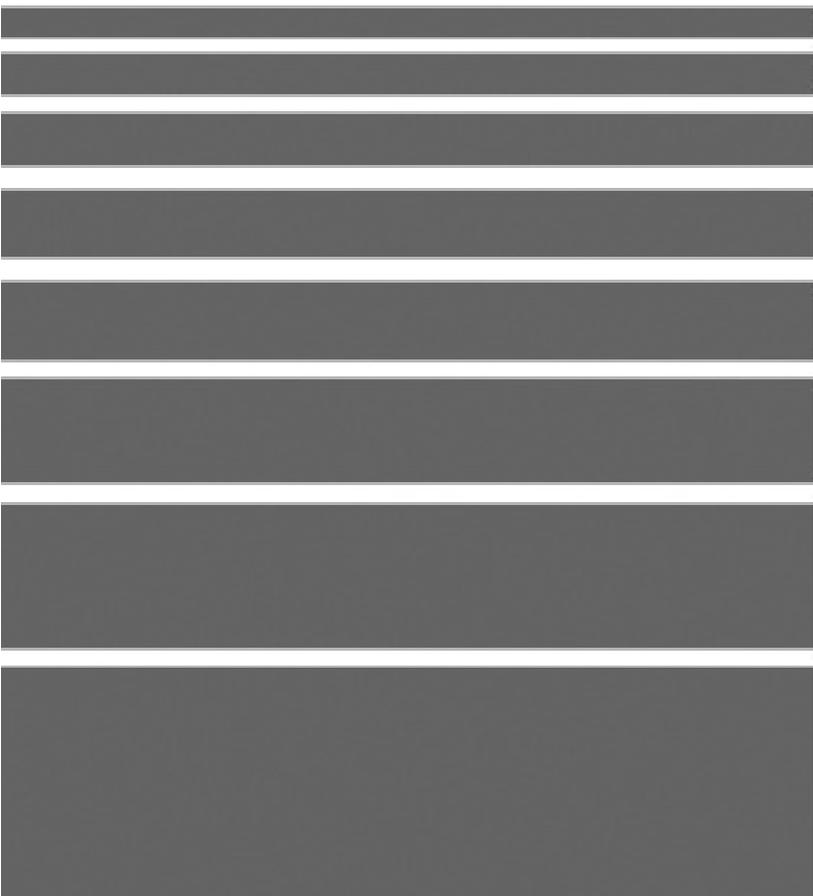




Encuesta de establecimientos hoteleros

Autores: Francisco Parra Rodríguez
Alberto Lezcano Lastra
Miguel Expósito Martín

DOC. Nº 1/2008
ISSN 2444 - 1627
Santander, Cantabria



ÍNDICE

I.- Introducción.....	2
2.- Muestra.....	3
3.- Fase1: Obtención de estimadores según ICANE.....	11
3.1.- Estimadores con la semana.....	13
3.2.- Estimadores con totales mensuales.....	14
3.3.- Tratamiento de la falta de respuesta.....	16
4.- Fase 2: Ajuste de los estimadores, para que coincidan con el INE.....	19
5.- Resultados obtenidos.....	28
7.- Implementación.....	30
Agradecimientos.....	32
Referencias.....	32

I.- INTRODUCCIÓN

El objetivo de esta investigación es obtener una valoración de la Encuesta de Ocupación Hotelera (EOH) del Instituto Nacional de Estadística (INE), a nivel comarcal en Cantabria, para lo cual se han obtenido estimadores de la muestra de la encuesta para las zonas turísticas de Cantabria así definidas por la Consejería de Turismo, Cultura y Deportes.

Los estimadores obtenidos permiten el estudio del turismo a nivel zonal bien por el lado de la demanda, ya que se estiman los viajeros, pernoctaciones y estancia media distribuidos por la categoría de los establecimientos que se ocupan y el país de residencia de los viajeros o Comunidad Autónoma de procedencia en el caso de los viajeros españoles; del lado de la oferta se proporciona el número de establecimientos abiertos estimados, plazas estimadas, y grados de ocupación según la categoría del establecimiento.

La obtención completa de los estimadores consta de dos fases. En la **Fase 1** se calculan los estimadores realizando una elevación estadística para cada una de las zonas turísticas de Cantabria. El resultado para Cantabria será la suma de los resultados obtenidos para cada zona. Como las zonas consideradas en ICANE son diferentes de las que contempla el INE, los resultados de ICANE y del INE difieren. En realidad, aunque considerásemos las mismas zonas que el INE, los resultados también serían diferentes, porque los cálculos detallados que realiza tanto el ICANE como el INE no son exactamente los mismos, y no hemos conseguido en ICANE reproducir el proceso de elevación del INE. Tan sólo hemos logrado, considerando las mismas zonas que el INE, obtener los mismos valores para los establecimientos abiertos y las plazas ofertadas, pero no para los viajeros ni para las pernoctaciones. Es por ello que necesitamos la **Fase 2**, que se encarga de realizar una modificación de los estimadores obtenidos al final de la Fase 1, para que el total para Cantabria concuerde con el INE.

2.- MUESTRA

El directorio de la EOH es el registro de establecimientos hoteleros de la Consejería de Turismo, Cultura y Deportes, que a nivel comarcal en Cantabria presenta los siguientes movimientos para los 6 primeros meses de 2008:

DIRECTORIO DE TURISMO						
ZONA_TURISTICA	2008					
	1	2	3	4	5	6
ASÓN-AGÜERA	49	49	47	47	48	48
BESAYA	29	29	29	29	29	29
CAMPOO	17	17	17	16	16	16
COSTA CENTRAL	125	125	125	125	125	124
LIÉBANA	35	35	35	36	37	37
PAS-PISUEÑA-MIERA	19	19	19	19	19	19
SAJA-NANSA	56	56	56	56	56	57
SANTANDER	99	99	99	99	99	99
TRASMIERA	124	124	124	123	123	123
Total general	553	553	551	550	552	552

DIRECTORIO DE TURISMO. ESTABLECIMIENTOS ABIERTOS						
ZONA_TURISTICA	2008					
	1	2	3	4	5	6
ASÓN-AGÜERA	38	37	39	39	40	42
BESAYA	28	29	29	29	29	29
CAMPOO	16	16	15	15	15	15
COSTA CENTRAL	67	68	85	89	88	97
LIÉBANA	24	25	32	34	35	37
PAS-PISUEÑA-MIERA	15	15	16	16	16	16
SAJA-NANSA	38	39	42	43	43	51
SANTANDER	75	75	81	81	81	84
TRASMIERA	60	60	76	79	81	98
Total general	361	364	415	425	428	469

La encuesta tiene dos muestras, una a la que se le envía cuestionario en donde deben de reflejar los movimientos a lo largo de una semana escogida del mes, y otra, en donde se incluyen los establecimientos hoteleros de mayor categoría, que deben de responder a un cuestionario mensual en donde se les solicita información global sobre los movimientos a lo largo del mes.

