

Análisis Descriptivo de las Variables de Influencia para los indicadores del CMBD del 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. Análisis Descriptivo
 - 2.1. Complejidad
 - 2.2. Severidad
 - 2.3. Riesgo de Mortalidad (ROM)
 - 2.4. Categoría Diagnóstica Mayor (CDM)
 - 2.5. Tipo de GRD
 - 2.6. Edad
 - 2.7. Sexo
 - 2.8. Tipo de Ingreso
 - 2.9. Tipo de Alta
 - 2.10. Tipo de Hospital
 - 2.11. Edad de la Madre
3. Conclusiones

RESUMEN: En el presente informe se analizan las variables que han sido consideradas como posibles factores de influencia para los indicadores analizados. Este estudio proporcionará un conocimiento descriptivo de sus características, analizando sus distribuciones y la conveniencia de realizar transformaciones adecuadas con el objeto de mejorar la significación de los estudios estadísticos posteriores, así como facilitar la construcción de modelos apropiados para el estudio de dependencias entre variables de influencia e indicadores. 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

1. Estancia Media (1)
2. Estancia Media Preoperatoria (1)
3. Tasa de Mortalidad (17)
4. Tasa de Reingresos (1)
5. Tasa de Infección Nosocomial (1)
6. Tasa de Cesáreas (1)
7. Tasa de Complicaciones (15)
8. Tasa de Ambulatorización Quirúrgica (2)
9. Frecuentación en Hospitalización (1)
10. Tasa de Realización (11)

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 modelo de indicadores*" para más información). Algunos de estos indicadores se refieren a resultados continuos, como número de días, mientras que otros se refieren a resultados discretos, como ocurrencias.

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

RELACIONADOS CON LA ENFERMEDAD

1. Complejidad (peso español GRD-AP v18)
2. Severidad (GRD refinados)
3. Riesgo de Mortalidad, ROM
4. Cat. Diagnóstica Mayor (GRD-AP v18)
5. Tipo de GRD: médico o quirúrgico

PACIENTE / HOSPITAL

6. Edad
7. Sexo
8. Tipo de ingreso
9. Tipo de alta
10. Tipo de hospital
11. Edad de la madre
(sólo para Tasa de Cesáreas)

Tabla 2. Factores de riesgo/influencia

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*).

El CMBD incluye un conjunto genérico de variables de influencia, o factores de riesgo, que influyen en esta casuística y están relacionados tanto con la enfermedad y su diagnóstico (basados en GRDs), como con el paciente y con el funcionamiento hospitalario (Tabla 2).

Con la excepción de la complejidad, todas las variables de influencia son discretas, o se utilizan en una variante

discretizada (como los grupos de edad).

En el presente informe se realiza un análisis descriptivo de las variables de influencia, 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"*).

2. Análisis Descriptivo

2.1. COMPLEJIDAD

Esta variable de influencia está medida a través del peso español. Se trata de una variable continua positiva cuya distribución tiene un gran sesgo y apuntamiento (ver Fig. 1, la fila inferior es una ampliación de la zona donde se concentran los valores de complejidad). Esta figura muestra que la distribución de la variable es muy estable para los distintos años, siendo las diferencias casi imperceptibles a simple vista.

