

Resumen

En el ámbito de la Salud Pública, existe la necesidad de conocer en profundidad las características de las poblaciones y los problemas de salud. De esa manera se puede intervenir para mejorarlos. Es fundamental tener una idea de la situación de partida y para eso se recurre a las fuentes de datos existentes, fundamentales en los sistemas de vigilancia epidemiológica. Sin embargo pueden existir limitaciones en los indicadores generalmente utilizados en la epidemiología y salud pública, ya que muchas veces no toman en cuenta la estructura multivariada de la información o si la toman, lo hacen a través de algoritmos de cálculo que generan indicadores univariados para ganar en simplicidad, y no miden por lo tanto, correctamente los fenómenos bajo estudio. Se propone construir un conjunto de indicadores alternativos y complementarios a los que ya existen en salud oral. Se reformularán los índices recomendados de la Organización Mundial de la Salud (CPO, SIC); en el estudio de la salud oral de la población habitualmente se usa el CPO (número total de piezas cariadas, perdidas y obturadas). La aplicación se hace con información proveniente del relevamiento en necesidades de tratamiento y demanda de servicios de salud bucal, de la población de Trabajo por Uruguay (TPU) a agosto de 2007. Se presentan las limitaciones que tiene trabajar con un *índice agregado univariado*. Como alternativa se propone modelar proporciones construidas a partir del CPO, para lo que se utilizan distintos tipos de *modelos lineales generalizados (GLM)* adaptados para modelar proporciones. Se utilizan modelos de *regresión beta* (mediante una reparametrización adecuada), propuesta por Ferrari y Cribari-Neto (2004) para modelar variables de respuesta continua a valores en el intervalo (0, 1). En este caso las variables explicativas tienen que ver con aspectos sociodemográficos individuales (edad, sexo, educación), de contexto (región, barrio) y con la historia de salud bucal de cada individuo (el motivo de su consulta, la cantidad de prótesis, el tiempo sin concurrir al dentista, etc)

Presentación del problema

Índice CPO: Es un índice *unidimensional* que cuenta el número de dientes cariados (C), perdidos (P) y obturados (O). Ha sido utilizado durante mucho tiempo como una forma de determinar la historia de *caries* de un conjunto de individuos. Los valores bajos de (CPO) indican un buen 'status' de salud oral, mostrando que las piezas dentales están sanas.

$$CPO_i = C_i + P_i + O_i \quad (1)$$

La población de estudio se compone de 1185 personas mayores de 18 años correspondiente al tercer llamado de TPU en Montevideo. Se obtuvo una muestra aleatoria mediante un muestreo multietápico (n=308) de individuos provenientes tres regiones de Montevideo. La encuesta fue realizada por docentes de la Cátedra de Odontología Social de la Facultad de Odontología, la que incluyó un examen clínico.

Distribución de los componentes del CPO

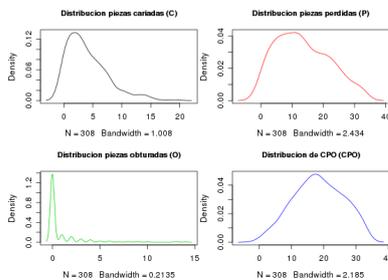


Figura : densidades componentes del CPO

Objetivos que se plantean en este línea de estudio

- Usando la misma información que habitualmente se releva en las encuestas de salud oral se pretende reformular el problema usando otra aproximación metodológica
- Relativizar los componentes del CPO convirtiéndolos en tasas o proporciones usando los 3 componentes del CPO. Por ejemplo
 - $\frac{\sum_{i=1}^n C_i}{\sum_{i=1}^n (C_i + P_i + O_i)}$ nivel de cobertura de la enfermedad previa a la entrada al programa (*prop1*)
 - $\frac{\sum_{i=1}^n P_i}{\sum_{i=1}^n (C_i + P_i + O_i)}$ indicador de estado de la enfermedad en el momento actual (*prop2*)
 - $\frac{\sum_{i=1}^n O_i}{\sum_{i=1}^n (C_i + P_i + O_i)}$ indicador de salud en caries en el momento actual (*prop3*)
 - $\frac{\sum_{i=1}^n P_i}{\sum_{i=1}^n (C_i + P_i + O_i)}$ indicador de necesidad de prótesis en el momento actual (*prop4*)
- Evaluar como se puede ampliar su uso a otros problemas de la epidemiología y la salud pública

Distribución de los componentes del CPO en tasas

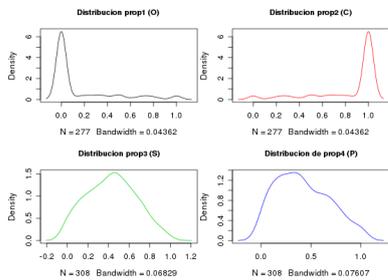


Figura : densidades componentes del CPO en tasas

Distribución Beta

Cualquiera de las proporciones de la figura 2 Y puede tener la siguiente función de densidad de tipo (BETA)

$$f(y; p, q) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} y^{p-1} (1-y)^{q-1}, \quad 0 < y < 1, \quad (2)$$

