

# Análisis de variables mediante curvas ROC y modelos categóricos

Manuel Pelegrina, M. Ruiz-Soler, E. López y A. Wallace  
Universidad de Málaga

Mediante este trabajo proponemos establecer una relación formal entre los modelos basados en las curvas características operativas del receptor ROC y los modelos de análisis para datos categóricos. Tradicionalmente ha existido una separación entre el análisis propuesto en la teoría de la detección de señales (TDS) y el propuesto por los modelos lineales generalizados (MLGs). Sin embargo, diferentes autores han sugerido algún tipo de relación específica (v.g. Dorfman, y Alf, 1968; Swets, 1986; DeCarlo, 1998; y Tosteson y Begg, 1988, entre otros). Así, por ejemplo, los modelos categóricos generan tablas de contingencia similares a las respuestas condicionadas de la TDS. En consecuencia, es posible incluir medidas estandarizadas de asociación derivadas de los modelos estadísticos (Bishop Fienberg and Holland, 1975). En este sentido, algunas medidas son funciones de la razón de productos cruzados (independientes de los totales marginales) como LOR,  $\eta$  and  $Q$ . Estos índices son también consistentes con la existencia de criterio de decisión (Swets, 1986, 1996) y mediante ellos es posible la aplicación del análisis ROC. Hay además otros índices en los que es posible aplicar también un análisis ROC, pero que implican un modelo de umbral. Resumiendo, mediante esta investigación proponemos evaluar empíricamente los datos en un sentido complementario: TDS y MLGs.

*Variable analysis by means of ROC curves and categorical models.* Formal relationship between the relative (or receiver) operating characteristics (ROC) models and the analytical models for categorical data is proposed. Traditionally, some differences have been established between signal detection theory (TDS) models and Generalized Linear Models (GLMs). However, some authors have suggested some specific relations (v.g. Dorfman y Alf, 1968; Swets, 1986; DeCarlo, 1998 and Tosteson y Begg, 1988). For example, the categorical models generate results in a contingency table similar to the conditional responses to TDS. Therefore, it is possible to include standard measures of association derived from statistical models (Bishop Fienberg and Holland, 1975). Hence, same measures are function of the cross-product ratio (independent of marginal totals) as LOR,  $\eta$  and  $Q$ . These indices are also consistent with a variable-criterion model (Swets, 1986, 1996), and ROC analysis can be applied. There are also other indices consistent with ROC analysis that imply a threshold model. In short, we propound that it is possible to evaluate the empirical data by using two models that can be complementary: TDS and GLMs models.

Poco antes del surgimiento de la teoría de la detección de señales (TDS), Wald (1950) elabora lo que denominó funciones estadísticas de decisión. La curva característica operativa del receptor o *Receiver Operating Characteristics (ROC) analysis* surgió en el contexto de las citadas funciones y supuso un fundamento importante para la formalización de la TDS unos años después. Las curvas ROC representan la ejecución de un observador que clasifica un suceso «each stimulus must be classified, or placed, in one of two categories, sn or n» (Egan, 1975, p. 7). A esta clasificación que corresponde a procedimientos o tareas SI-NO se añaden los procedimientos de escalas de estimación y de elección forzada que permitan establecer algún tipo de diferencia entre dos señales. En este ámbito se ha valorado la calidad industrial (Baker, 1975), la precisión en el diagnóstico

(Swets, 1988), la evaluación del aprendizaje (Brickley, Prytherch, Kay y Shepherd, 1995), la valoración subjetiva del dolor (Coppola & Gracely, 1983) y variables propias de la psicofísica clínica (Grossberg, & Grant (1978), entre otras muchas posibilidades.

El objetivo que se plantea, desde cualquiera de los tres procedimientos clásicos citados anteriormente es el estudio del resultado obtenido en el rendimiento de un observador al valorar una señal. El más común corresponde a respuestas SÍ-NO y a escalas de estimación. La tabla 1 representa un punto de partida basado en tablas de contingencia 2x2 o de clasificación de la respuesta y permite su extensión a tablas a x b. Dentro de las tablas de contingencia las procedentes de la TDS son tablas de respuestas condicionadas por la presencia de señal o ruido (o matrices de transición, tabla 1).