Las encuestas semanales realizadas por el INE a lo largo del 2008 con respuesta son las siguientes:

a) Hoteles abiertos

ENCUESTAS						
ZONA_TURISTICA	2008					
	1	2	3	4	5	6
ASÓN-AGÜERA	15	17	20	20	21	23
BESAYA	11	12	13	14	14	14
CAMPOO	8	7	8	7	7	7
COSTA CENTRAL	20	24	40	45	48	58
LIÉBANA	3	5	13	16	16	18
PAS-PISUEÑA-MIERA	7	8	8	9	9	7
SAJA-NANSA	8	7	10	13	11	15
SANTANDER	39	42	43	46	48	49
TRASMIERA	24	30	37	47	48	60
Total general	135	152	192	217	222	251

b) Hoteles cerrados

ENCUESTAS

ZONA_TURISTICA	2008					
	1	2	3	4	5	6
ASÓN-AGÜERA	6	4	3	3	3	1
BESAYA	2	2	1	0	0	0
CAMPOO	0	1	0	1	1	1
COSTA CENTRAL	18	15	13	12	10	7
LIÉBANA	10	9	3	2	3	3
PAS-PISUEÑA-MIERA	1	0	0	0	0	0
SAJA-NANSA	4	7	6	4	5	5
SANTANDER	6	3	7	5	2	2
TRASMIERA	16	10	16	11	12	7
Total general	63	51	49	38	36	26

Quedando una pequeña parte de la muestra pendiente de respuesta, la existencia de encuestas pendientes de responder da lugar a que en esta estadística existan resultados provisionales, que se hacen definitivos una vez que se han recibido y depurado todas las encuestas:

ENCUESTAS PENDIENTE DE RESPUESTA

ZONA_TURISTICA	2008					
	1	2	3	4	5	6
ASÓN-AGÜERA	0	0	0	0	0	0
BESAYA	0	0	0	0	0	0
CAMPOO	1	1	1	0	0	0
COSTA CENTRAL	1	1	6	4	3	4
LIÉBANA	0	0	0	0	0	0
PAS-PISUEÑA-MIERA	0	0	1	0	0	2
SAJA-NANSA	1	0	1	0	1	1
SANTANDER	0	0	1	0	1	1
TRASMIERA	1	1	3	0	0	5
Total general	4	3	13	4	5	13

La encuestas semanales disponibles serán para los seis primeros mes del 2008 las siguientes:

MUESTRA SEMANAL						
ZONA_TURISTICA	2008					
	1	2	3	4	5	6
ASÓN-AGÜERA	21	21	23	23	24	24
BESAYA	13	14	14	14	14	14
CAMPOO	9	9	9	8	8	8
COSTA CENTRAL	39	40	59	61	61	69
LIÉBANA	13	14	16	18	19	21
PAS-PISUEÑA-MIERA	8	8	9	9	9	9
SAJA-NANSA	13	14	17	17	17	21
SANTANDER	45	45	51	51	51	52
TRASMIERA	41	41	56	58	60	72
Total general	202	206	254	259	263	290

Sobre el directorio se puede observar fracciones de muestreo cercanas al 50% en cada zona, exceptuando SAJA-NANSA, en donde se encuesta a un porcentaje próximo al 50% de los establecimientos hoteleros. Como se puede apreciar la muestra semanal cubre con suficiencia la distribución de los establecimientos hoteleros para realizar un análisis como el que se presenta.

MUESTRA SEMANAL						
ZONA_TURISTICA	2008					
	1	2	3	4	5	6
ASÓN-AGÜERA	42,9%	42,9%	48,9%	48,9%	50,0%	50,0%
BESAYA	44,8%	48,3%	48,3%	48,3%	48,3%	48,3%
CAMPOO	52,9%	52,9%	52,9%	50,0%	50,0%	50,0%
COSTA CENTRAL	31,2%	32,0%	47,2%	48,8%	48,8%	55,6%
LIÉBANA	37,1%	40,0%	45,7%	50,0%	51,4%	56,8%
PAS-PISUEÑA-MIERA	42,1%	42,1%	47,4%	47,4%	47,4%	47,4%
SAJA-NANSA	23,2%	25,0%	30,4%	30,4%	30,4%	36,8%
SANTANDER	45,5%	45,5%	51,5%	51,5%	51,5%	52,5%
TRASMIERA	33,1%	33,1%	45,2%	47,2%	48,8%	58,5%
Total general	36,5%	37,3%	46,1%	47,1%	47,6%	52,5%

La muestra mensual distribuida por zonas es la que figura en la tabla adjunta:

MUESTRA MENSUAL						
ZONA_TURISTICA	2008					
	1	2	3	4	5	6
ASÓN-AGÜERA	7	7	7	7	7	7
BESAYA	4	4	4	4	4	4
CAMPOO	4	4	4	4	4	4
COSTA CENTRAL	22	23	30	30	30	32
LIÉBANA	3	3	4	4	5	5
PAS-PISUEÑA-MIERA	3	3	3	3	3	3
SAJA-NANSA	4	4	5	5	5	7
SANTANDER	23	23	26	26	26	26
TRASMIERA	15	15	18	19	19	19
Total general	85	86	101	102	103	107

3.- FASE I: OBTENCIÓN DE ESTIMADORES SEGÚN ICANE

La estimación de las variables de la encuesta por zonas turísticas se ha realizado elevando la muestra del INE por las zonas turísticas de Cantabria. Dicha elevación la hemos realizado tal y como describe el INE en sus documentos metodológicos [1]. Se parte diferenciando los establecimientos hoteleros en dos grupos:

- Establecimientos de 3, 4 y 5 estrellas de oro.
- Resto de establecimientos.

La información que se dispone para cada grupo es semanal, pero con diferente desglose de países en viajeros y pernoctaciones, en los establecimientos de 3, 4 y 5 estrellas de oro, para los que además se dispone una parte de la información para todas las semanas del mes. Es decir, hay dos encuestas: una semanal, que consiste en las respuestas que el establecimiento da a cada uno de los 7 días de una semana-muestra escogida del mes (esta respuesta incluye la nacionalidad de los viajeros); y otra mensual, que consta tan sólo de un valor global de viajeros y pernoctaciones para todo el mes.

Partiendo de las siguientes variables:

E = número de establecimientos abiertos en el mes, existentes en el directorio

ET = número de establecimientos existentes en el directorio

e = número de establecimientos que contestan a la encuesta

e' = número de establecimientos que contestan a la encuesta mensual

c = número de establecimientos de la muestra que están cerrados dentro de su periodo de apertura

c' = número de establecimientos que están cerrados dentro de su periodo de apertura en la encuesta mensual

P = número de plazas según directorio

N' = número de plazas supletorias utilizadas

V = número de viajeros entrados

VM = número de viajeros entrados en el total del mes

N = número de plazas ocupadas (pernoctaciones)

NM = número de plazas ocupadas (pernoctaciones) en el total del mes

ES = estancia media

GP = grado de ocupación por plazas

D = número de días del mes de referencia (28, 29, 30, 31)

dm = número de días que el establecimiento ha estado abierto en el mes de referencia

GE = grado de ocupación por habitaciones esperado para un mes concreto

Y con los subíndices siguientes:

i = establecimiento hotelero

j = zona turística

k = categoría de establecimiento hotelero (5, 4, 3, 2, 1 oro: 3, 2 y 1 plata)

m = lugar de residencia (nacionalidad: país o comunidad autónoma) del viajero

Se obtienen dos grupos de estimaciones:

1.- Estimadores con la semana:

- Se calculan para todas las categorías de establecimientos.
- Sólo se realiza para una muestra de establecimientos.
- Incluye la nacionalidad del viajero.
- Total del mes = (total semana / 7) * días del mes {28 o 30 o 31}

2.- Estimadores con totales mensuales:

- Se calculan para sólo para los establecimientos de 3, 4 y 5 estrellas de oro.
- Se encuestan todos los establecimientos (datos censales).
- No incluye la nacionalidad del viajero.
- Sólo nos informa de una respuesta global para todo el mes (A viajeros, B pernnoctaciones).

