, los modelos más complejos reciben penalizaciones más altas.

Como criterio de selección de modelos, de la lista de posibles modelos-candidatos se selecciona aquel modelo con menor AIC.

RESULTADOS

Descripción de la muestra

La muestra fue levantada en las vacaciones de fin de año 2015, fueron logradas 700 entrevistas pero debido a la no respuesta se redujo a 465 entrevistas completas. Este es el tamaño de muestra que se considera.

El software que se utilizó en los cálculos de la modelación es el paquete estadístico R. Este es un software libre que se está utilizando cada vez más en la Estadística, en el presente trabajo, para realizar los cálculos se requirió de además de las instrucciones básicas, los paquete denominados “MASS” y “nnet”, el primero para realizar el ajuste del modelo de regresión logística ordinal y el segundo para ajustar a los mismos datos un modelo multinomial. Esto último se utilizó para verificar el supuesto del modelo.

La modelación que vamos a realizar es sobre los servicios que se ofrecen a los turistas en hospedaje y en alimentos y bebidas.

Servicios de hospedaje.

Para el servicios de hospedaje la variables de interés es la respuesta a la pregunta: *¿Cómo califica usted la experiencia general con el establecimiento de hospedaje?* (STPI211) y las posibles respuestas de los turistas son: Excelente, Bueno, Malo y Muy malo, cuyas frecuencias observadas en la muestra fueron de 102, 334, 27 y 2, respectivamente. Las variables independientes que se consideran son

- 1.- *¿Cómo califica usted el servicio ofrecido por el personal del establecimiento de hospedaje?* (STPI212).
- 2.- *¿Cómo califica usted la relación calidad/precio del establecimiento de hospedaje?* (STPI213)
- 3.- *¿Cómo califica usted la variedad de la oferta de hospedaje?* (STPI214)
- 4.- *¿Cómo califica usted la información obtenida del establecimiento de hospedaje?* (STPI215)

Una primera etapa en la modelación, es la verificación de que existe una relación entre la variable de respuesta y las variables independientes, esto se realiza con la prueba de independencia. Esta prueba se efectúa mediante la construcción de una tabla de contingencia y se calcula el estadístico Chi-cuadrada, esto se realizó con el comando de R:

```
Indep1 <- chisq.test(Var211, Var212).
```

Para la variable dependiente (STPI211) y las variables independientes STPI212, STPI213, STPI214 y STPI215. Los valores de la Chi-cuadrada dados por las pruebas fueron: 809.49, 392.811, 327.57 y 321.54, respectivamente. Con estos valores se rechaza ampliamente la hipótesis de independencia a favor de la dependencia, ya que el valor teórico de la Chi-cuadrada con un nivel de significancia de 0.05 y 9 grados de libertad es 16.91898.

Estimación de los parámetros de los modelos especificados.

Para seleccionar el modelo adecuado a la variable de interés, se ajustaron cuatro modelos. El primer modelo ajustado es la variable dependiente (STPI211) contra la variable independiente STPI212, utilizando la siguiente instrucción de R:

```
modelo1 <- polr(Var211 ~ Var212, method="logistic")
```

El modelo 2, corresponde a este modelo más la variable independiente STPI213; el siguiente modelo M3 se agrega la variable independiente STPI214 y el último, es el modelo M3 más la variable STPI215. De esta forma se tienen los cuatro modelos reportados en la Tabla 1, con sus correspondientes estadísticos.

Modelo	Bondad ajuste Deviance (p-valor)	AIC	Pronósticos (%)
M1: STPI211 ~ STPI212	419.761 (0.9160158)	427.761	85.5913
M2: STPI211 ~ STPI212 + STPI213	365.1964 (0.9995932)	375.1964	86.6666
M3: STPI211 ~ STPI212 + STPI213 + STPI214	330.9514 (0.9999989)	342.9514	86.0215
M4: STPI211 ~ STPI212 + Var213 + STPI214 + STPI215	328.0693 (0.9999999)	342.0693	86.4516

Tabla 1. Modelos ajustados y sus correspondientes estadísticos.

En los cuatro modelos ajustados resultaron significativos los coeficientes de las variables independientes excepto en el último modelo, la variable independiente STPI215 tiene un p-valor de 0.09039, el cual es mayor a 0.05.

