

## Análisis Descriptivo de los indicadores del CMBD del Sistema Nacional de Salud (SNS).

Dr. José Manuel Gutiérrez  
*Instituto de Física de Cantabria (IFCA)*  
*CSIC - Universidad de Cantabria (UC)*  
*gutierjm@unican.es*

Dr. Antonio S. Cofiño  
*Dpto. de Matemática Aplicada y C.C.*  
*Universidad de Cantabria (UC)*  
*cofinoa@unican.es*

### CONTENIDOS:

1. Introducción
2. Estancia Media
3. Estancia Media Preoperatoria
4. Indicadores en Forma de Tasa
5. Eliminación de Hospitales con Valor Fuera de Rango

**RESUMEN:** En el presente informe se proporciona un análisis descriptivo de los indicadores, analizando sus distribuciones y la conveniencia de realizar transformaciones adecuadas (por ejemplo, para normalizar la variable) con el objeto de mejorar la significación de los estudios estadísticos posteriores. Los resultados de este informe son relativos a la población correspondiente a los años 2005 y 2007, que han sido elegidos como años ilustrativos del período 2001-2007, con un total de 3.541.107 y 3.686.655 altas, respectivamente. En este informe se muestra que las conclusiones son robustas para los dos años analizados, no observándose ningún cambio significativo en la distribución de los factores.

## 1. Introducción

El registro de altas hospitalarias, también conocido como CMBD (Conjunto Mínimo Básico de Datos al alta), constituye la mayor base de datos administrativa sobre

- |  |
|--|
| <ol style="list-style-type: none"> <li>1. Estancia Media (1)</li> <li>2. Estancia Media Preoperatoria (1)</li> <li>3. Tasa de Mortalidad (17)</li> <li>4. Tasa de Reingresos (1)</li> <li>5. Tasa de Infección Nosocomial (1)</li> <li>6. Tasa de Cesáreas (1)</li> <li>7. Tasa de Complicaciones (15)</li> <li>8. Tasa de Ambulatorización Quirúrgica (2)</li> <li>9. Frecuentación en Hospitalización (1)</li> <li>10. Tasa de Realización (11)</li> </ol> |
|--|

**Tabla 1. Conjunto de indicadores**

pacientes hospitalizados (casi 27 millones de registros, a razón de 3,5 millones/año aprox.), siendo la fuente principal de información sobre morbilidad atendida, con información muy valiosa sobre múltiples aspectos de la actividad hospitalaria, incluyendo la calidad y variabilidad de la práctica asistencial.

Esta base de datos recopila la información de 283 hospitales del Sistema Nacional de Salud (SNS).

El Ministerio de Sanidad y Política Social es responsable de la gestión del CMBD estatal, generando con periodicidad anual diversas estadísticas oficiales. Recientemente, se ha desarrollado un modelo de explotación de esta información basado en un conjunto reducido de indicadores de máximo valor explicativo que permita profundizar en el análisis de las características de la atención hospitalaria de los pacientes ingresados en el Sistema Nacional de Salud. Este modelo se basa en 51 indicadores, agrupados en diez familias genéricas mostradas en la Tabla 1 (el número de indicadores en cada familia se muestra entre paréntesis; ver documento de "*descripción del modelos de indicadores*"). Algunos de estos indicadores se refieren a resultados continuos (número de días), mientras que otros se refieren a resultados discretos (ocurrencias), dando lugar a indicadores en forma de promedio o tasa, respectivamente.

Para poder comparar el valor de los indicadores en los distintos hospitales primero es necesario corregir, o ajustar, los resultados brutos teniendo en cuenta las distintas casuísticas de los pacientes tratados en los distintos hospitales (case-mix), considerando así la complejidad de los servicios prestados en cada caso (ver *documentación técnica sobre metodología de ajuste para más detalles*).

En el presente informe se realiza un análisis descriptivo de estos indicadores, analizando sus distribuciones y eliminando casos anómalos para preparar los datos (conjunto de altas) de cara a su uso para el ajuste de los indicadores (ver *documentación técnica "metodología de ajuste de los indicadores"*). También se analiza la conveniencia de transformar alguno de ellos para facilitar y simplificar su tratamiento estadístico, su modelización y su ajuste.

