

EAA 2320/EAA331A: Marketing Analytics Prueba 2

Profesora: Daniela Zúñiga

Código de Honor:

Como miembro de la comunidad de la Pontificia Universidad Católica de Chile, me comprometo a respetar los principios y normativas que la rigen. Asimismo, me comprometo a actuar con rectitud y honestidad en esta evaluación.

Adicionalmente declaro estar en condiciones de salud adecuadas para rendir esta evaluación y que me presento a ésta bajo mi responsabilidad. En caso de sentirme mal o tener alguna complicación, deberé informarlo inmediatamente al ayudante o profesor.

Declaro haber leído este código de Honor y estoy de acuerdo con éste. Además, me comprometo a no comunicarme con compañeros de este curso en cualquiera de sus secciones durante la prueba

Tiempo: 90 minutos

Pregunta 1 (5 pts.)

En el siguiente cuadro puede ver la Matriz TF de 5 blogs que hicieron reportajes sobre la llegada de Apple Pay a Chile:

TF	Seguridad	Tecnología	Visa	Mastercard	Iphone	Apple Watch	Apple Wallet	Comodidad	Google Wallet	Bancos	Total
Blog1	0.20	0.20	0.20	0.20	0.20	0.00	0.00	0.00	0.00	0.00	1.00
Blog2	0.00	0.17	0.00	0.00	0.17	0.17	0.17	0.17	0.17	0.00	1.00
Blog3	0.33	0.17	0.00	0.00	0.17	0.17	0.00	0.00	0.00	0.17	1.00
Blog4	0.00	0.00	0.25	0.25	0.00	0.25	0.00	0.25	0.00	0.00	1.00
Blog5	0.25	0.00	0.00	0.00	0.25	0.25	0.00	0.25	0.00	0.00	1.00
IDF	1.22	1.22	1.40	1.40	1.10	1.10	1.70	1.22	1.70	1.70	

- a) (3 pts) ¿Qué significa que el término Seguridad tenga un TF de 0,33 en el Blog 3? ¿Podemos decir que se debe a que hay más menciones del término Seguridad en ese blog vs. el resto?

Significa que en ese Blog, el 33% de las palabras corresponden a Seguridad (1.5 puntos).

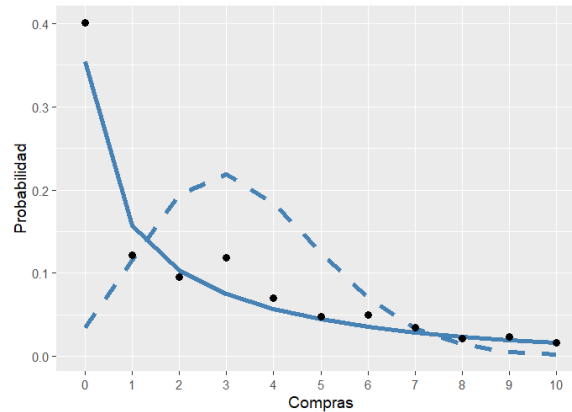
Que tenga un TF más alto que los otros Blogs significa que en términos relativos, ese término fue nombrado más veces en ese documento que en los demás. No podemos decir que es así en términos absolutos, ya que puede que en todos los blogs la palabra haya aparecido sólo una vez y que la razón de que su TF sea mayor sea que es el Blog 5 tiene menos palabras. (1.5 punto)

- b) (2 pts) ¿Cómo interpreta usted que el IDF de la palabra Iphone sea de los más bajos junto con la palabra Apple Watch?

La interpretación de un IDF bajo, cercano a 1 es que se trata de una palabra que aparece comúnmente en los Blogs, y que por lo tanto no es muy característica de ninguno de ellos. (1 punto) En este caso, podemos interpretar que la palabra Apple Watch y Iphone son muy comunes y no nos permiten diferenciar el contenido al que hacen referencia los blogs. (1 punto)

Pregunta 2 (10 pts.)