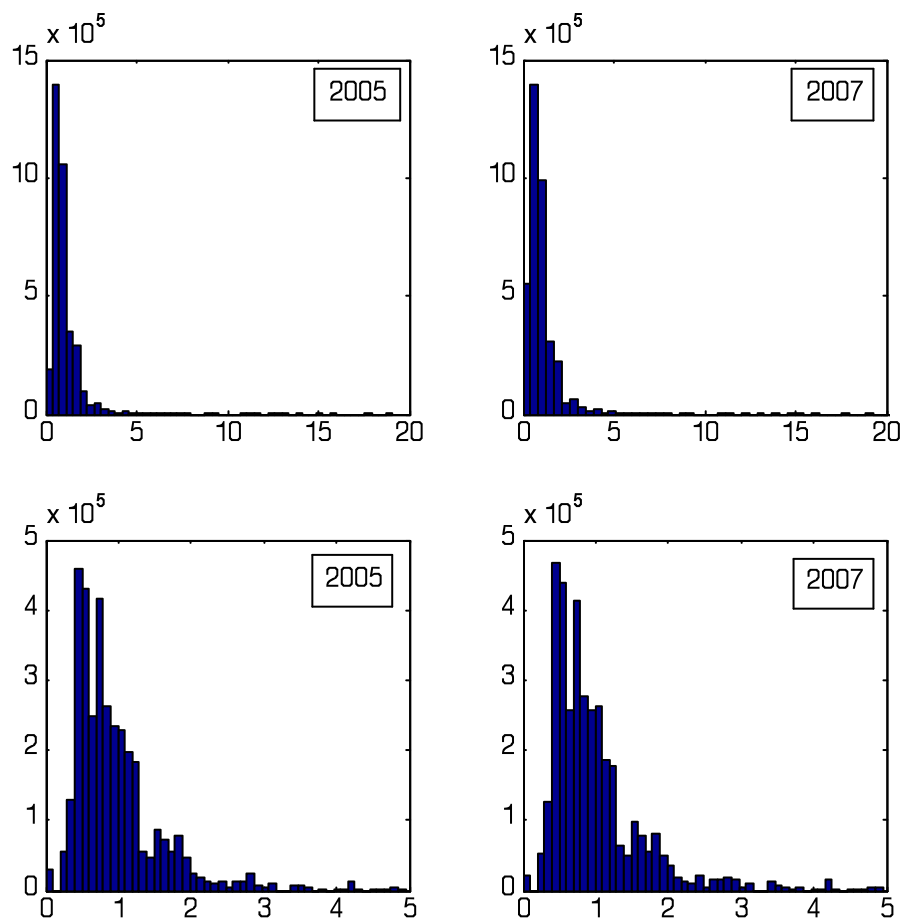


Figura 1. Frecuencias absolutas de la variable *complejidad* para los años 2005 (izquierda) y 2007 (derecha).

En estos casos es conveniente transformar la variable original para que su distribución se aproxime mejor a la normal. En este trabajo se ha aplicado la transformación potencial de Box-Cox:

$$T(X) = Y = \begin{cases} \frac{X^\lambda - 1}{\lambda} & \lambda \neq 0 \\ \ln X & \lambda = 0 \end{cases} \quad (1)$$

El parámetro λ es estimado por máxima verosimilitud, $\lambda = -1,27$ (2005) y $\lambda = -1,34$ (2007), que permite transformar la variable aplicando la Ec. (1) dando lugar a las distribuciones que se muestran en la Figura 2. A simple vista se puede comprobar que la transformada normalizada conserva reminiscencias del carácter bimodal de la distribución original y, por tanto, no es una buena aproximación normal.

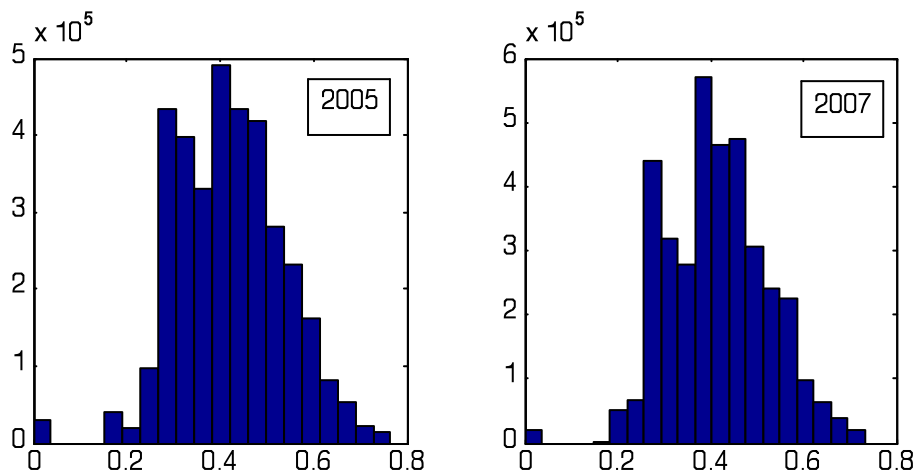


Figura 2. Frecuencias de la variable transformada por Box-Cox.

Para comprobar estadísticamente la validez de la hipótesis de normalidad de la variable resultante, se procedió a realizar un test de Kolmogorov-Smirnov, resultando un rechazo de la hipótesis de ajuste a la distribución normal con una confianza del 99.9%; por tanto, la aproximación normal no es válida en este caso.