## 2. Estancia Media

La Estancia Media (EM) se obtiene como el valor medio de la *estancia* (diferencia entre la fecha de alta y de ingreso, medida en días). La estancia es una variable discreta con valor mínimo cero y máximo 999 para las altas del año 2007 (984 en el 2005). El valor medio de la *estancia* en la población total de altas es 7,4849 (7,5874 en el 2005). La *estancia* puede ser considerada como una variable discreta de forma natural pues sólo puede tomar un número finito de valores; sin embargo, debido a su

elevado número de estados, esta variable también puede tratarse como continua. El problema es que las frecuencias absolutas de los sucesivos días de estancia van decayendo rápidamente (de forma potencial, como muestra la siguiente figura). Por ejemplo, en el año 2007 las estancias superiores a un año son sólo 174. Un comportamiento idéntico se da también con las estancias del resto de años.

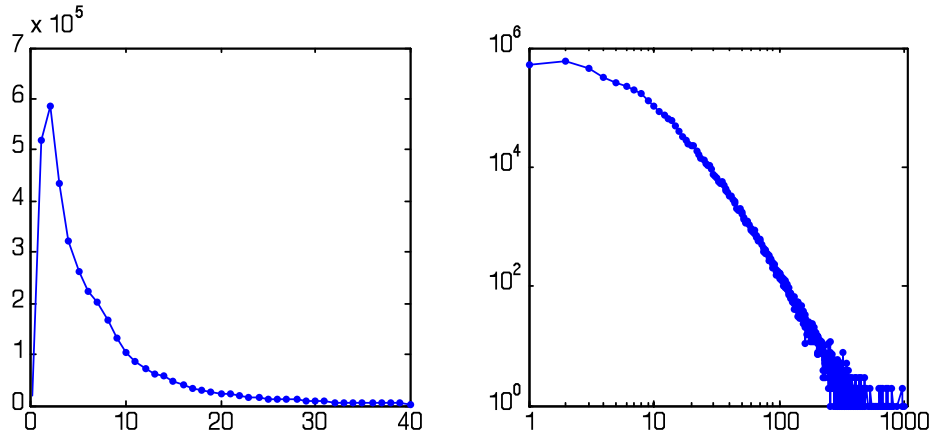


Figura 1. (izquierda) Diagrama de frecuencias absolutas de valores de estancia ordenados. (derecha) Diagrama logarítmico, mostrando una relación potencial (altas del 2007).

Como en el caso de la variable de influencia *complejidad* (ver documentación técnica “análisis descriptivo de variables de influencia”), una opción para resolver este problema es considerar una transformación potencial de Box-Cox para normalizar la distribución de la variable. En este caso se aplica la Eq. (1) a la variable  $EM+1$  considerando las altas del 2007 (obsérvese que la transformación de Box-Cox requiere que los datos sean positivos; por ello, se traslada la estancia una unidad para satisfacer este requerimiento), ajustando el parámetro óptimo por el método de máxima verosimilitud, resultando un valor  $\lambda=-0,29$ , que permite transformar la distribución de la variable en base a la ecuación.

$$TEM = -((EM+1)^{-0,29} - 1) / 0,29, \tag{1}$$

que relaciona la variable transformada ( $TEM$ ) y la original ( $EM$ ).

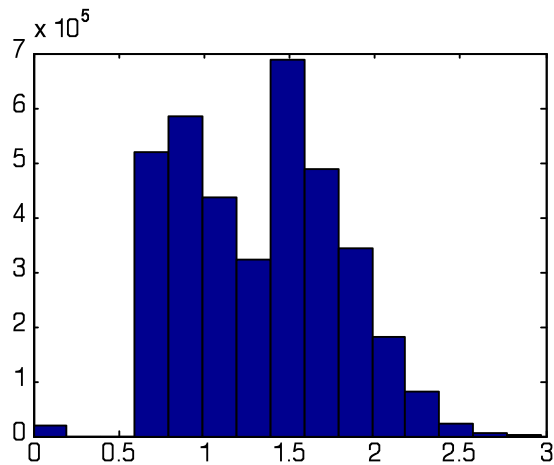


Figura 2. Histograma de las frecuencias absolutas de  $TEM$ .

Como puede observarse en la figura anterior, la transformación hace que los valores altos en la cola derecha de la distribución de EM se agrupen en valores cercanos a 2,5 para la nueva variable. Sin embargo, la nueva variable no supera un test de normalidad (Kolmogorov-Smirnov) a ningún nivel de significación razonable (también puede observarse visualmente la clara bimodalidad de la distribución resultante).