El ejercicio realizado en ayudantía sobre las compras trimestrales de arroz estima un modelo Poisson - Gamma con parámetros $r=0.50$ y $s=0.15$. La figura siguiente muestra la frecuencia empírica de compras mediante los puntos en la figura, y los modelos Poisson - Gamma (Binomial Negativo), en línea continua y el modelo Poisson en línea discontinua.



- (i) (3 pts.) ¿Qué se puede desprender de la figura en relación a los dos modelos presentados: modelo Poisson – Gamma (Binomial Negativo) y modelo Poisson? Explique la figura con los hogares que compran 2 kilos de arroz en un trimestre.

Se puede observar en la figura que el modelo PG (BN) se ajusta mucho mejor a los datos observados que el modelo P. Por ejemplo, el 10% de los hogares compra 2 kilos lo cual es casi exactamente igual al modelo PG mientras que el modelo P estima que el 20% compra 2 kilos.

- (ii) (5 pts.) Si un hogar no ha comprado arroz en el último trimestre, ¿cuál es la probabilidad de que compre arroz en el próximo año?

Si un hogar no ha comprado en el último trimestre, entonces sus parámetros son: $r^* = 0.51 + 0 = 0.51$ y $s^* = 0.15 + 1 = 1.15$. La probabilidad de que no compre en el próximo año es ($T^*=4$) igual a $(1.15/(1.15+4))^{0.51} = 0.466$

- (iii) (2 pts.) Calcule la compra esperada anual de los hogares que no compraron arroz el último trimestre.

Si un hogar no ha comprado en el último trimestre, entonces sus parámetros son: $r^* = 0.51 + 0 = 0.51$ y $s^* = 0.15 + 1 = 1.15$. Compra esperada = $(0.51/1.15)*4=1.77$

Fórmulas:

$$Penetración = 1 - \left(\frac{s}{s+T} \right)^r \quad Compra Esperada = \frac{r}{s} T$$

Ajuste de parámetros por información histórica

$$r^* = r + n$$

$$s^* = s + T$$

Pregunta 3 (10 pts.)

El fabricante de los helados San Francisco quiere estudiar las conductas de compra de los consumidores de helados en formato cassata o familiar para implementar acciones orientadas a aumentar su participación de mercado. Para esto utiliza un panel de 400 consumidores en el cual se registran mediante un scanner las compras de helados tipo cassata en el canal supermercado durante 1 año. Para cada compra de un helado cassata se registra la marca, el precio pagado en miles de pesos, si el producto está en *display* en un espacio destacado del supermercado y si se han distribuido en la sala de ventas *flyers* o folletos promocionando el producto. Estas dos últimas variables tienen la forma si/no correspondiente a los valores 1 y 0 respectivamente. Además, se consideran en el modelo las constantes de marca.

La tabla siguiente muestra la estimación de dos modelos de regresión logística (logit) para explicar la elección de marca en esta categoría en función de las variables independientes mencionadas. La marca considerada como base (constante = 0) es San Francisco. El Modelo 1 en la tabla considera sólo las constantes de marca y el Modelo 2 agrega las otras variables definidas.

Estimaciones Modelo Logístico Elección de Marca

Variable	Modelo 1			Modelo 2*		
	Coeficiente	Error Estándar	t	Coeficiente	Error Estándar	t
Precio (miles \$)				-1.25	0.18	-6.9
Display (0/1)				0.88	0.46	1.9
Flyer/Folleto (0/1)				0.19	0.05	3.8
Constante Savory	0.84	0.25	3.4	0.83	0.39	2.1
Constante Bresler	0.73	0.23	3.2	0.44	0.25	1.8

(*) Pseudo R2 = 52%, porcentaje de clasificaciones correctas en la muestra igual a 68%

Suponga que en el formato cassata o familiar sólo existen las tres marcas consideradas en el estudio. Responda las siguientes preguntas.

- a) (5 pts.) Según el modelo 2 ¿Cuál es el diferencial de participación de mercado de Savory respecto de un escenario donde no existe diferencia alguna entre las 3 marcas?