Dado que la aproximación normal no ofrece resultados satisfactorios, se procede a discretizar la variable de forma que se pierda la menor información posible y se pueda llevar a cabo los estudios de asociación con los distintos indicadores considerando un modelo estadístico discreto. Para ello se consideran clases tomando los percentiles de la población como puntos de corte. Se consideraron 10 clases en base a los deciles, resultando los puntos de corte:

0, 0,4100, 0,5097, 0,5819, 0,7128, 0,7991, 0,9285, 1,0754, 1,2326, 1,7827, 19,0786

En base a estos valores se obtiene una variable discreta *complejidad* con diez estados sobre los cuales se tiene una distribución uniforme de frecuencias; se tomó este número de estados para que existiese similitud con el resto de factores del estudio. Como marca o valor representativo de cada una de las categorías se considerará la media de los valores agrupados (dado que los valores son similares en los distintos años, se considera la serie del 2007 para definir los intervalos y las marcas de las clases, como se muestra en la siguiente tabla):

Clase	1	2	3	4	5	6	7	8	9	10
marca	0,343	0,416	0,551	0,672	0,795	0,934	1,076	1,264	1,696	4,015

2.2. SEVERIDAD

Variable discreta con cinco valores posibles correspondientes a 1:"No se especifica clase", 2:"Menor", 3:"Moderado", 4:"Mayor" y 5:"Extremo", respectivamente, con frecuencias absolutas 19.822, 1.988.520, 1.137.257, 461.261, 79.795 (27.899, 1.767.494, 1.188.923, 460.341, 96.450 para 2005). En los análisis posteriores de variables de influencia y mejores resultados (*ver documentación técnica*) no se considera el caso "no se especifica clase" (19.822 altas en el 2007).

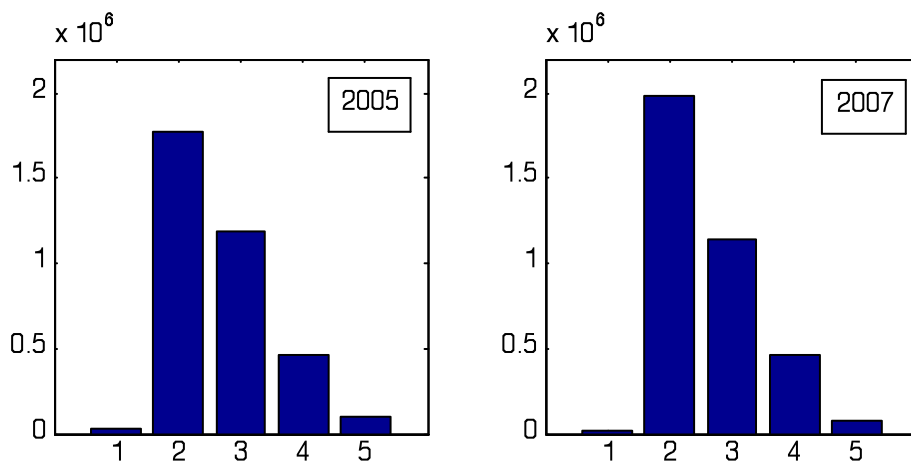


Figura 3. Frecuencias absolutas de la severidad.

2.3. RIESGO DE MORTALIDAD (ROM)

Variable discreta con cinco estados posibles correspondientes a 1:"No se especifica clase", 2:"Menor", 3:"Moderado", 4:"Mayor" y 5:"Extremo", respectivamente, con frecuencias absolutas 19.822, 2.591.790, 705.771, 287.941, 81.331 (27.899, 2.527.657, 550.504, 306.167, 128.880 para 2005). En los análisis posteriores de variables de influencia y mejores resultados (*ver documentación técnica*) no se considera el caso "no se especifica clase" (19.822 altas en el 2007).

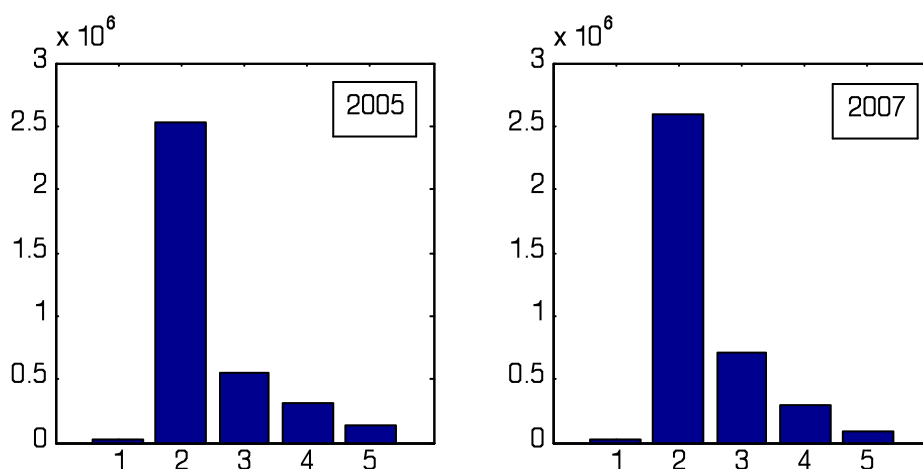


Figura 4. Frecuencias absolutas del Riesgo de Mortalidad (ROM).