3.1.- ESTIMADORES CON LA SEMANA

Estimación del número de establecimientos abiertos en el mes

$$\hat{E}_{jk} = E_{jk} \cdot \frac{e_{jk}}{e_{jk} + c_{jk}}$$

Estimación del número de plazas de los establecimientos abiertos en el mes

$$\hat{P}_{jk} = \sum_{i=1}^{E_{jk}} P_{ijk} \cdot \frac{\sum_{i=1}^{e_{jk}} P_{ijk}}{\sum_{i=1}^{e_{jk}} P_{ijk} + \sum_{i=1}^{c_{jk}} P_{ijk}}$$

Estimación del número de viajeros entrados en el mes

$$\hat{V}_{jkm} = \sum_{i=1}^{e_{jk}} V_{ijkm} \cdot \frac{D}{7} \cdot \frac{\hat{P}_{jk}}{\sum_{i=1}^{e_{jk}} P_{ijk}}$$

Estimación de las plazas ocupadas (Pernoctaciones)

$$\hat{N}_{jkm} = \sum_{i=1}^{e_{jk}} N_{ijkm} \cdot \frac{D}{7} \cdot \frac{\hat{P}_{jk}}{\sum_{l=1}^{e_{jk}} P_{ijlk}}$$

Estimación de las plazas ocupadas supletorias o pernoctaciones en dichas plazas

$$\hat{N}'_{jk} = \sum_{i=1}^{e_{jk}} N'_{ijk} \cdot \frac{D}{7} \cdot \frac{\hat{P}_{jk}}{\sum_{i=1}^{e_{jk}} P_{ijk}}$$

Estancia media por categorías

$$\hat{ES}_{jk} = \frac{\sum_m \hat{N}_{jkm}}{\sum_m \hat{V}_{jkm}}$$

Estancia media por país de residencia

$$\hat{ES}_{jm} = \frac{\sum_k \hat{N}_{jkm}}{\sum_k \hat{V}_{jkm}}$$

Grado de ocupación neto por plazas (porcentaje)

$$\hat{GP}_{jk} = \frac{\sum_m \sum_{i=1}^{e'_{jk}} N_{ijkm}}{7 \sum_{i=1}^{e'_{jk}} P_{ijk} + \sum_{i=1}^{e'_{jk}} N'_{ijk}} \cdot 100$$

Para el conjunto de categorías

$$\hat{GP}_j = \frac{\sum_k \hat{GP}_{jk} \cdot \hat{P}_{jk}}{\sum_k \hat{P}_{jk}}$$

3.2.- ESTIMADORES CON TOTALES MENSUALES

Estimación del número de establecimientos abiertos en el mes

$$\hat{EM}_{jk} = E_{jk} \cdot \frac{\sum_{i=1}^{e'_{jk}} \frac{dm_{ijk}}{D}}{e'_{jk} + c'_{jk}}$$

Estimación del número de plazas de los establecimientos abiertos en el mes

$$\hat{PM}_{jk} = \sum_{i=1}^{E_{jk}} P_{ijk} \cdot \frac{\sum_{i=1}^{e'_{jk}} P_{ijk} \cdot \frac{dm_{ijk}}{D}}{\sum_{i=1}^{e'_{jk}} P_{ijk} + \sum_{i=1}^{c'_{jk}} P_{ijk}}$$

Estimación del número de viajeros entrados en el mes

$$\hat{VM}^*_{jk} = \sum_{i=1}^{e'_{jk}} VM_{ijk} \cdot \frac{\hat{PM}_{jk}}{\sum_{i=1}^{e'_{jk}} P_{ijk} \cdot \frac{dm_{ijk}}{D}}; \hat{VM}^*_{jkm} = \hat{VM}^*_{jk} \cdot \frac{\hat{V}_{jkm}}{\sum_m \hat{V}_{jkm}}$$

Obsérvese aquí que, como la encuesta mensual sólo nos informa del nº total de viajeros entrados en el mes, sin decir nada de la nacionalidad (m) de los mismos, tenemos que acudir a la encuesta semanal, y tomar de ella la proporción de viajeros entrados de cada nacionalidad. Obtenemos así el resultado «los viajeros entrados en el mes, según su nacionalidad», combinando dos informaciones:

- De la encuesta mensual tomamos el total de viajeros entrados en el mes (este dato es evidentemente más preciso que el correspondiente de la encuesta semanal, que sólo toma información de una semana del mes).
- De la encuesta semanal se toma la proporción de viajeros entrados de cada nacionalidad.

Estimación de las plazas ocupadas (Pernoctaciones)

$$\widehat{NM}^*_{jk} = \sum_{i=1}^{e'_{jk}} NM_{ijk} \cdot \frac{\widehat{PM}_{jk}}{\sum_{i=1}^{e'_{jk}} P_{ijk} \cdot \frac{dm_{ijk}}{D}}; \widehat{NM}^*_{jkm} = \widehat{NM}^*_{jk} \cdot \frac{\widehat{N}_{jkm}}{\sum_m \widehat{N}_{jkm}}$$

Son válidas aquí las mismas observaciones que para los viajeros.

Estimación de las plazas ocupadas supletorias o pernoctaciones en dichas plazas

$$\widehat{N}'^*_{jk} = \sum_{i=1}^{e_{jk}} N'_{ijk} \cdot \frac{D}{7} \cdot \frac{\widehat{PM}_{jk}}{\sum_{i=1}^{e_{jk}} P_{ijk}}$$

Estancia media por categorías

$$\widehat{ES}^*_{jk} = \frac{\sum_m \widehat{NM}^*_{jkm}}{\sum_m \widehat{VM}^*_{jkm}}$$

Estancia media por país de residencia

$$\widehat{ES}^*_{jm} = \frac{\sum_k \widehat{NM}^*_{jkm}}{\sum_k \widehat{VM}^*_{jkm}}$$

Estimación del grado de ocupación por plazas (en porcentaje)

$$\widehat{GP}^*_{jk} = \frac{\widehat{NM}^*_{jk}}{D \cdot \widehat{PM}_{jk} + \widehat{N}'^*_{jk}} \cdot 100$$

Para el conjunto de categorías

$$\widehat{GP}^*_j = \frac{\sum_k \widehat{GP}^*_{jk} \cdot \widehat{PM}_{jk}}{\sum_k \widehat{PM}_{jk}}$$

3.3.- TRATAMIENTO DE LA FALTA DE RESPUESTA

Para poder realizar una estimación en todas las categorías ha habido que realizar un tratamiento para la falta de respuesta que se resume a continuación:

- 1) Si en un estrato (categoría dentro de una zona) sólo hay establecimientos cerrados, se supone que todos los establecimientos de la zona y categoría están cerrados.
- 2) Si en un estrato no se ha encuestado ningún establecimiento en el mes en cuestión o se han recibido respuestas negativas (falta de respuesta), se imputan los resultados tomados del primer estrato no vacío que se encuentra descendiendo o ascendiendo según la categoría más próxima, pero sin salirnos de la zona.