- Cribari y Neto [Cribari-Neto and Zeileis, 2010] hacen una reparametrización $\mu = \frac{p}{p+q}$ y $\phi = p+q$ con $0 < \mu < 1$, $\phi > 0$. Se puede escribir $y \sim B(\mu, \phi)$. Por lo tanto, $E(y) = \mu$, $VAR(y) = \mu(1-\mu)/(1+\phi)$.
- El parámetro ϕ se conoce como parámetro de precisión, ya que para μ fijo, cuanto más grande es ϕ más pequeña es la varianza de y ; ϕ^{-1} es un parámetro de dispersión.

$$f(y; \mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu\phi-1} (1-y)^{(1-\mu)\phi-1}, \quad 0 < y < 1, \quad (3)$$

Diferentes modelos de distribución BETA

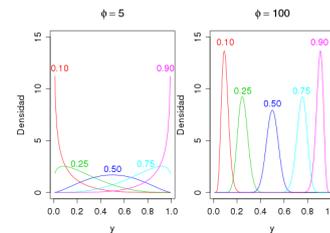


Figura : densidades BETA en el intervalo (0,1) al cambiar μ y ϕ

Formulación del modelo predictivo

A partir de los trabajos de Salinas et al [Salinas-Rodríguez et al., 2009] podemos decir que

- Si se tiene y_1, \dots, y_n una muestra aleatoria tal que $y_i \sim B(\mu_i, \phi)$, $i = 1, \dots, n$.
- El **modelo de regresión BETA (MRB)** es $\mu_i = g^{-1}(x_i^T \beta)$ donde β , el vector de parámetros de regresión, se estima por máxima verosimilitud (ML).

En varias ocasiones el parámetro de precisión no es constante a través de todas las observaciones, con lo cual es necesario modelarlo, tal cual se hizo con la **media**. En particular $y_i \sim B(\mu_i, \phi_i)$, $i = 1, \dots, n$, y

$$\begin{aligned} g_1(\mu_i) &= \eta_{1i} = x_i^T \beta, \\ g_2(\phi_i) &= \eta_{2i} = z_i^T \gamma, \end{aligned} \quad (4)$$

donde $\beta = (\beta_1, \dots, \beta_k)^T$, $\gamma = (\gamma_1, \dots, \gamma_h)^T$, $k+h < n$, son los **coeficientes de regresión** de ambas ecuaciones, η_{1i} and η_{2i} son predictores lineales, x_i y z_i son los vectores de regresión, los que se estiman por (ML). La función de **link** puede ser **logit, probit, log(log)**

Resultados Preliminares

Modelos de Regresión del CPO transformado en tasa (MRL)

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.5084	0.0217	23.41	0.0000
edad.rec.De 35 a 44	0.1446	0.0292	4.95	0.0000
edad.rec.Mayor a 45	0.2401	0.0316	7.59	0.0000
zona2	-0.0336	0.0328	-1.02	0.3063
zona3	-0.0607	0.0279	-2.18	0.0303

Modelos de Regresión del CPO transformado en tasa (MRB) (mean model with logit link)

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.04452	0.09685	0.460	0.6458
edad.rec.De 35 a 44	0.65666	0.13163	4.989	6.08e-07
edad.rec.Mayor a 45	1.17688	0.14708	8.002	1.23e-15
zona2	-0.12570	0.14783	-0.850	0.3952
zona3	-0.22743	0.12530	-1.815	0.0695

Se supone parámetro de precisión ϕ constante

	Estimate	Std. Error	t value	Pr(> t)
(ϕ)	3.1688	0.2297	13.79	2e-16 ***

Varias alternativas metodológicas a seguir

- Para el **CPO** el (MRL) parece funcionar adecuadamente. Hay que probar otras variables explicativas
- Sin embargo esta población es muy especial, ya que muestra la distribución del **CPO** una 'simetría' que solo se explica al ser esta una población muy enferma
- Existe a su vez el problema de que al expresar las tasas (para los componentes del CPO) aparecen muchos 0 y 1
- Se resuelve mediante la transformación de *Smithson-Verkulien*

$$y_i^* = \frac{y_i(n-1) + 0.5}{n}$$

- Conceptualmente para poder estimar los (MRB) no es correcto a priori sustituir los 0 y 1
- Por lo tanto es importante tratar de trabajar con los modelos que consideran eso, que se denominan *zero inflated models*
- Otro camino es seguir con la extensión de esta clase de modelos **BETA** [Grün et al., 2011], donde se consideran
 - Corrección y reducción del sesgo
 - Estimación de mejores modelos mediante árboles de regresión (Combinación de los métodos CART y los (MRB))
 - (MRB) con variable de clase latente que explican mejor la *heterogeneidad*

Referencias Bibliográficas

- Cribari-Neto, F. and Zeileis, A. (2010). Beta regression in r. *Journal of Statistical Software*, 34(2):1-24.
- Grün, B., Kosmidis, I., and Zeileis, A. (2011). Extended beta regression in R: Shrink, smooth, mixed, and partitioned. Working Paper 2011-22, Working Papers in Economics and Statistics, Research Platform Empirical and Experimental Economics, Universität Innsbruck.
- Salinas-Rodríguez, A., Manrique-Espinoza, B., and Sosa-Rubi, S. G. (2009). Análisis estadístico para datos de conteo: aplicaciones para el uso de los servicios de salud. *Salud Pública de México*, 51:397-406.