2.4. CATEGORÍA DIAGNÓSTICA MAYOR (CDM)

Variable discreta con 26 estados distintos (incluyendo la categoría no válida, denotada por 0). Cada uno de estos estados corresponde a un grupo de enfermedades o trastornos y no guardan ningún orden entre sí:

- | | |
|--|---|
| 0 Pre-CDM | 12 Enfermedades y trast. del aparato reproductor masculino |
| 1 Enfermedades y trastornos del sistema nervioso | 13 Enf. y trastornos del aparato reproductor femenino |
| 2 Enfermedades y trastornos del ojo | 14 Embarazo, parto y puerperio |
| 3 Enfermedades y trast. de oído, nariz, boca y garganta | 15 Recién nacidos y neonatos con condición Roig. en periodo perinatal |
| 4 Enfermedades y trastornos del aparato respiratorio | 16 Enf. y trast. de sangre, órganos hematopoy. y sist. inmunitarios |
| 5 Enfermedades y trastornos del aparato circulatorio | 17 Enf. y trast. mieloproliferativos y neoplasias mal diferenciadas |
| 6 Enfermedades y trastornos del aparato digestivo | 18 Enf. infecciosas y parasitarias, sistémicas o no localizadas |
| 7 Enf. y trast. de hígado, sistema biliar y páncreas | 19 Enfermedades y trastornos mentales |
| 8 Enf. y trast. sistema musculoesquelético y tejido conectivo | 20 Uso drogas/alcohol y trast. ment. induc. por drogas/alcohol |
| 9 Enf. y tras. de piel, tejido subcutáneo y mama | 21 Lesiones, envenenamientos y efectos tóxicos de fármacos |
| 10 Enf. y trast. del sist. endocrino, nutricional y metabólico | 22 Quemaduras |
| 11 Enfermedades y trastornos de riñón y vías urinarias | 23 Otras causas de atención sanitaria |
| | 24 Infecciones VIH |
| | 25 Trauma múltiple significativo |

La siguiente figura muestra un histograma de las frecuencias absolutas de cada una de las categorías en los años 2005 y 2007. Como puede observarse, la distribución de CDMs es muy estable para distintos años.

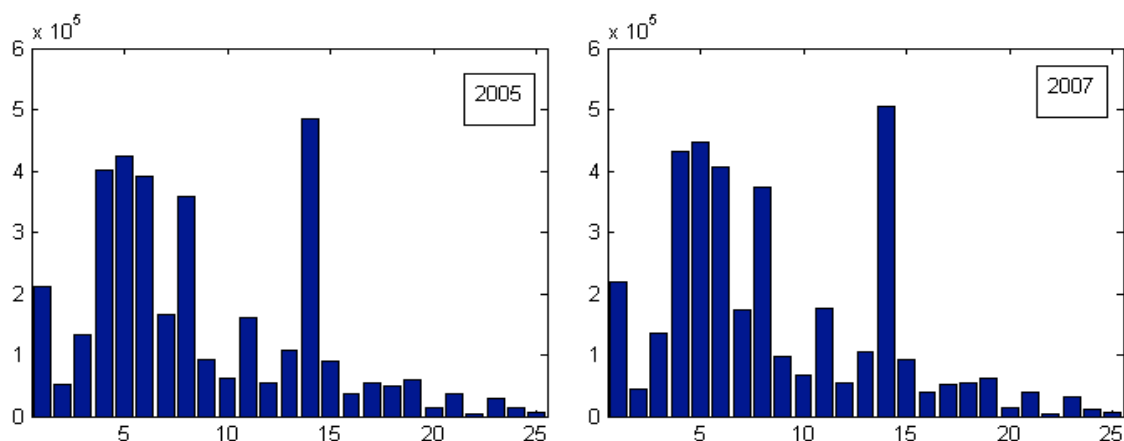


Figura 5. Frecuencias absolutas de la categoría diagnóstica mayor (CDM).