Suponiendo que el estrato k de la zona j está vacío, bien porque no se ha encuestado a ningún establecimiento del estrato o por haberse recibido respuesta negativa, y que lo vamos a imputar por los datos del estrato k' :

- La estimación de las variables del número de establecimientos abiertos en el mes y del número de plazas de los establecimientos abiertos en el mes coincide con el dato del directorio.
- Para la variable de viajeros (para pernoctaciones es similar):

$$\hat{V}_{jk} = \sum_{i=1}^{e_{jk'}} V_{ijk'} \cdot \frac{D}{7} \cdot \frac{\hat{P}_{jk}}{\sum_{i=1}^{e_{jk'}} P_{ijk'}} = \underbrace{\left(\sum_{i=1}^{e_{jk'}} V_{ijk'} \cdot \frac{D}{7} \cdot \frac{1}{\sum_{i=1}^{e_{jk'}} P_{ijk'}} \right)}_{\text{Información estrato k'}} \cdot \underbrace{\hat{P}_{jk}}_{\text{Información estrato k.}}$$

Esta fórmula expresa que lo que se toma prestado del estrato donante no es, evidentemente, el valor absoluto de sus viajeros, sino el ratio «viajeros por plaza» registrado en el mismo. Por ejemplo, si en la zona Z1 tenemos que para la categoría de 2 estrellas de oro no hay ninguna respuesta, aunque sabemos que en esa zona según el directorio de Hoteles se ofrecen 50 plazas, entonces acudiremos a examinar la categoría de 1 estrella de oro de la misma zona Z1. Si en dicha categoría se ofrecen 100 plazas y sí se registraron viajeros en ella (por ejemplo 4 franceses y 10 madrileños), terminaremos concluyendo que para la categoría de 2 estrellas de oro hubo 2 franceses y 5 madrileños. Ésa es la forma de rellenar el estrato vacío. En el mismo ejemplo, si vemos que tampoco hay respuestas para la categoría de 1 estrella de oro, acudiremos a examinar la de 3 estrellas de oro, etc. Es decir, que a partir del estrato vacío, vamos moviéndonos en zig-zag alrededor del mismo, hasta encontrar un estrato con respuesta: primero nos

movemos hacia la categoría inmediata inferior, luego hacia la inmediata superior, luego hacia la 2ª inmediata inferior, etc. Eso sí, nunca abandonamos la zona turística, moviéndonos sólo de categoría en categoría, hasta encontrar el estrato donante con respuesta que andamos buscando.

Para saber si un estrato dado puede ser el donante, nos encontramos frente a dos casos. Ante todo, recordar que el estrato vacío lo es porque no registra respuestas a la encuesta semanal, existiendo además establecimientos abiertos en él según el directorio.

Caso 1.- Si nuestro estrato vacío no registra respuesta a la encuesta mensual (o la registra, pero con 0 viajeros), entonces nos servirá como donante cualquier estrato que tenga establecimientos abiertos. En el ejemplo anterior, si en la categoría de 1 estrella de oro había establecimientos abiertos según el directorio, pero se descubrió que todos ellos estaban cerrados, entonces la conclusión que se sacó para esa categoría es que no hubo viajeros. Y eso es lo que trasladaremos al estrato vacío: tampoco habrá viajeros para él.

Caso 2.- En cambio, si para nuestro estrato vacío consta que hay viajeros para la encuesta mensual (es decir sabemos que hubo X viajeros en total en todo el mes, aunque desconocemos la nacionalidad de los mismos, porque ésta sólo se registra en la encuesta semanal), entonces nuestro problema no es tanto saber cuántos viajeros hubo en el estrato vacío, sino de qué nacionalidad eran. En este caso, el estrato donante tiene que tener viajeros. Y trasladaremos al estrato vacío la proporción de viajeros de cada nacionalidad registrada en el estrato donante. Esta situación puede darse si un hotel abre 10 días del mes: a la encuesta semanal puede responder que estuvo cerrado (por lo que no tendremos información sobre la nacionalidad de los viajeros), mientras que a la encuesta mensual responderá con el total de viajeros que entraron en ese mes (durante los 10 días estuvo abierto).

Al margen del tratamiento anterior de la falta de respuesta, y aunque no está directamente relacionado con ello, nos hemos encontrado con una situación que creemos merece la pena mencionar en este apartado. Observando los datos para diciembre de 2006, en la zona turística de Trasmiera para la categoría de 1 estrella de plata vimos que había 100 pernoctaciones y 0 viajeros. Creyendo que se trataba de un error, examinamos los microdatos. En esa zona, había 3 establecimientos de 1 estrella de plata en la muestra. Dos de ellos estaban cerrados, y el otro estaba abierto, y sus respuestas para la semana-muestra efectivamente nos llevaban a deducir que en ese mes hubo 100 pernoctaciones pero no hubo viajeros. En concreto, en esa semana-muestra hubo 15 pernoctaciones de Castilla y León (60 en el mes) y 10 pernoctaciones de Asturias (40 en el mes). Como en la encuesta semanal el establecimiento tiene que registrar, para cada uno de los 7 días elegidos, cuántos viajeros entran ese día, y cuántos viajeros duermen, bien puede ocurrir que en una semana no entren viajeros, y que los que estén alojados hayan entrado la semana anterior del mismo mes, por ejemplo. El tratamiento que hemos decidido dar a esta situación consiste en que, si un establecimiento no registra viajeros y sí pernoctaciones en la semana-muestra, entonces nosotros consideraremos que sí ha habido viajeros, siendo éstos el resultado de dividir las pernoctaciones entre la estancia media (días que duerme un viajero) para Cantabria según el INE, para el año, mes y categoría de establecimiento. Si no tuviésemos cargado en ICANE los datos del INE para ese año, tomaríamos los del mismo mes pero del año anterior; si no tuviésemos datos del INE de ningún año, tomaríamos que la estancia media es 2. No presentamos esto como una solución definitiva, ni mucho menos, sino tan sólo como una posible heurística («arte de inventar») para lidiar con información incompleta como es la existente en la encuesta semanal. Creemos que el INE no realiza nada referente a la situación anterior, y quizás eso sea lo correcto. Relacionado con esto, está la información que publica el INE para Cantabria en Junio de 2006: hay 1163 pernoctaciones de Países africanos, frente a sólo 78 viajeros. La estimación correspondiente obtenida en ICANE, antes de realizar la Fase 2 de ajuste, es de 501 viajeros africanos en lugar de 78. Dando por correctas las 1163 pernoctaciones, ¿qué dato para los viajeros es el correcto? ¿78 viajeros que dan una estancia media de 15 días? ¿O 501 viajeros con una estancia media de 2,32 días? ¿No se tratará más bien de un error al grabar las pernoctaciones? Quién sabe.

