**Ayudantía 8 – Estadísticas para la Gestión II**

**Felipe Herrera, Hernán Romo y David Toro**

1. Responda V o F. Justifique las falsas.
2. \_\_\_V\_\_\_ Un modelo de regresión logística puede ser utilizado para explicar variables que en su naturaleza sean cuantitativas continuas.

 Por ejemplo el peso. Podría recodificarse con 0 cuando el peso es menor a 50 kilos, y 1 cuando éste es igual o mayor a 50 kilos.

1. \_\_\_F\_\_\_ El modelo de regresión logística tiene un carácter lineal y constante, igual que el modelo de regresión lineal múltiple, pero se diferencia del modelo de regresión lineal simple en que éste último sólo tiene una variable dependiente.

El modelo de regresión logística no tiene un carácter lineal.

1. \_\_\_F\_\_\_ Los betas de la regresión logística pueden ser tratados de la misma manera que en la RLM, donde la evaluación de X junto a los betas arroja el valor estimado por el modelo.

La evaluación de X junto a los betas arroja un valor que debe reemplazarse en la siguiente fórmula para obtener la probabilidad:

1. \_\_\_V\_\_\_ El pseudo R^2 de un modelo de regresión logística tiene una interpretación equivalente a la de la RLM.
2. \_\_\_F\_\_\_ Cuando una variable obtiene un valor igual o menor a 0,05 en su “P value” conviene quitarla pues ésta no tiene peso suficiente en el modelo.

Cuando esto sucede, no debería quitarse la variable porque es estadísticamente significativa. Debe quitarse cuando su “P value” es superior a 0,05 (afirmación válida solo para un 95% NC).

1. \_\_\_F\_\_\_ Si bien un modelo de regresión logística puede predecir la probabilidad de ocurrencia de una categoría de una variable dicotómica, la efectividad de éste en su estimación no puede ser medida.

El test con Chi cuadrado permite evaluar cuán adecuado es un modelo. Si éste arroja un valor igual o menor a 0,05, el modelo podría considerarse estadísticamente significativo, por lo que tendría pertinencia en las probabilidades que predice.

1. Considerando la base de datos (Participación y Aborto.dta), responda las siguientes preguntas:

|  |  |
| --- | --- |
| **Variable** | **Codificación** |
| participacion | 0 = En contra1 = A favor |
| aborto | 0 = En contra1 = A favor |
| quintil | 1 = Primer quintil2 = Segundo quintil3 = Tercer quintil4 = Cuarto quintil5 = Quinto quintil |
| indigena | 0 = Indígena1 = No indígena |

1. Genere un modelo de Regresión Logística que explique la Participación Ciudadana a partir de las variables quintil, edad e indígena:



Logit (participación)= -2.723\*quintil+-0.073\*edad+-0.273\*indigena+14.445

Hay que tener en cuenta que la variable dependiente del modelo que estima stata no es la participación (como variable dicotómica) sino que el logit de la participación, o transformación logística de la razón de probabilidad (en la siguiente pregunta aparece bien escrito)

1. ¿Cuál es la probabilidad estimada de estar a favor de la Participación Ciudadana para una persona de 23 años de edad, que pertenezca el tercer quintil y que no sea indígena?

Logit=-2.723\*3+-0.073\*23+-0.273\*1+14.445=4.324

 p=1/(1+2.71^-4.324)=0.986

 La probabilidad que esté a favor de la participación ciudadana dados los datos es de un 98,6%

1. ¿El modelo predice las probabilidades y en consecuencia puede considerarse adecuado?

El modelo tiene una significancia menor a 0.05, por lo que tenemos baja probabilidad de cometer el error tipo I (rechazar la hipótesis nula cuando no debiese ser rechazada), es decir, rechazamos la hipótesis nula de que el modelo no predice las probabilidades, en consecuencia, el modelo debe considerarse como adecuado.

En otras palabras: si Prob> chi2 es menor al nivel de significancia (0.05 en este caso), el modelo es adecuado para predecir las probabilidades de ocurrencia de la variable dependiente. Si Prob> chi2 es MAYOR al nivel de significancia, el modelo NO es adecuado para predecir las probabilidades de ocurrencia de la variable dependiente.

<https://www.youtube.com/watch?v=4h1vKWLIsNc> minuto 36:00

1. En base a la misma base de datos, responda:
2. Genere un intervalo de confianza para la proporción de las personas que están a favor del aborto para cada partido político (80%NC)

bysort partido: ci aborto, bin wald level(80)

1. ¿Se puede afirmar que la proporción de las personas que están a favor del aborto es menor al 60% (a una significancia del 10%)?.



Dado que el valor Pr(Z<z)=0.0076 es menor que el nivel de significación (0.10), la probabilidad de cometer un error tipo I es menor que el 10% por lo que se debe rechazar la hipótesis nula, o en otras palabras no se debe rechazar la hipótesis alternativa. Se puede afirmar, con un 90% de nivel de confianza, que la proporción de personas que están a favor del aborto es menor al 60%

1. Genere un modelo de regresión logística que explique la disposición de las personas al aborto, explicado por las variables quintil, edad, indígena y su tendencia política (izquierda=0 o derecha=1)

gen tendencia=0

replace tendencia=1 if partido=="rn"

replace tendencia=1 if partido=="udi"

logit aborto quintil edad indigena tendencia



Logit Aborto=-1.804\*quintil+-0.022\*edad+-18.701\*indígena+-20.885\*tendencia+26.665

1. Calcule las siguientes probabilidades, suponiendo que la persona es indígena y pertenece al segundo quintil.

Logit Aborto=-1.804\*3+-0.022\*edad+-18.701\*0+-20.885\*tendencia+26.665



|  |  |  |
| --- | --- | --- |
| Edad | Izquierda | Derecha |
| 20 | Logit=22.617Prob=0,999 | Logit=1.732Prob=0,848 |
| 50 | Logit=21.957Prob=0,999 | Logit=1.072Prob=0,744 |
| 80 | Logit=21.297Prob=0,999 | Logit=0.412Prob=0.601 |

1. ¿El modelo predice las probabilidades y en consecuencia puede considerarse adecuado?



Sí, puede considerarse como adecuado.

1. Si el modelo puede considerarse adecuado, ¿es conveniente retirar variables de él? (90% NC)