Desde el modelo de la TDS partimos del supuesto de que cuando el sujeto elige un valor de la escala utiliza como criterio alguna razón monótona basada en la máxima verosimilitud: Es decir, los valores de sensación de la señal sobre los valores de sensación del ruido. Ello se puede representar mediante la fórmula siguiente:

$$L_{SR} = \frac{p(s/S)}{p(s/R)} \tag{1}$$

donde  $L_{sr}$  es la razón de verosimilitud,  $p(s/S)$  es la probabilidad de sensación dada la señal y  $P(s/R)$ , es la probabilidad de sensación dado el ruido sólo. No es adecuado el ajuste de estos datos mediante mínimos cuadrados, sino mediante la razón de verosimilitud, entre las que se incorpora la razón logística de verosimilitud (Reales y Ballesteros, 1997). Esta formulación matemática no representa la ejecución del observador sino el modelo ideal de decisión. Dicha fórmula es una optimización de la ejecución. La máxima verosimilitud representa el observador ideal, la «hipótesis» basada en el criterio del sujeto ante lo evidente, a partir del cual se compara el criterio óptimo esperado. Por ello, se espera que el criterio del sujeto será algún valor de dicha razón de verosimilitud (o próximo a ella).

A partir de este modelo, y dependiendo de la hipótesis de partida, podemos planificar los correspondientes diseños. En cualquier caso y para una tabla 2x2 obtenemos (ver tabla 2):

- A= Número de respuesta sí a la ocurrencia de señal.
- B= Número de respuestas negativas ante la ocurrencia de señal.
- C= Número de respuestas positivas ante la no ocurrencia de señal.
- D= Número de rechazos correctos a la no ocurrencia de señales.
- N= A+B+C+D, donde N = número total de respuestas.
- A+B es la frecuencia de respuestas a las ocurrencias de señal.
- C+D es la frecuencia de respuestas a las no ocurrencias de señal.

Mediante los datos anteriores podemos conocer diferentes probabilidades de respuesta:

- Proporción de respuestas positivas verdaderas o aciertos.
- Proporción de respuestas negativas falsas o rechazos incorrectos.
- Proporción de respuestas positivas falsas o falsas alarmas.
- Proporción de respuestas negativas verdaderas o rechazos correctos.

-Otros índices que analizaremos posteriormente en función de los objetivos de la investigación, (Swets, 1986a; Swets y Pickett, 1982).

Variable del estímulo	Variable de respuesta		Total
	Sí ( $b_1$ )	No ( $b_2$ )	
Señales ( $a_1$ )	$b_1/a_1$ A	$b_2/a_1$ B	A+B
Ruidos ( $a_2$ )	$b_1/a_2$ C	$b_2/a_2$ D	C+D
Total	A+C	B+D	N= A+B+C+D

Respuesta Señal	SI	NO
	Ocurrencias	A (aciertos) <i>Sensibilidad: A/A+B</i>
No ocurrencias	C (falsas alarmas)	D (rechazos correctos) <i>Especificidad: D/C+D</i>

Un segundo procedimiento corresponde a tareas basadas en escalas de estimación. Éstas se caracterizan porque el sujeto valora la intensidad de su respuesta (por ejemplo, mediante una escala de 1 a 5, ver tabla 3). También denominada con los términos de clasificación, apreciación, evaluación, categorización, etc. (v.g. Parducci & Wedell, 1986).

Se parte de la aceptación del supuesto de que a cada punto de la escala corresponde una razón de respuestas basada en la máxima verosimilitud, así como de la independencia entre los valores de la escala con otros procesos cognitivos (v.g. Irwin & Hautus, 1997; Penney & Balsom, 1993) ajenos a la discriminación. Mediante este procedimiento asumimos que el sujeto asigna razones de probabilidad a los valores de la escala condicionado por tales razones de probabilidad. Este procedimiento fue ya utilizado en psicofísica por Jersild (1929) y en tareas de detección por Swets Tanner y Birdsall (1961), Pollack y Decker (1958), Egan y Clarke (1956) y Egan, Schulman y Greenberg (1959). En nuestro trabajo ha sido el procedimiento más usual, tanto con estímulos lingüísticos como mediante estímulos pictóricos o dibujos (v.g., Pelegrina, 1988b).