4.- FASE 2: AJUSTE DE LOS ESTIMADORES, PARA QUE COINCIDAN CON EL INE

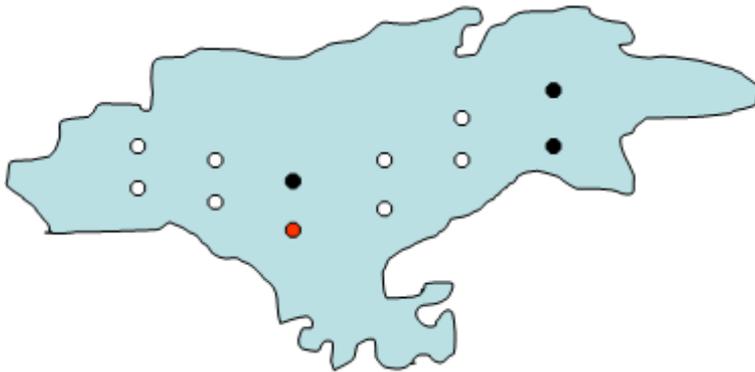
El sistema de elevación considerando la zona turística no garantiza que los resultados regionales obtenidos coincidan con los estimados por el INE.

Las zonas turísticas consideradas en ICANE para realizar las estimaciones, son distintas de las que considera el INE:

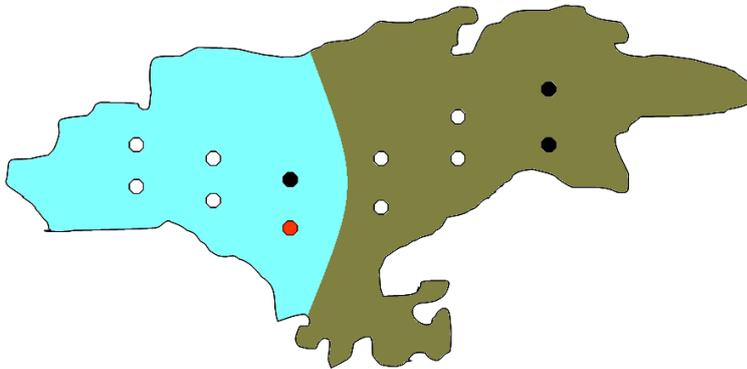
Zonas INE	Zonas ICANE
Eje transversal	Asón-Agüera
Santander	Besaya
Zona Occidental	Campoo
Zona Oriental: Costa Esmeralda	Costa Central
Zona Oriental: Resto Zona	Liébana
	Pas-Pisueña-Miera
	Saja-Nansa
	Santander
	Trasmiera

Como la elevación de la información muestral se realiza de manera independiente para cada zona interior de Cantabria, el resultado final de las estimaciones obtenidas para el conjunto Cantabria será diferente, dependiendo de la manera en que se haya decidido dividir Cantabria en zonas. Si consideramos las zonas del INE obtendremos **UNAS** estimaciones. Si consideramos las zonas del ICANE obtendremos **OTRAS**. Aunque esto será evidente para quien tenga experiencia en tratamientos estadísticos, nos gustaría ilustrarlo con un sencillo ejemplo.

Supongamos que tenemos 12 hoteles en toda Cantabria encuestándose sólo 4 de ellos. En el siguiente gráfico los hoteles no encuestados se representan mediante un círculo blanco. Un círculo negro indica un hotel encuestado y que resulta estar abierto, mientras que el círculo rojo corresponde a un hotel encuestado y que se descubrió cerrado.



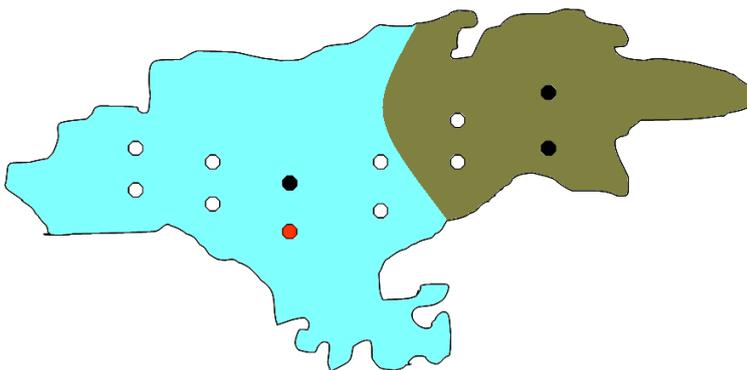
Consideremos ahora dos divisiones distintas de Cantabria, y veamos qué resultado final obtenemos para el parámetro estadístico «Número de hoteles abiertos».



Zona azul:
 50 % → **3** abiertos

Zona marrón:
 100 % → **6** abiertos

TOTAL = 3 + 6 = **9** abiertos



Zona azul:
 50 % → **4** abiertos

Zona marrón:
 100 % → **4** abiertos

TOTAL = 4 + 4 = **8** abiertos

Debemos mencionar que, aunque las estimaciones que realizamos en ICANE las hubiésemos hecho utilizando las mismas zonas turísticas que considera el INE, sólo hubiéramos conseguido coincidir en el nº de establecimientos abiertos y en el nº de plazas ofertadas. Para los viajeros y las pernoctaciones el INE da unos datos y nosotros damos otros. Esto se debe a que, aún utilizando las mismas zonas que el INE, hay una casuística muy compleja y no hemos conseguido en ICANE emular del todo el proceso de elevación que realiza el INE.

Para lograr la igualdad de las estimaciones realizadas en ICANE con las que publica el INE en su página web, se ha realizado un ajuste biproportional [2] o algoritmo RAS. Se trata de un tratamiento mecánico, que va modificando una matriz de datos en sucesivas iteraciones hasta que ésta cumple un objetivo perseguido y marcado con antelación. Las filas de la matriz representan las categorías de hoteles, mientras que las columnas son las nacionalidades (país o comunidad autónoma) de los viajeros. El algoritmo RAS es bastante simple y directo, y se comprenderá muy bien con un ejemplo.

Para simplificar, consideraremos sólo 2 categorías de hoteles y 3 nacionalidades de viajeros. Explicaremos cómo realizar el ajuste o calibrado del parámetro «Viajeros», siendo el tratamiento idéntico para las «Pernoctaciones».

Las celdas (1,1), (1,2) y (1,3) muestran los viajeros Alemanes, Andaluces y Asturianos alojados en hoteles de 3 estrellas, en Cantabria, en un mes dado.

Análogamente, las celdas (2,1), (2,2) y (2,3) son los viajeros para hoteles de 2 estrellas.

Esas 6 celdas muestran, en la matriz correspondiente al paso 1, los estimadores deducidos al final de la Fase 1 por el proceso de elevación por zonas desarrollado en ICANE. Constituyen la situación de partida, el punto de arranque del algoritmo. Son esas 6 celdas (con fondo blanco) las que van a ir variando a lo largo de la ejecución del algoritmo.