Por tanto, se procedió a discretizar la variable *estancia media* para poder llevar a cabo los estudios de asociación con los distintos factores considerando un modelo estadístico discreto. Para ello se han calculado los centiles de la población, resultado los siguientes valores (omitiendo los repetidos): 0 (mínimo), 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16, 19, 22, 27, 36, máximo. Por tanto, se puede categorizar la variable *estancia* considerando 21 categorías cuyos límites inferior y superior vendrían dados por cada par sucesivo de los valores anteriores y posteriores, respectivamente. El histograma de la variable resultante se muestra a continuación:

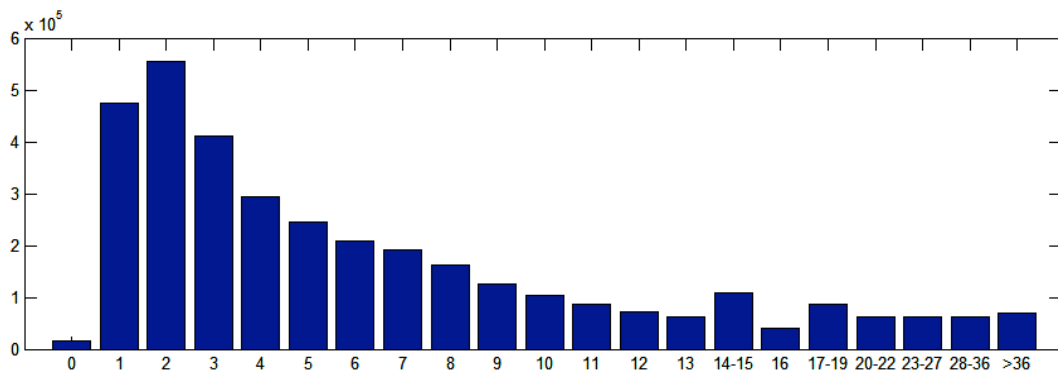


Figura 3. Histograma de la estancia media discretizada en 21 categorías.

Como marca, o valor representativo de cada clase, se considera la media de los valores agrupados respectivos:

Clase	0	1	2	3	4	5	6	7	8	9	10
Marca	0	1	2	3	4	5	6	7	8	9	10
Clase	11	12	13	14-15	16	17-19	20-22	23-27	28-36	>36	
Marca	11	12	13	14,451	16	17,903	20,944	24,785	31,359	60,140	

### 3. Estancia Media Preoperatoria

La Estancia Media Preoperatoria (EMP) se trata de la misma forma que la estancia media, discretizándola en un número apropiado de clases. Sin embargo, en este caso, la población total se reduce a aquella que se ha sometido a algún tratamiento quirúrgico (altas con GRD de tipo quirúrgico y fecha de intervención válida en el registro del CMBD: 911.573 altas en el año 2007). El valor medio de la *EMP* en esta población es 2,06 y su distribución de frecuencias presenta las mismas características que el caso de estancia media analizado en la sección anterior.

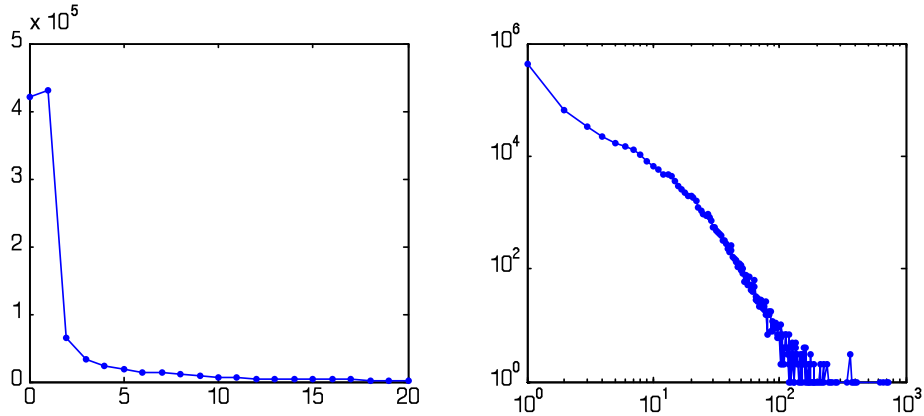


Figura 4. (izquierda) Diagrama de frecuencias absolutas de valores de estancia preoperatoria ordenados. (derecha) Diagrama logarítmico (altas del 2007).