La secuencia de presentación de un ensayo sigue el mismo proceso que en el procedimiento SI-NO, salvo que en el intervalo de respuesta el sujeto tiene varias posibilidades en lugar de las dos respuestas típicas del procedimiento SÍ - NO. La ventaja del procedimiento de escalas de estimación consiste en que sin repetir el diseño con diversos niveles de las variables, podemos obtener un punto en la curva ROC para cada nivel de la escala. Además, en este procedimiento los valores de la VD se consideran acumulados. El último punto representa el porcentaje acumulado de todos los demás y, por tanto, es igual a la unidad (ver tabla 4). Ello indica que una respuesta situada en el grado dos de la escala incluye la respuesta uno, etc. Es decir, si el sujeto discrimina con una seguridad cinco ello incluye todas las discriminaciones menores de cinco.

Las respuestas de la tabla 4 pueden ser consideradas desde la perspectiva del procedimiento SI-NO. Para ello basta con elegir un punto de corte, o bien la mediana, en caso de obtener una variabilidad alta de aciertos y falsas alarmas respectivamente.

Respuesta Señal	Niveles altos	Niveles bajos
Ocurrencia	escala	escala
No ocurrencia	escala	escala

	Escala de estimación				
	1	2	3	4	5
Acu.	0.10	0.42	0.65	0.75	1
A	(0.10)	(0.32)	(0.23)	(0.10)	(0.25)
Acu.	0.09	0.30	0.56	0.70	1
FA	(0.09)	(0.21)	(0.26)	(0.14)	(0.30)

Cuadro 4. Representación de las proporciones de aciertos (A) y falsas alarmas (FA) mediante una escala de estimación de cinco puntos, así como sus correspondientes valores acumulados.

Sin embargo, es posible integrar en un mismo ensayo los procedimientos SÍ-NO y de escalas de estimación, instruyendo al sujeto para que en cada respuesta «SÍ» - «NO» incluya la valoración de la escala. Con ello se obtienen diez puntos con nueve grados de libertad (tabla 5).

Ello supone la aceptación de una continuidad y no simetría entre A y FA y entre RI y RC. Es decir, la distribución anterior supone que el sujeto o el observador categoriza o clasifica las respuestas de una manera cuando responde «SÍ» y de otra manera cuando responde «NO». La TDS clásica considera complementarias ambas respuestas de manera que se analizan desde este modelo las dos posibilidades a considerar, los A y FA que con complementarios con los RC y RI. En consecuencia, en la tabla 5 la curva ROC se calcularía mediante las respuestas «SÍ», es decir mediante A y FA.

La ventaja del procedimiento anterior es que en cualquier experimento, registro observacional o encuesta en el que el sujeto (o encuestador) responda (o registre) «SÍ» o «NO» o bien mediante una escala de estimación, o combinando ambos, podemos representar la curva ROC y comprobar la distribución obtenida. Lo anterior se puede generalizar a tablas 2 x 2 (v.g. Green y Birdsall, 1978). Así, representando las escalas de estimación mediante tablas de contingencia podemos observar la simetría o asimetría con respecto a múltiples señales: es decir, ver si el sujeto escoge o valora unos intervalos con preferencia a otros (Luce,1959), o ver si el sujeto eligió de acuerdo con la razón de verosimilitud.

En este sentido Swets (1986a) propone una serie de índices, propios de tablas de contingencia (v.g. Pelegrina, Salvador y Ortiz, 1999). Entre ellos el índice *Kappa* de Cohen, 1960 y el índice *Q* de Yule citado por Bishop, Fienberg y Holland (1975). Consideramos además de una manera especial el hecho de que las respuestas «SÍ» - «NO» a la señal y el ruido formen distribuciones de respuesta ordinales, siendo la matriz de confusión una tabla de contingencia con respuestas condicionadas ordinales. Ello da lugar a la incorporación de medidas no paramétricas (en el sentido de Macmillan, 1993), pero también a la incorporación de las tablas de contingencia propias de los modelos categóricos y observacionales (Anguera, 1981, Agresti, 1984, 1989, 1990, Ato y López, 1996,

Cradit, Tashchian, y Hofacker, 1994; Kennedy, 1983, Wickens y Olzak (1989) y Reynolds, 1977), así como su inclusión en los modelos lineales generalizados (MLGs) (DeCarlo, 1998).