Hay que mencionar que detrás de cada número alojado en esas celdas en realidad hay varios sumandos, correspondientes a las zonas turísticas en que se ha dividido Cantabria. Por ejemplo, la celda (1,2) contiene los 40 Andaluces alojados en hoteles de 3 estrellas en toda Cantabria. Ese número 40 podría ser el resultado de:

- 7 viajeros en la zona «Liébana»
- 13 viajeros en la zona «Besaya»
- 20 viajeros en la zona «Santander»

Cuando al final de la ejecución exitosa del algoritmo se deduzca que hay que transformar ese número 40 en 40.82, sabremos que el factor de corrección a aplicar a la celda (1,2) es $40.82/40 = 1.0205$, y ese mismo factor lo aplicaremos a los sumandos que componían el número 40:

- $7 * 1.0205 = 7.1435$ viajeros en la zona «Liébana»
- $13 * 1.0205 = 13.2665$ viajeros en la zona «Besaya»
- $20 * 1.0205 = 20.4100$ viajeros en la zona «Santander»

La fila 3 es simplemente el total de la suma de columnas. Así por ejemplo la casilla (3,1) es el total de viajeros Alemanes alojados en Cantabria, según ICANE.

De manera similar, la columna 4 es el total de la suma de filas. Por ejemplo, la casilla (2,4) son los viajeros totales alojados en hoteles de 2 estrellas en toda Cantabria, según el ICANE.

Es importante ver que tenemos así dos proyecciones de la información contenida en la matriz:

- La fila 4 nos dice los viajeros de cada nacionalidad que se alojan en Cantabria, sin decirnos nada acerca de la categoría de los hoteles en que se alojan.
- La columna 4 son los viajeros que se alojan en cada categoría de hotel en Cantabria, sin aludir para nada a la nacionalidad de los mismos.

Precisamente esos dos puntos de vista para los viajeros son los que proporciona el INE. El INE nunca nos dice cuántos viajeros de una nacionalidad concreta se alojan en una categoría de hotel dada. Si tuviésemos ese dato combinado del INE (nacionalidad y categoría), el ajuste no requeriría un algoritmo iterativo. En lugar de esto, el INE publica de manera separada:

- Cuántos viajeros hay según el lugar de residencia (nacionalidad: país o CCAA) de los mismos.
- Cuántos viajeros hay según la categoría del hotel elegido para alojarse.

Esas dos informaciones del INE las recogemos como dos vectores (uno horizontal y otro vertical). En nuestro ejemplo tenemos que el INE nos dice que hubo, por una parte, 40 viajeros Alemanes, 100 Andaluces y 120 Asturianos; y por otra, 100 viajeros alojados en hoteles de 3 estrellas, y 160 en hoteles de 2 estrellas.

Un paso previo para que el algoritmo funcione bien y termine convergiendo en una solución, es que los dos objetivos que se le marcan sean el mismo. En el ejemplo tenemos que el INE dice que hubo 260 viajeros en total. Ese dato de 260 se obtiene sumando los componentes del vector-objetivo vertical de viajeros por categoría del hotel (100+160). Y también debe obtenerse el mismo valor 260 sumando los componentes del vector-objetivo horizontal de viajeros por nacionalidad (40+100+120). Si este último vector fuese por ejemplo «50, 110, 130» en lugar de «40, 100, 120» tendríamos que modificarlo para que sumase 260. La corrección consistiría en multiplicar cada componente por $260/(50+110+130) = 0.8966$ con lo que se obtendría el vector-objetivo «45, 99, 116» que sí da una suma de 260 viajeros. Este paso es necesario porque debido a los redondeos que se producen cuando el INE publica los viajeros como números enteros para cada una de las nacionalidades, sucede que la suma de esos números enteros redondeados puede diferir en 1 o 2 unidades respecto de la suma de viajeros por categorías. Si dejásemos esto así, sin hacer la corrección del vector-objetivo horizontal, le estaríamos dando dos objetivos contradictorios al algoritmo:

- Por una parte le decimos que los viajeros tienen que sumar 260 (100+160).
- Por otra le decimos que los viajeros tienen que sumar 290 (50+110+130).

Esto ocasionaría una falta de convergencia del algoritmo, ocasionada por la imposibilidad lógica de hallar una solución que cumpla a la vez ambos objetivos.

La corrección la realizamos en el vector-objetivo de viajeros por nacionalidades, porque al tener más componentes creemos que tendrá más pérdidas por redondeos parciales, por lo que su suma será más imprecisa. En cambio el vector-objetivo de viajeros por categorías tiene siempre sólo 7 componentes, que son los correspondientes a las categorías de hoteles existentes. Es decir, damos prioridad al dato global por categorías frente al dato global por nacionalidades, y si éste no coincide con aquél, lo corregimos.

Una vez que tenemos la matriz con las estimaciones calculadas en ICANE, y los dos vectores-objetivos del INE, comienza la ejecución del algoritmo, el cual consta de un bucle en el que se van modificando todas las celdas de la matriz, de manera que al finalizar el algoritmo con éxito se cumplirán las dos igualdades siguientes:

- Totales de filas de la matriz = Vector-objetivo vertical
- Totales de columnas de la matriz = Vector-objetivo horizontal

En las iteraciones pares del bucle se ajustan las Filas; en la impares las columnas.

Paso 1. Inicio, situación de partida. Diferencias = 60 → Iteración 1ª (ajuste de columnas)

	Alemania	Andalucía	Asturias	TOTAL	Objetivo INE	Diferencias
3 estrellas	10	40	60	110	100	10
2 estrellas	20	50	80	150	160	10
TOTAL	30	90	140	260		

Objetivo INE	40	100	120
--------------	----	-----	-----

260

Diferencias	10	10	20
-------------	----	----	----

60

Estamos en la primera iteración (impar), por lo que hay que ajustar las columnas. Se trata de modificar la matriz entera para que los totales de columnas «30, 90, 140» coincidan con el vector-objetivo horizontal «40, 100, 120». Esta modificación consiste en tres modificaciones sucesivas:

- Modificamos la 1ª columna «10, 20» de manera que su suma sea el objetivo «40». Multiplicando cada componente por $40/(10+20) = 1.33333$, llamado factor de corrección, obtenemos «13.33, 26.67».
- Modificamos la 2ª columna «40, 50» para que sume el objetivo «100». El factor de corrección es $100/(40+50) = 1.11111$ y obtenemos «44.44, 55.56».
- Modificamos la 3ª columna «60, 80» para que sume el objetivo «120». Con el factor de corrección $120/(60+80) = 0.85714$ se obtiene «51.43, 68.57».

Como se ve, la iteración primera ha logrado satisfacer completamente el vector-objetivo horizontal, pero hemos desatendido el vector-objetivo vertical. De él se ocupará la siguiente iteración.