Sin diferenciación: $P(\text{Savory}) = 33\%$

Con diferenciación (asumiento igual precio, display, flyer) la prob sólo se verá influenciada por la cte de marca.

Ut Savory= 0.83

$P(\text{Savory}) = \exp(0.83) / (\exp(0.83) + \exp(0) + \exp(0.44)) = 0.47$

Diferencia: $0.47 - 0.33$. Considerar correcto si pusieron supuestos en precio/display/flyer

- b) (5 pts.) ¿Qué le recomendaría usted a San Francisco si esta marca quiere aumentar su participación de mercado: disminuir el precio en \$1 mil pesos o poner los helados San Francisco en *display*? Explique.

Caso 1: Cambio Utilidad = -1.25*(-1) = 1.25

Caso 2: Cambio utilidad = 0.88 aunque casi no significativo

Pueden haber otras consideraciones: efectos de largo plazo de bajar precios, otros.

Fórmula:

$$\Pr(y_{ij} = 1) = \frac{e^{v_{ij}}}{\sum_{k=1}^J e^{v_{ik}}}$$

Pregunta 4 (15 pts.)

MasterClass es un servicio de *streaming* online que ofrece clases de diversos temas impartidas por personas expertas en sus respectivas áreas. Los y las instructoras van desde escritores ganadores del Premio Nobel hasta chefs con estrellas Michelin, directores de cine premiados con el Oscar y más. Las clases cubren una amplia gama de temas como escritura, cine, cocina, música, deportes, entre otros. En cada clase, los instructores comparten sus experiencias, habilidades y consejos a través de videos de alta calidad y materiales complementarios.

Actualmente, Masterclass ofrece tres planes: individual, duo y familiar; y está pensando modificar el plan de suscripción individual y para eso diseñó e implementó un análisis conjunto basado en ratings con los atributos y niveles presentados en la tabla 1.

1000 encuestados evaluaron distintos perfiles utilizando una escala de 1 a 10. Una vez obtenidas las respuestas, se estimó un modelo de regresión lineal obteniéndose los siguientes resultados:

Tabla 1. Resultados de la estimación

Atributo	Nivel	Segmento 1 (40%)	Segmento 2 (60%)
Precio Mensual	\$9	0	0
	\$11	-0,3	-0,7
	\$13	-0,5	-0,9
Acceso a todas las clases	No	0	0
	Si	0,3	0,1
Interacción con instructores	No	0	0
	Si	0,6	0,2

Material extra de aprendizaje	No	0	0
	Si	0,4	0,2
Descargas offline	No	0	0
	Si	0,4	0,2

1. Los valores en la tercera y cuarta columna son los coeficientes promedio del modelo de regresión para cada segmento. Estos representan las utilidades parciales (part-worths).
2. Los niveles con coeficientes igual a cero corresponden al nivel base (y por lo tanto no fueron estimados)
3. Todos los coeficientes estimados son estadísticamente significativos al 5% ($p < 0.05$)
4. Notar que la probabilidad de pertenecer al segmento 1 es 0,4.

Basado en estos resultados, responda a las siguientes preguntas:

1. (6 pts) El plan básico actual de la empresa (que cuesta \$9) no ofrece hoy la posibilidad de interactuar con los instructores. La gerenta comercial de la empresa evalúa ofrecer este servicio y subir el precio del plan básico a \$10 ¿usted recomendaría mantener el servicio original, ofrecer el nuevo servicio u ofrecer ambos? Determine la disposición a pagar por este atributo y utilícelo en su argumentación.