En este sentido, los modelos de detección de señales pueden ser formulados como una subclase de los (MLGs) y el resultado es una clase de detección de señales rica basada en modelos con diferentes distribuciones subyacentes (DeCarlo, 1998). Entre tales modelos cabe destacar la regresión logística, la cual proporciona una vía para estimar los parámetros de la detección de señales y puede ser utilizada para respuestas binarias. El cálculo de la discriminación en un modelo basado en la distribución logística viene determinado por la fórmula siguiente:

$$d = \frac{\Psi_s - \Psi_r}{\tau} \tag{2}$$

donde  $\Psi_s$  y  $\Psi_r$  son las modas de las distribuciones de la señal y el ruido y  $\tau$  es el parámetro escala (ver McMillan y Creelman, 1991). En una distribución normal  $d$  se corresponde con  $d'$  y  $\tau$  con  $\sigma$  (DeCarlo, 1997).

Método

Sujetos

Los datos proceden de una muestra de 60 sujetos del área de Metodología de edad comprendida entre 18 y 24 años.

Material

Se utilizó la plantilla de recogida de datos elaborada por Pelegrina et al (1999).

Procedimiento

El procedimiento consistió en realizar una tarea típica de memoria de reconocimiento. En primer lugar se pasó una lista de contenidos propios de las asignaturas de metodología Pelegrina et al

NO						SI					
R	5	4	3	2	1	R	1	2	3	4	5
E	1	2	3	4	5	E	6	7	8	9	10
RC	1	3	10	40	50	A	54	42	15	4	1
p	0.004	0.01	0.04	0.18	0.22	P	0.24	0.19	0.06	0.02	0.004
PA	0.004	0.014	0.054	0.234	0.454	PA	0.694	0.884	0.944	0.964	0.968
RI	1	2	6	35	42	RO	0.24	0.43	0.49	0.51	0.514
p	0.004	0.009	0.003	0.17	0.20	<b>FA</b>	<b>50</b>	<b>42</b>	<b>20</b>	<b>3</b>	<b>1</b>
PA	0.004	0.013	0.016	0.186	0.386	<b>P</b>	<b>0.25</b>	<b>0.21</b>	<b>0.1</b>	<b>0.01</b>	<b>0.004</b>
						<b>PA</b>	<b>0.636</b>	<b>0.846</b>	<b>0.946</b>	<b>0.956</b>	<b>0.96</b>
						<b>RO</b>	<b>0.25</b>	<b>0.46</b>	<b>0.56</b>	<b>0.57</b>	<b>0.574</b>
<--- Evidencia sobre ausencia de señal Especificidad						Evidencia sobre presencia de señal ----> Sensibilidad					

(1999), escrita en letra escrita en *France* de tamaño 11. La tarea consistió en leer lo más rápidamente posible las frases allí presentadas. El tiempo de lectura fue de un segundo por frase. Pasada una semana, a la misma hora se pasó el cuestionario (la misma plantilla citada). Cada alumno/a contestó «SÍ» o «NO» y añadió su seguridad en la respuesta mediante una escala de estimación de uno a cinco. Cinco indicaba el mayor nivel de seguridad, cuatro un nivel alto de seguridad, tres un nivel medio dos un nivel bajo y uno un nivel muy bajo de seguridad en la respuesta.