Paso 2. Diferencias = 18.40 → Vamos a la iteración 2ª (ajuste de filas)

	Alemania	Andalucía	Asturias	TOTAL	Objetivo INE	Diferencias
3 estrellas	13.33	44.44	51.43	109.20	100	9.20
2 estrellas	26.67	55.56	68.57	150.80	160	9.20
TOTAL	40	100	120	260		

Objetivo INE	40	100	120
---------------------	----	-----	-----

260

Diferencias	0	0	0
--------------------	---	---	---

18.40

En esta 2ª iteración nos ocuparemos de ajustar las filas, es decir, de modificar todas las celdas de la matriz para que sus totales de filas «109.20, 150.80» coincidan con el vector-objetivo vertical «100, 160». Primero modificamos una fila y luego la otra:

- Modificamos la 1ª fila «13.33, 44.44, 51.43» para que sume su objetivo «100». El factor de corrección $100/(13.33+44.44+51.43) = 0.91575$ nos transforma esa 1ª fila en «12.21, 40.70, 47.10».
- Modificamos la 2ª fila «26.67, 55.56, 68.57» para que sume su objetivo «160». El factor de corrección $160/(26.67+55.56+68.57) = 1.06101$ nos lleva a la nueva 2ª fila «28.30, 58.95, 72.75».

Ya hemos visto las dos primeras iteraciones del algoritmo. Primero hemos ajustado las columnas y luego las filas. Estos dos pasos se repetirán sucesivamente, ocupándonos alternativamente de la suma de columnas y de filas, hasta que las diferencias con los dos vectores-objetivo sean nulas.

Paso 3. Diferencias = 1.02 → Vamos a la iteración 3ª (ajuste de columnas)

	Alemania	Andalucía	Asturias	TOTAL	Objetivo INE	Diferencias
3 estrellas	12.21	40.70	47.10	100.01	100	0.01
2 estrellas	28.30	58.95	72.75	160	160	0
TOTAL	40.51	99.65	119.85	260.01		

Objetivo INE	40	100	120
---------------------	----	-----	-----

260

Diferencias	0.51	0.35	0.15
--------------------	------	------	------

1.02

Los factores de corrección para las 3 columnas son 0.98741, 1.00351 y 1.00125.

Paso 4. Diferencias = 0.12 → Vamos a la iteración 4ª (ajuste de filas)

	Alemania	Andalucía	Asturias	TOTAL	Objetivo INE	Diferencias
3 estrellas	12.06	40.84	47.16	100.06	100	0.06
2 estrellas	27.94	59.16	72.84	159.94	160	0.06
TOTAL	40	100	120	260		

Objetivo INE	40	100	120
---------------------	----	-----	-----

260

Diferencias	0	0	0
--------------------	---	---	---

0.12

Los factores de corrección para las 2 filas son 0.99940 y 1.00038.

Paso 5. Diferencias = 0 → Final con éxito del algoritmo

	Alemania	Andalucía	Asturias	TOTAL
3 estrellas	12.05	40.82	47.13	100
2 estrellas	27.95	59.18	72.87	160
TOTAL	40	100	120	260

Objetivo INE
100
160

Diferencias
0.06
0.06

Objetivo INE	40	100	120

260

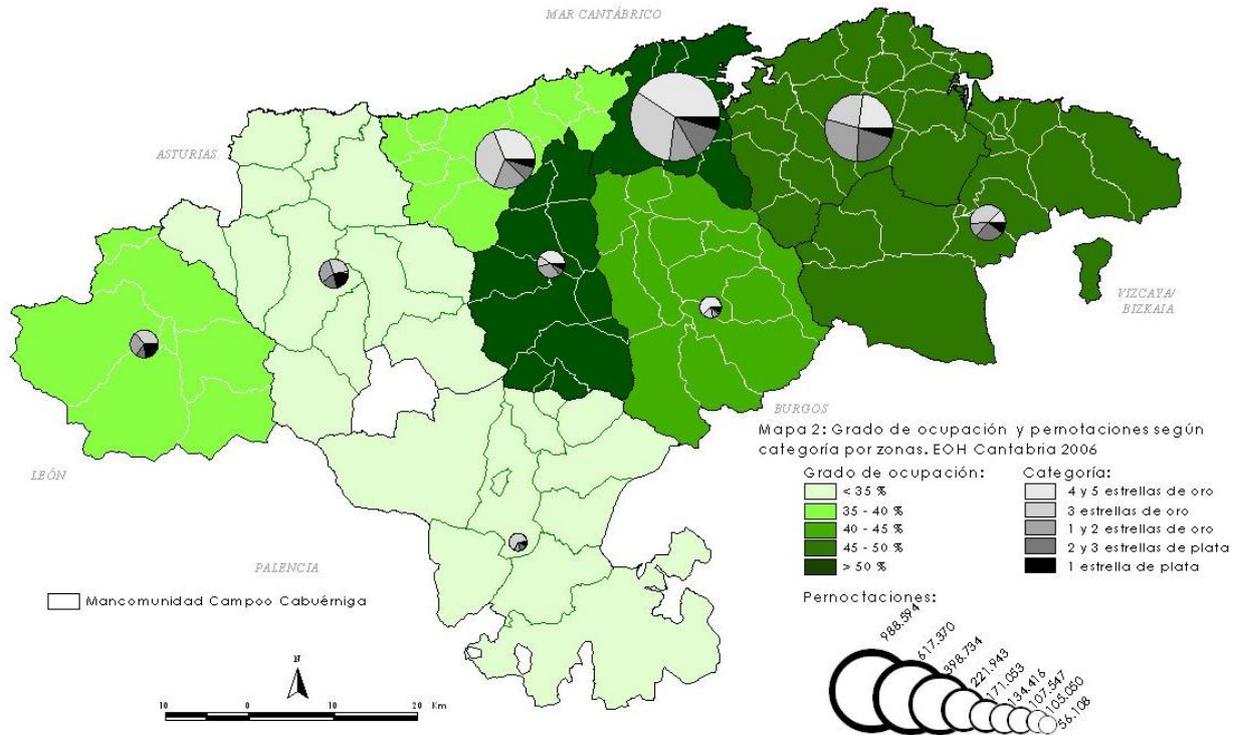
Diferencias	0	0	0

0.12

Como se ve, en sólo 5 iteraciones hemos conseguido realizar el calibrado o ajuste de la matriz. «Ireland, C.T. & Kullback» demuestran en su artículo [2] que el algoritmo siempre termina convergiendo. Nosotros nos detenemos cuando la suma de diferencias, considerando sólo 2 decimales, es cero (0.00). Además hemos limitado a 100 el número de iteraciones, y damos un error si el ajuste no se logra.