2.5. TIPO DE GRD

Variable discreta con tres estados posibles correspondientes a 1: "No se especifica clase", 2: "Médica", 3: "Quirúrgica", respectivamente, con frecuencias absolutas 20.024, 2.355.879, 1.310.752 (28.244, 2.249.595, 1.263.268 para 2005). En los análisis posteriores de variables de influencia y mejores resultados (*ver documentación técnica*) no se considera el caso "no se especifica clase" (20.024 altas en el 2007).

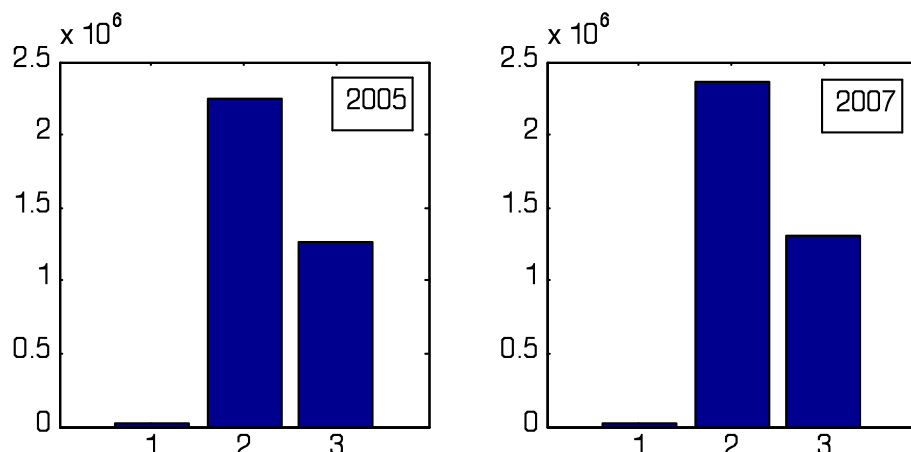


Figura 6. Frecuencias absolutas del tipo de GRD.

2.6. EDAD

La variable de influencia *edad* se considera en su variante categorizada, considerando 6 estados posibles ([0,1), [1,15), [15,45), [45,65), [65,75), >75). Las frecuencias absolutas de estos casos son 154.554, 210.602, 1.088.931, 763.410, 566.358, 902.800, para el año 2007. Los porcentajes de incremento poblacional en las distintas categorías respecto del año 2005 fueron: 1,42%, 2,59%, 2,05%, 5,48%, -2,26%, y 11,02%. Como hecho más destacado, la mayor diferencia entre los dos años se da en el crecimiento de altas asociadas al mayor umbral de edad (>75).

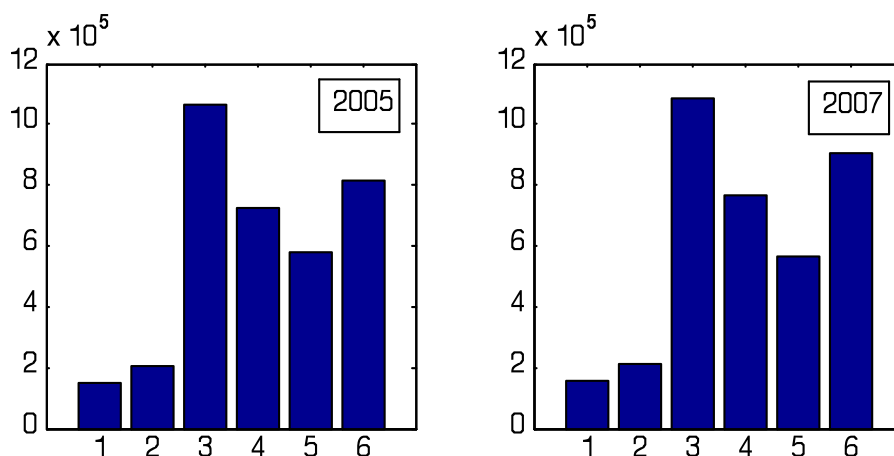


Figura 7. Frecuencias absolutas del factor *edad*.

La estructura bimodal de esta distribución de edades tiene su origen en la distribución de las edades originales, que presenta distintos picos en torno a los recién nacidos, a los 30, a los 75 y un pequeño modo en torno a los 60 años. De hecho la distribución de edades podría ser aproximada por la suma de cuatro distribuciones gaussianas

centradas en las edades indicadas (ver Figura 2.4). Por tanto, las categorías de edades consideradas parecen adecuadas para representar las distintas distribuciones suyacentes.