Para el S1:

Cambio en ut/cambio precio: $0.3/(11-9) = 0.15$ utiles / \$

Aumento por interacción con instructores: $+0.6 \rightarrow$ Disposición a pagar $= 0.6/0.15 = \$4$ (2 pts)

Para el S2:

Cambio en ut/cambio precio: $0.7/(11-9) = 0.35$ utiles / \$

Aumento por interacción con instructores: $+0.2 \rightarrow$ Disposición a pagar $= 0.2/0.35 = \$0.57$ (2 pts)

Como quieren subirlo a \$10, el S1 sí lo pagaría pero el S2 no. Le conviene ofrecer ambos planes. (2 pts)

2. (6 pts.) Suponga que la empresa decide ofrecer 2 planes individuales con las siguientes características:
Plan A: \$9, Acceso a todas las clases, Sin interacción con instructores, Con material extra, Sin descargas
Plan B: \$13, Acceso a todas las clases, Con interacción con instructores, Con material extra, Con descargas

Utilizando el modelo logit, estime una participación de mercado para ambos planes **en el segmento 2**.

Fórmula:

$$\Pr(y_{ij} = 1) = \frac{e^{v_{ij}}}{\sum_{k=1}^J e^{v_{ik}}}$$

$$Ut A = 0 + 0.1 + 0 + 0.2 + 0 = 0.3$$

$$Ut B = -0.9 + 0.1 + 0.2 + 0.2 + 0.2 = -0.2$$

$$P(A) = \exp(0.3)/\exp(0.3)+\exp(-0.2)=0.62$$

$$P(B) = \exp(-0.2)/\exp(0.3)+\exp(-0.2)=0.38$$

3. (3 pts) ¿Por qué los ejecutivos utilizan la metodología de análisis conjunto, que es una de las metodologías indirectas y no se pregunta directamente, a los potenciales consumidores, por las utilidades que les reporta cada nivel de cada atributo? Explique.
- **Personas no dicen la verdad o es difícil juzgar importancia de atributos (1)**
 - **Sesgos de deseabilidad social (1)**
 - **No permite incluir interacciones (1)**

Pregunta 5 (10 pts)

Una empresa realiza una promoción con descuentos significativos a quienes compren sus productos via online. Un modelo de regresión logística (logit) resume los resultados de la campaña en términos de los clientes según las variables Recencia, Frecuencia y Monto. El modelo, para la probabilidad de respuesta o compra igual a p , se puede escribir como

$$\ln(p/(1-p)) = 0,012 - 0,034 \text{ Recencia} + 0,126 \text{ Frecuencia} + 0,004 \text{ Monto}$$

(0,003) (0,001) (0,025) (0,002)

Entre paréntesis se presentan los errores estándar.

1. (3 pts.) ¿Estaría de acuerdo usted en que la variable más relevante para predecir la respuesta o compra es la Frecuencia? Explique por qué si o por qué no.

No porque la importancia no se mide con los coeficientes sino con los estadísticos t que son: 34, 5 y 2 respectivamente. La variable más importante es la Recencia.

2. (5 pts.) ¿Cuál es la estimación para un cliente con $R = 4$, $F = 12$ y $M = 340$? Calcule e interprete la probabilidad p y los odds ($p/(1-p)$).

El logit es 2,748, los odds son 15,6 (aproximadamente 16 a 1 lo que quiere decir que de cada 17 casos en 16 compra y en 1 no). La probabilidad es 0,939 (en la escala 0 a 100 la probabilidad es 93,9%)

3. (2 pts) Si se quiere maximizar el uso de su presupuesto enfocándose en la certeza de las predicciones positivas, ¿qué métrica de la matriz de confusión debería priorizar al evaluar su modelo?

Debería priorizar la métrica de precisión en su modelo. La precisión se refiere a la proporción de verdaderos positivos frente al total de positivos predichos (es decir, el número de clientes que realmente realizaron una compra dividida por el total de clientes que el modelo predijo que harían una compra). Priorizar la precisión ayudará a asegurarse de que el presupuesto de marketing se gaste en clientes que tienen más probabilidades de realizar una compra, minimizando el costo de dirigirse a aquellos que no comprarán. Precision = $TP / (TP + FP)$

Fórmulas métricas

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Specificity} = \frac{TN}{TN + FP}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$