5.- RESULTADOS OBTENIDOS

En el mapa adjunto se pueden observar los resultados obtenidos para el 2006 de los grados de ocupación y las pernoctaciones por categoría de establecimiento:



En la tabla adjunta se presentan los datos correspondientes a diciembre del 2006:

VIAJEROS		ASÓN- AGÜERA	BESAYA	CAMPOO	COSTA CENTRAL	LIÉBANA	PAS- PISUEÑA- MIERA	SAJA- NANSA	SANTANDER	TRASMIERA	CANTABRIA
ESPAÑOLES	Andalucía	174	10	45	250	22	61	-	1.101	60	1.723
	Asturias (Principado de)	294	389	80	282	499	142	192	1.374	821	4.073
	Cantabria	172	662	107	423	136	187	499	3.487	616	6.290
	Castilla y León	249	205	128	241	71	164	171	1.641	521	3.392
	Cataluña	394	294	72	146	32	32	26	2.923	252	4.171
	Galicia	282	160	311	292	19	52	138	1.398	270	2.922
	Madrid (Comunidad de)	1.263	819	264	558	654	310	274	4.718	1.083	9.942
	País Vasco	827	541	130	783	106	590	317	1.682	1.824	6.800
	Resto de España	1.003	849	217	1.044	116	298	380	4.372	1.361	9.639
	Total	4.658	3.930	1.354	4.019	1.656	1.836	1.997	22.696	6.807	48.953
EXTRANJEROS	Alemania	8	10	1	78	-	-	-	296	4	398
	Francia	98	42	2	214	-	3	9	343	20	731
	Italia	20	37	2	14	-	-	-	622	-	695
	Portugal	10	28	124	21	10	-	10	182	-	386
	Reino Unido	100	18	-	100	-	5	-	410	15	649
	Resto Unión Europea	19	6	23	4	-	-	-	260	20	332
	Resto Europa	-	22	20	-	-	24	11	209	34	320
	Países africanos	-	5	-	-	-	-	-	28	-	33
	América	-	15	5	212	140	5	-	418	8	803
	Resto del mundo	7	-	6	-	-	-	-	581	125	720
Total	263	184	183	644	150	37	30	3.348	228	5.067	

7.- IMPLEMENTACIÓN

Mencionaremos en este apartado la plataforma tecnológica empleada en el ICANE para la explotación de EOH.

Los ficheros mensuales con los microdatos (directorio del mes, y respuestas de cada establecimiento hotelero a la encuesta) nos los envía el INE cifrados en un correo electrónico, de la manera habitual. Se cargan tal cual en una base de datos relacional Oracle, mediante SqlLoader. A continuación ejecutamos un programa escrito en el lenguaje PL/SQL de Oracle, que se encarga de obtener los estimadores a partir los microdatos. Dicho programa se ha desarrollado utilizando la herramienta gratuita SqlDeveloper, de Oracle. Los resultados de la ejecución del programa quedan en dos tablas relacionales. Una de ellas almacena los parámetros a nivel de «Año, Mes, Zona Turística, Categoría» (establecimientos abiertos, plazas, plazas supletorias). La otra almacena los parámetros a nivel de «Año, Mes, Zona Turística, Categoría, Nacionalidad del viajero [País, Comunidad Autónoma]» (viajeros y pernoctaciones). El grado de ocupación y la estancia media se infieren directamente a partir de los estimadores anteriores:

- Grado ocupación =
$$\frac{\text{Pernoctaciones}}{(\text{Plazas ofrecidas} + \text{Plazas supletorias ocupadas})}$$
- Estancia media =
$$\text{Pernoctaciones} / \text{Viajeros}$$

Una vez que tenemos los estimadores calculados, se los pasamos al personal de estadística del ICANE, para que confeccionen una hoja Excel que presente los resultados en el formato adecuado. Actualmente (Ocubre de 2008) se están desarrollando en ICANE programas en Crystal Reports para que este último paso quede más automatizado. La idea es que ejecutando un programa Crystal Reports se obtenga directamente el documento pdf listo para su difusión.

Hay dos tipos de informes que generamos en ICANE con los resultados EOH: un informe mensual y otro anual.

El informe mensual muestra los estimadores tal y como se obtienen en ICANE, es decir, sin la obligación de que coincidan con el INE. Esto es así porque al tratarse de datos provisionales, no merece la pena ejecutar el proceso de calibrado.

En cambio, el informe anual lo generamos cuando el INE nos comunica que para un año dado los resultados EOH son definitivos (por ejemplo, en Agosto de 2008, se dieron

por definitivos los datos EOH para 2007). ¿Cómo logramos que nuestras estimaciones ICANE coincidan con las del INE? Primero tenemos que extraer los resultados que publica el INE en su página web, e introducirlos en dos tablas relacionales Oracle. Una contiene los establecimientos abiertos, las plazas ofertadas, los viajeros y las pernoctaciones registradas, por «Año, Mes, Categoría». La otra contiene los viajeros y las pernoctaciones por «Año, Mes, Nacionalidad del viajero [País, Comunidad Autónoma]». Adicionalmente, tenemos una tabla para establecer una correspondencia entre las Nacionalidades del viajeros según el INE, y las que emplearemos para agregar los resultados en ICANE. Por ejemplo, aunque el INE muestra las nacionalidades «Polonia», «Suecia» etc. como separadas, nosotros decidimos agruparlas bajo el mismo título «Resto Unión Europea». Del mismo modo, no mostramos todas las comunidades autónomas de España como elementos diferenciados, sino que agrupamos en «Resto de España» las CCAA que consideramos con escaso volumen de viajeros. El paso que consiste en extraer la información que publica el INE con los resultados definitivos, e introducirlos en nuestra base de datos Oracle, es un proceso manual y muy pesado, y lo consideramos el punto débil de todo este proyecto. Hay multitud de problemas; por ejemplo, el INE no publica los viajeros de una nacionalidad cuando éste número es pequeño y por tanto poco significativo. Sin embargo en ICANE necesitamos saber exactamente cuántos viajeros de cada nacionalidad tenemos que conseguir, para lo que utilizamos un procedimiento solver de Excel para rellenar los huecos, y hay que terminar de cuadrar a ojo. Otro problema es que hay que acudir a varias páginas webs del INE para extraer la información: los viajeros extranjeros están en un sitio, mientras que los viajeros españoles están en otro, dando en este caso no el valor absoluto de viajeros de cada CCAA de procedencia, sino un porcentaje, que además puede no sumar exactamente 100, y debemos por tanto normalizar, etc. La parte buena de todo este proceso es que sólo hay que realizarlo una vez al año: cuando el INE eleva a definitivos sus estimadores para el año anterior.

Una vez que hemos cargado los resultados definitivos del INE en nuestras tablas Oracle, para obtener en ICANE los estimadores por zonas turísticas, de manera que la suma de las zonas coincida con el resultado para Cantabria según el INE, tan sólo tenemos que ejecutar el mismo programa PL/SQL que se ejecuta cada mes, pero ahora indicándole en sus parámetros de entrada que queremos obtener los estimadores para todo un año (parámetro MES con valor null), y que además se quiere realizar el calibrado con el INE (un parámetro booleano adhoc con valor true). Entonces dicho programa realiza, además de la Fase 1 que siempre se lleva a cabo (obtención de estimadores elevando por zonas), la Fase 2 (ajuste biproporcional o algoritmo RAS), la cual requiere de la presencia de información en las dos tablas Oracle mencionadas con los datos publicados por el INE.

AGRADECIMIENTOS

Agradecemos a Cèlia Gomà Aznar, del Institut d'Estadística de Catalunya (Idescat) la ayuda que ha prestado al personal del Instituto Cántabro de Estadística (Icane) durante las fases iniciales de este proyecto. Gracias a ella hemos podido descifrar el significado de la información contenida en los diversos campos de los ficheros del INE.

REFERENCIAS

[1] www.ine.es/ [Servicios > Encuesta de Ocupación Hotelera]

Metodología empleada por el INE en la encuesta de Ocupación Hotelera.

[2] Ireland, C.T. & Kullback, S. «Contingency tables with given marginals»

Biometrika 1968 55(1): 179-188