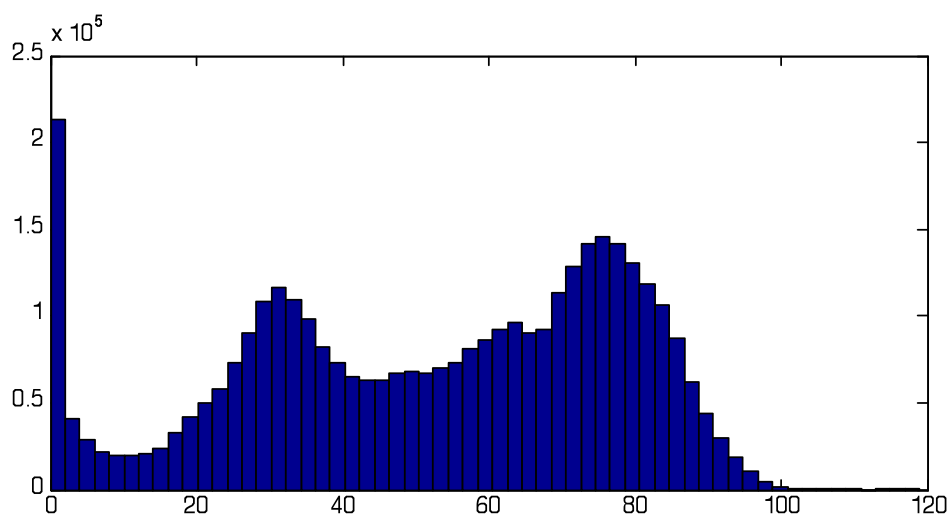


Figura 8. Frecuencias de edades del año 2007 (las altas con edad cero son 154554).

2.7. SEXO

Variable discreta con cuatro categorías (Varón, Mujer, Indeterminado, y Otros) con frecuencias absolutas: 1.720.559, 1.965.820, 125, 151, respectivamente (1.653.665, 1.887.352, 75, 15 en el 2004). En los estudios posteriores sólo se consideran los estados de sexo *varón* y *mujer*.

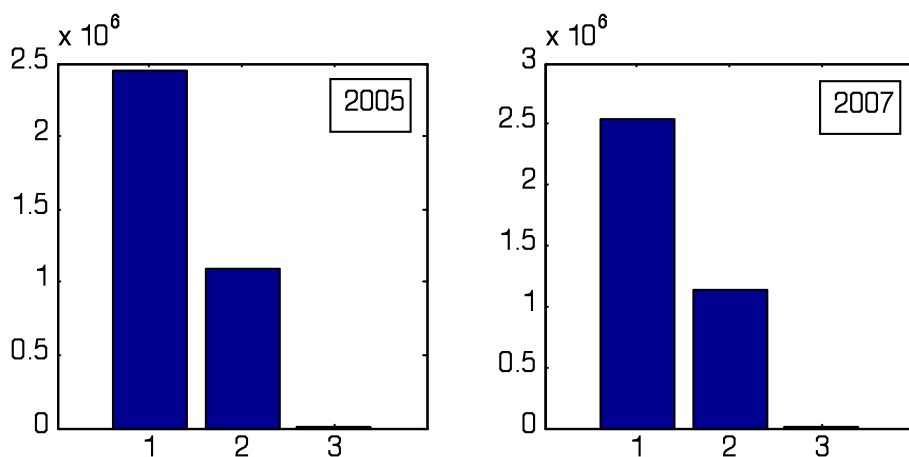


Figura 9. Frecuencias del factor sexo.

2.8. TIPO DE INGRESO

Variable discreta con tres estados (programado, urgente y otros). Las frecuencias absolutas de estos casos para el año 2007 son 2.545.075, 1.129.349, 12.231 (las correspondientes al año 2005 son 2.446.146, 1.085.680, 9.281). Como se puede

apreciar, las frecuencias de los dos años analizados son muy similares. En los análisis posteriores no se considera el último estado de la variable “otros” (12.231 altas en el año 2007).