Por otro lado, la prueba de bondad de ajuste que se realiza con el estadístico deviance residual y su p-valor que son reportan en la segunda columna de la Tabla 1. Considerando estos valores de las pruebas estadísticas se puede asegurar que los modelos ajustados son los adecuados ya que sus correspondientes p-valores son mayores a 0.05.

Selección de modelos

Para seleccionar el modelo más adecuado que explica el comportamiento de los datos, se consideran los elementos antes mencionado, el estadístico AIC y su capacidad de predicción.

El modelo número cuatro considerando el estadístico AIC, sería el mejor ya que su valor es el menor de los cuatro modelos de la Tabla 1, sin embargo, tiene un coeficiente que no es estadísticamente significativo y en consecuencia, un parámetro más que el modelo tres. Aun cuando el modelo M4 el porcentaje de valores predichos ligeramente superior que el modelo tres, se selecciona el modelo tres como el más adecuado. Por tanto los parámetros estimados del modelo seleccionado son:

$$\log \left[\frac{P(Y \leq y_i | X)}{P(Y > y_i | X)} \right] = \alpha_i - 2.804(STPI212) - 1.553(STPI213) - 1.784(STPI214); \quad i = 1, 2, 3.$$

Los valores ajustados de los intercepto son: $\alpha_1=8.2309$, $\alpha_2=14.3982$ y $\alpha_3=21.2788$

Prueba de las rectas paralelas

Como se menciona arriba, la prueba se realiza comparando el modelo tres que se seleccionó con el modelo de regresión logística multinomial. Este último modelo considera para cada categoría de la variable dependiente las tres variables independientes, para realizar el ajuste, esto es, prácticamente ajusta tres modelos, como el siguiente:

$$STPI211 \sim \alpha_i + \beta_{1i}STPI212 + \beta_{2i}STPI213 + \beta_{3i}STPI214 \quad \text{para } i = 1, 2, 3$$

Los resultados del ajuste realizado en R se proporcionan en la Tabla 2.

Categoría	Intercepto	STPI212	STPI213	STPI214
1	-23.97663	20.81906	4.013394	-4.4355889
Error Estándar	4.143403	1.638050	4.083768	4.404914
2	-35.64502	22.18864	6.093980	-2.6667798
Error Estándar	4.230329	1.543282	4.066799	4.424951
3	-57.14251	25.38718	7.399188	-0.9882062
Error Estándar	4.499920	1.541837	4.063944	4.444243

Tabla 2. Parámetros estimados y sus correspondientes errores estándar del modelo multinomial especificado arriba.

Para efectuar la prueba se calcula el estadístico mediante comandos de R

$$G = -2\{\log\text{Verosimilitud}(\text{modelo3}) - \log\text{Verosimilitud}(\text{modelo multinomial})\}$$

Dando un valor de $G=9.452622$. El valor teórico de $X^2_{(0.95,6)}=12.59159$, por lo cual no hay evidencia estadística para rechazar la hipótesis de que los modelos comparados sean distintos, esto es, que los coeficientes estimados en el modelo de regresión logística ordinal sean distintos para los interceptos.

Interpretación del modelo seleccionado.

Según el modelo seleccionado, la que tiene mayor impacto sobre las categorías de la variable dependiente es la variable STPI212, que corresponde al servicio ofrecido por el personal del establecimiento y le sigue en importancia la variable STPI214 que corresponde a la variedad de la oferta de hospedaje.

Servicio de alimentación y bebidas.

Para el caso de la modelación del servicio de alimentación y bebidas se consideran siete variables de las cuales la pregunta: *De acuerdo a su experiencia personal, ¿Cómo califica usted los servicios recibidos en el establecimiento de alimentos y bebidas?*, cuyas respuestas es la variable de interés. Las posibles respuestas son: Excelente, Bueno, Malo y Muy malo, las frecuencias en la base de datos es: 83, 348, 32 y 2, respectivamente. Las variables independientes se muestran en la Tabla 3.

En un primer análisis previo a la modelación, verificar que las variables independientes que se consideran tengan influencia sobre la variable de interés. Para tal efecto se realiza la prueba de independencia. El valor del estadístico de prueba para las seis distintas pruebas se proporcionan en la última columna de la Tabla 3, el valor teórico de la Chi-cuadrada con una significancia del $\alpha=0.05$, es de 16.91898, con nueve grados de libertad, por tanto, en todos los casos se rechaza la hipótesis de independencia a favor de la existencia de dependencia de la variable de interés con respecto a cada una de las variables independientes.