Por tanto, esta variable puede ser considerada como una variable discreta de forma natural (obsérvese que la estancia se mide en número de días); para ello es conveniente reagrupar la variable en un número menor de categorías, por ejemplo, usando los centiles de la población (omitiendo los repetidos): 0 (mínimo), 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 14, 18, 24, máximo, se optó por categorizar la variable *EMP* considerando 7 clases cuyos límites inferior y superior y marcas utilizadas en los cálculos vienen dados en la siguiente tabla

Clase	0	1	2-4	5-8	9-14	15-24	>24
Marca	0	1	2,6545	6,3189	11,1265	18,5399	39,2823

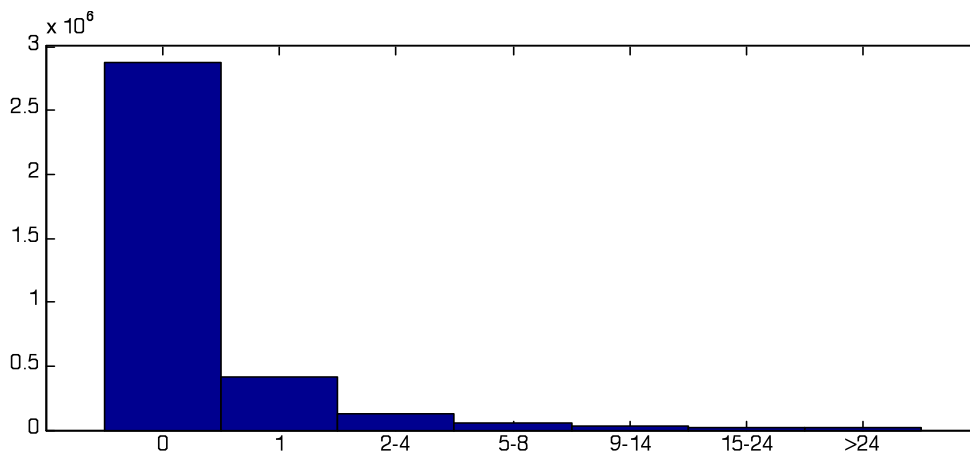


Figura 5. Histograma de la estancia media preoperatoria discretizada en 7 categorías

#### 4. Indicadores en Forma de Tasa

En el caso de los indicadores en forma de tasa (del 3 al 10 en la Tabla 1), no se requiere ningún tratamiento especial, ya que la variable ocurrencia asociada es una variable binaria que puede tomar solamente dos estados distintos. Los estadísticos descriptivos básicos de estos indicadores se pueden consultar directamente en la aplicación en el *eje de análisis "descriptivo básico"*.

#### 5. Eliminación de Hospitales con Valor fuera de Rango (Outliers)

En el análisis estadístico que se lleva a cabo para el ajuste de los valores del indicador en los distintos hospitales (*ver documentación técnica sobre metodología de ajuste para más detalles*), se ha trabajado a nivel de hospital, agregando los resultados de los pacientes tratados. En este caso, ha sido necesario llevar a cabo un control de calidad adicional, estudiando la distribución de los valores del indicador en los distintos hospitales (esta distribución se muestra, para cada indicador, en el eje de análisis "*ajuste de mejores resultados*"), y obteniendo los valores fuera de rango (u *outliers*) que corresponden a hospitales con un comportamiento atípico, muy alejado de los resultados promedio de los hospitales. Para una distribución arbitraria se suelen definir datos atípicos (*outliers*) desde 1,5 a 3 RI (donde RI es el Rango Intercuartílico) y datos atípicos extremos a partir de 3 RI. Dado que la distribución del indicador por hospitales suele tener una cola derecha pesada (*ver, por ejemplo, la Figura 1*), en este trabajo se consideró como valor umbral 3 RI para no descartar datos atípicos sino datos realmente extremos. Por tanto, aquellos hospitales con resultados superiores fueron descartados del análisis estadístico de mejores resultados, para no introducir un sesgo en el cálculo. Este cálculo se realizó independientemente para cada indicador analizado.

Los resultados de cada indicador en cada hospital se muestran en la aplicación, en la *tabla de corrección individual por hospitales* del eje de análisis "*ajuste de mejores resultados*", que muestra los resultados individuales para aquellos hospitales analizados en cada indicador; es decir, aquellos con un valor del indicador por debajo del umbral antes descrito (obsérvese que, dependiendo del nivel de autorización del usuario, los hospitales pueden figurar de forma anónima; Hospital 1, Hospital 2, ...).