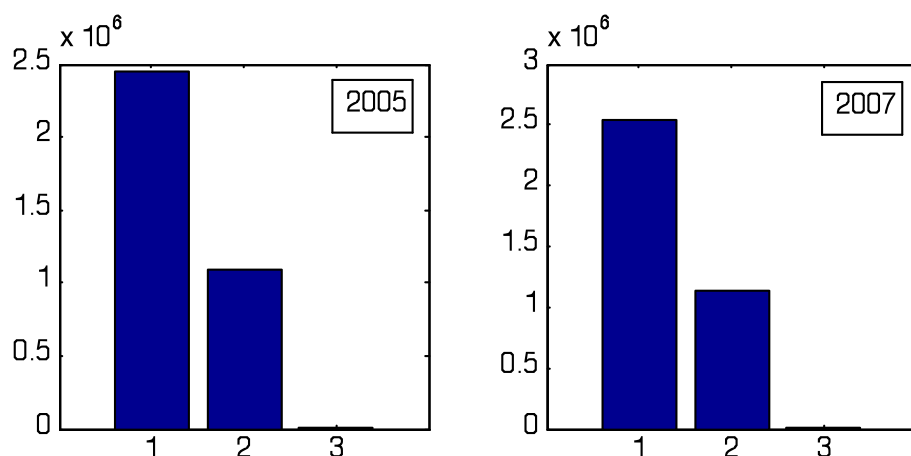


Figura 10. Frecuencias del factor *tipo de ingreso*.

2.9. TIPO DE ALTA

Variable discreta con seis estados (Domicilio, Traslado a otro Hospital, Alta voluntaria, Exitus, Traslado a centro sociosanitario, Otros). Las frecuencias absolutas son 3.404.837, 74.066, 19.872, 151.948, 22.297, 13.635, respectivamente (3.272.419, 81.734, 19.608, 145.879, 3.594, 17.873, para el año 2005). En los siguientes estudios no se considera el último estado de la variable “otros” (13.635 altas en el año 2007).

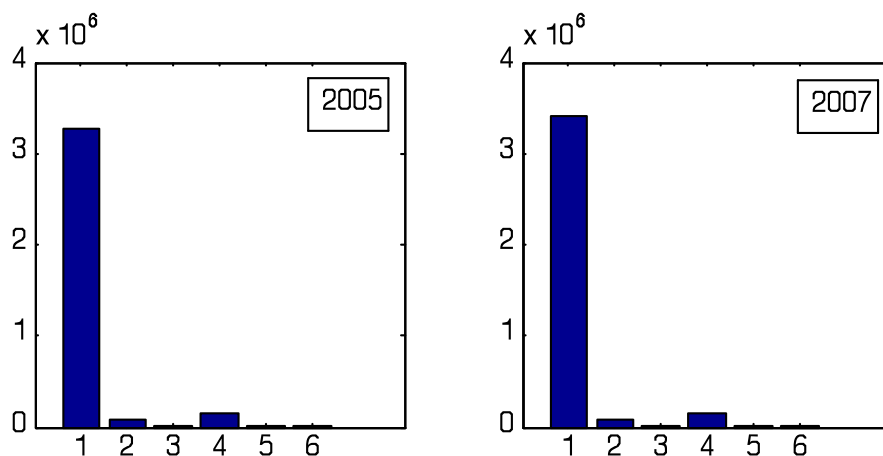


Figura 11. Frecuencias absolutas del factor *tipo de alta*.

2.10. TIPO DE HOSPITAL

Para describir la variable tipo de hospital se utilizan dos criterios diferentes: una clasificación basada en número de camas (grupo de hospital) y una agrupación de los hospitales en clases (cluster de hospital); ver *documentación técnica “resumen cluster hospitales”* para este segundo criterio.

El grupo de hospital es una variable discreta con cinco estados, correspondiente al número de camas del hospital (<200, [200,500), [500-1.000], >1.000). Las frecuencias absolutas de las altas del año 2007 según cada grupo de hospital fueron 590.853, 1.125.599, 1.062.872, 907.331 (529.469, 1.078.237, 1.031.160, 902.241 en el 2005). El total de hospitales registrados en el año 2007 es de 283, siendo 261 en el 2005.

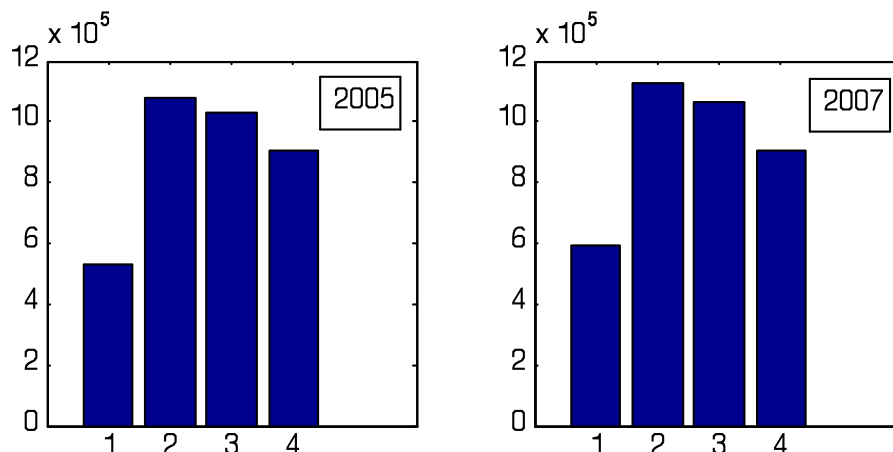


Figura 12. Frecuencias del factor *Grupo de Hospital*.