Variables	Descripción	Estadístico χ^2_c
STPI221	La experiencia en general con los establecimientos de alimentos y bebidas	
STPI222	La limpieza de los establecimientos de alimentos y bebidas.	407.5309
STPI223	La higiene de alimentos y bebidas.	374.1248
STPI224	El servicio ofrecido por el personal de los establecimientos de alimentos y bebidas.	783.2926
STPI225	La relación calidad/precio de los establecimientos de alimentos y bebidas.	403.3616
STPI226	La variedad de establecimientos de alimentos y bebidas.	416.1002
STPI227	La información sobre los establecimientos de alimentos y bebidas.	230.7897

Tabla 3. Descripción de las variables explicativas para la modelación del servicio de alimentos y bebidas y el estadístico para la prueba de independencia.

Estimación de los parámetros de los modelos.

En esta etapa de la modelación, se estiman los parámetros de los distintos modelos propuestos. Para efectuar los cálculos se utiliza los comandos del paquete estadístico R. Se ensayan distintos modelos para determinar el que mejor explique el comportamiento de los datos. En la Tabla 4 primera columna, se reportan las variables independientes de los cinco principales modelos ajustados. Además, se reporta los estadísticos de los modelos ajustados como Deviance residual y su p-valor para la bondad de ajuste, el AIC, el porcentaje de pronósticos que coinciden con los valores observados y el p-valor de la prueba de las rectas paralelas.

Modelo (Variables independientes)	Bondad de ajuste. Deviance (p-valor)	AIC	Pronósticos (%)	Prueba de las rectas paralelas. p-valor
M1: ~ STPI222+ STPI223+ STPI224+ STPI225+ STPI226+ STPI227	354.7458 (0.9998)	372.745	86.23	0.00131
M2: ~ STPI222 + STPI224+ STPI225+ STPI227	357.5974 (0.9998)	371.597	85.59	0.0006
M3: ~ STPI223 + STPI224 + STPI225	406.4535 (0.9627)	418.453	85.16	0.1441
M4: ~ STPI222 + STPI224 + STPI225	372.4561 (0.9988)	384.456	85.59	0.0019
M5: ~ STPI222 + STPI224 +STPI226	372.118 (0.9988)	384.118	86.23	0.0012

Tabla 4. Modelos ajustados para la variable dependiente *De acuerdo a su experiencia personal, ¿Cómo califica usted los servicios recibidos en el establecimiento de alimentos y bebidas?*(STPI221) y algunos de sus estadísticos.

Evaluación estadística de los modelos

La significancia estadística de los coeficientes de las variables independientes se realiza con estadístico de Wald. De los modelos reportados en la Tabla 4, el único modelo que tiene coeficientes estadísticamente igual a cero es el primero M1. Las variables independientes cuyos coeficientes tiene este problema son STPI223 y STPI226 con el p-valor de 0.8269 y 0.09287, respectivamente.

En la segunda columna se reporta el valor del estadístico Deviance y su p-valor para efectos de la prueba de bondad de ajuste. En esta prueba en la hipótesis H_0 se establece que el modelo es el adecuado, por lo que el p-valor debe ser mayor a 0.05 para no rechazar la hipótesis nula. Como se observa en los cinco modelos el p-valor es muy superior a 0.05. Por lo que los modelos son los adecuados.

Selección del modelos adecuado

En la Tabla 4, se reportan algunos de los modelos ajustados con sus respectivos estadísticos para realizar la selección del modelo adecuado a los datos. Si el criterio de selección fuera únicamente el AIC, se tendría que seleccionar el modelo M2 además, este modelo tiene el mayor p-valor de la prueba de bondad de ajuste. En cuanto al mejor porcentaje de coincidencia entre los valores pronosticados por el modelo y los valores observados, son los modelos M1 y M5 los que alcanzan los mayores valores. En la última columna de la Tabla 4, se proporciona el p-valor de la prueba de las rectas paralelas y se observa que el único modelo que satisface que este valor sea mayor a 0.05 es el modelo M3.