### Resultados y discusión

Las proporciones de los los datos obtenidos se representan en la tabla 6. Son datos para representar en una curva ROC Estos datos proceden de 2.124 estímulos señales y 2.124 estímulos «ruidos». Corresponden a un diseño en el que las celdillas dieron resultados asimétricos Ello ha sido obtenido en trabajos anteriores (Pelegrina, 1988; Pelegrina y Salvador, 1989). La comparación mediante diferentes modelos de análisis ofreció los resultados siguientes: Mediante  $\chi^2$  no resultó ser estadísticamente significativa, mediante una correlación de Spearman resultó una correlación de -0.85 entre S y R y  $p < 0.000$ . No obstante, aplicamos también un ajuste propio de los MLGs. Así, mediante la aplicación del paquete

estadístico IEM obtuvimos un modelo logit estimado mediante un modelo loglineal no estándar que se formula como:

$$C|AB \{AC,BC\} \quad (3)$$

Los estadísticos a destacar del modelo fueron: L-squared = 3.792 (0.4333) y 4 gl. BIC (L-squared) = -29.6134. Para C Wald (1) = 57.69,  $p < 0.000$ , para AC Wald (4) = 285.19, y  $p < 0.000$  y para BC Wald (1) = 7.34,  $p < 0.007$ . Vemos que el modelo queda bien ajustado, «C» es la respuesta SI-NO, «A» la escala de estimación y «B» la señal-ruido. Es decir A B afecta a C. y la escala de estimación (A) parece estar asociada (o influir) en la respuesta SI-NO (C). La señal-ruido (B) parece igualmente estar asociada (o influir) en la respuesta SI-NO (C).

No obstante, en el momento actual no estamos seguros de la familia de modelos, dentro de los MLGs, en los que podríamos centrar el análisis de las tablas de datos obtenidas en TDS. Sin embargo, tal vez podamos avanzar algunos conceptos claves: podemos empezar por la regresión logística cuando utilizamos el procedimiento SÍ-NO, los modelos logit cuando se utilizan escalas de estimación (ordinales) y los modelos loglineales cuando hay interdependencia entre las variables.

Un análisis comparativo de los datos obtenidos en nuestros trabajos anteriores permite mostrar que, en términos generales, la discriminación es un proceso inverso al incremento de la información. En el trabajo actual la discriminación ha resultado ser mínima, próxima al azar, pero se ha podido ajustar un modelo mediante el programa IEM. Entendemos que los MLGs pueden resultar ser un buen instrumento de análisis para datos obtenidos mediante TDS, aunque en el momento actual (y hasta nuevas réplicas) carecemos de datos suficientes para generalizar el modelo obtenido.

	Respuestas SI					Respuestas NO				
S	405	402	407	164	51	121	117	146	359	171
R	387	140	401	151	40	139	162	167	129	189

### Referencias

- DeCarlo, L.T. (1998). Signal detection theory and generalized linear models. *Psychological Methods*, 3 (2).
- Dorfman, D.D., & Alf, E. Jr. (1968). Maximun Likelihood estimation of parameters of signal detection theory. A direct solution. *Psychometrika*, 33, 117-124.
- McCullagh, P. (1980). Regression modelos for ordinal data. *Journal of the Royal Statistical Society, Series B*, 42, 109-142.
- Pelegrina, M.(1988a). Detección de señales en procesos perceptivos y cognitivos. En F. Salvador(Ed.). *Nuevas perspectivas metodológicas en procesos perceptuales y cognición* (pp. 21-44). Barcelona: PPU.
- Pelegrina, M., Ruiz, M., López, E., Ortiz, M. y Videra, A. (1999). *Mejoras obtenidas en el rendimiento de las asignaturas metodológicas median - te la participación activa de los alumnos*. Póster presentado en las Jornadas 1999 de Proyectos de Innovación Educativa Universitaria sobre mejora de la Práctica Docente (ICE). Universidad de Málaga.
- Reales, J. M., y Ballesteros, S. (1997). *TDS. Un Programa de ordenador para la teoría de la detección de señales*. Madrid: Unoversitas.
- Swets, J. A. (1986). Indices of discrimination or diagnostic accuracy: Their ROCs and implied models. *Psychological Bulletin*, 99 (1), 100-117.
- Swets, J. A. (1996). *Signal detection theory and ROC analysis in psychology and diagnostic: Colected paters*. Hillsdale, NJ: Erlbaum.
- Tosteson, A. N. A., & Begg, C.B. (1988). A general regression methodology for ROC curve estimation. *Medical Decision Making*, 91, (68-111).