Además de la clasificación anterior en grupos de hospitales por número de camas, también se considera la nueva clasificación basada en un análisis de conglomerados realizada por el Instituto de Información Sanitaria del Ministerio de Sanidad y Política Social en la que se han obtenido de forma objetiva 6 grupos de hospitales homogéneos, marcados por un código numérico del 1 al 6. Cada uno de los hospitales está asociado a uno de estos grupos. A este factor lo denominaremos *Hospital Cluster* para diferenciarle del anterior.

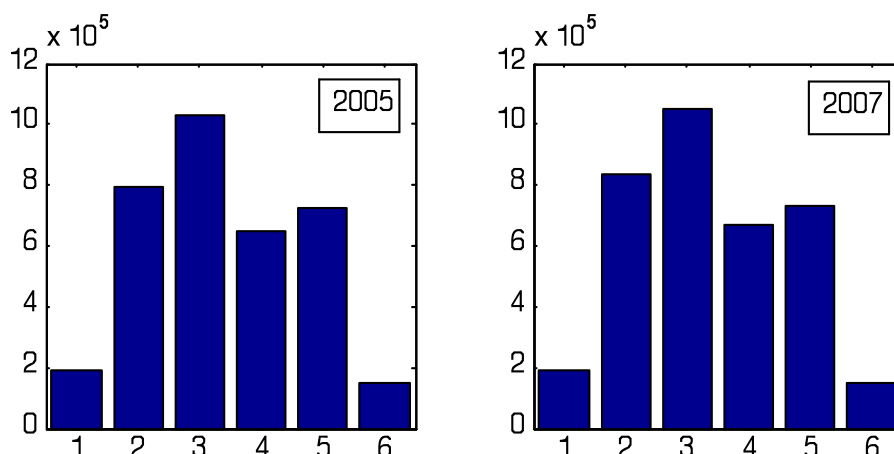


Figura 13. Frecuencias del factor *Hospital Cluster*.

2.11. EDAD DE LA MADRE

La edad de la madre es una variable discreta que ha sido agrupada en 11 categorías distintas: 1:0-14, 2:15-19, 3:20-24, 4:25-29, 5:30-34, 6:35-39, 7:40-44, 8:45-49, 9:50-54, 10:55-59, 11:>59. Las frecuencias absolutas de esta variable en aquellas altas

asociadas a partos en el año 2007 (un total de 388980 altas) son: 170, 14.504, 47.038, 99.325, 139.988, 73.986, 13.260, 670, 33, 5, 1.

Esta variable de influencia se utiliza únicamente para la modelización y ajuste del indicador *tasa de cesáreas*.

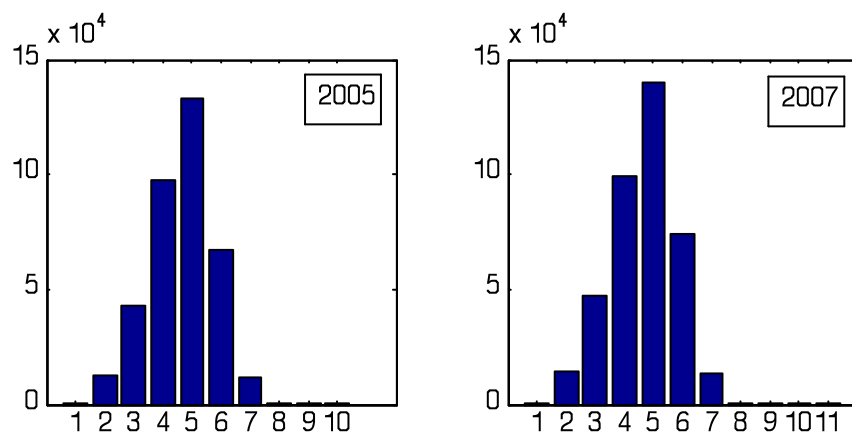


Figura 14. Frecuencias del factor *Edad de la Madre*.