Ante el análisis planteado de los modelos ajustados, se tienen básicamente dos opciones seleccionar el modelo M1 que es el mejor considerando dos de los cuatro criterios (bondad de ajuste y mejor pronósticos) pero tiene el inconveniente de no satisfacer el supuesto de las rectas paralelas y tener el mayor número de variables independientes. La otra opción es seleccionar el modelo M3 que es el único que satisface el supuesto del modelo y por considerar solamente tres variables independientes. Ante esta situación, se opta por el modelo M3 por satisfacer el supuesto del modelo y contener el menor número de variables independientes aunque, no sea el mejor en los demás criterios.

Los parámetros ajustados del modelo seleccionado se proporcionan en la siguiente expresión.

$$\log \left[\frac{P(Y \leq y_j | X)}{P(Y > y_j | X)} \right] = \alpha_j - 1.680(STPI223) - 1.776(STPI224) - 1.402(STPI225); \text{ con } j = 1, 2, 3.$$

Los valores ajustados de los interceptos son: $\alpha_1=6.4256$, $\alpha_2=10.8653$ y $\alpha_3=17.1230$.

Los valores del estadístico Wald, para la significancia de los coeficientes fueron: -5.460, -5.682 y -4.757 para STPI223, STPI224 y STPI225, respectivamente. Con estos valores del estadístico se comprueba la significancia de los coeficientes de las variables independientes.

Pronósticos con el modelo seleccionado.

Nuevamente utilizando instrucciones de R, se calcula los valores pronosticados con el modelo seleccionado. En la Tabla 5 se proporcionan tanto los valores pronosticados como los valores observados para cada una de las categorías de la variable de interés.

Observados/Predichos	1	2	3	4	Totales
1	1	1	0	0	2
2	0	12	19	1	32
3	0	2	332	14	348
4	0	0	32	51	83
Totales	1	15	383	66	365

Tabla 5. Pronósticos con el modelo seleccionado, según categoría.

Análisis con el modelo seleccionado

Según el modelo seleccionado, la variable que tiene mayor impacto sobre la variable de interés es STPI224 que corresponde al servicio ofrecido por el personal de los establecimientos de alimentos y bebidas; la siguiente variable es STPI223 que corresponde a la higiene de alimentos y bebidas y la última es STPI225 que corresponde a la relación calidad/precio de los establecimientos de alimentos y bebidas. Estas son las tres variables que influyen de manera determinante sobre la percepción de la calidad del servicio de alimentos y bebidas.

CONCLUSIONES Y PERSPECTIVAS

Los modelos de regresión logística ordinal ajustados a los datos de satisfacción del turista resultan satisfactorios. En primer lugar, se satisface el supuesto del modelo de la recta paralelas y su evaluación estadística es buena. En segundo lugar, mediante los modelos se han podido identificar aquellas variables que afectan directamente la satisfacción del turista, como es el caso de la modelación del servicio de alimentos y bebidas, donde se identifican estas variables como el servicio ofrecido por el personal, la higiene y la relación calidad/precio.

Perspectivas. Para trabajos posteriores se requiere de un mayor análisis del modelo en función de los datos, esto es, dado el modelo seleccionado cuantificar el impacto sobre la variable de interés de las distintas variables independientes.

REFERENCIAS

1. Agresti, A. "Categorical Data Analysis", USA, John Wiley & Sons, 2002.

2. Akaike, H. (1973). Information theory and extension of the maximum likelihood principle. In Petrov, B. and Csáki, F. (editors), *Second International Symposium on Information Theory*, page 267-281. Akadémiai Kiadó, Budapest.
3. Heredia J., Rodríguez A., Vilalta J. “Empleo de la regresión logística ordinal para la predicción del rendimiento académico”. *Revista Investigación Operacional*, Vol. 33, No. 3, 252-267, 2012.
4. Hosmer D., Lemeshow S. “Applied Logistic Regression”, USA, John Wiley & Sons, 2000.
5. Mc Cullagh, P. and Nelder, J.A. “Generalized Linear Models”, 2nd. ed. Chapman and Hall, 1989.
6. R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
7. Secretaria de Turismo, H. Ayuntamiento de Acapulco. “Encuesta sobre el grado de satisfacción del turista y derrama económica”, 2015.