Hotel Price Index (HPI). Base 2008
(as of January 2016)

Methodological note

1. **Background**
2. **Sample design**
3. **Estimators**
4. **Base change 2008-2001**
5. **Statistical Secrecy**
6. **Variation quotients**
7. **Data collection. XML tourism**

INE National Statistics Institute

1. Background

The *Hotel Price Index*, HPI, is a statistical measurement of the monthly evolution of the prices that hotel businesspersons apply to their clients.

Information from the Tourist Accommodation Occupancy Survey is used to obtain this: Hotel Establishments (HOS), with the information that is collected monthly, from some 10,900 establishments in the summer and some 9,200 in the winter, to which a questionnaire is sent. From this questionnaire, information is obtained regarding hotel occupancy (guests registered, overnight stays, occupancy rate, etc.), structure (bedplaces, personnel, etc.) and the rest of the variables of interest, with a broad breakdown by geography and by establishment category. The questionnaire requests, among other variables, the ADR (Average Daily Rate), or average daily fee, applied to different types of client, for a double room with a bathroom. These fees are broken down according to the type of client to which they have been applied:

- Traditional tour operators
- Traditional travel agencies (including hotel vouchers and checks)
- Companies
- Individuals
- Groups
- Direct hiring on the hotel website and/or the hotel chain website
- Online tour operators
- Online travel agencies
- Others

This breakdown by type of client, as well as the introduction of the ADR concept in the section on prices in the HOS questionnaire, is an innovation introduced in January 2008. Until that time, data had been requested regarding prices, broken down by type of fee (see the methodology of the HPI, base 2001).

Both improvements have been introduced in order to respond to the changes occurred in the sales and distribution channels of the hotel sector that have been generated by Internet use (for example, online travel agencies and tour operators, or direct hiring on hotel websites), and to employ the price variables used by said sector and which are available in the management systems of hotel establishments.

These innovations require a base change in the HPI, to introduce in the computation of the index, a new breakdown by type of client, as well as the use of the ADR, the average daily fee, as an indicator of price.

Unlike the Consumer Price Index (CPI), the HPI is an indicator from the perspective of supply, as it measures the evolution of the prices that hotel establishments actually receive, invoiced to all types of client (the CPI only considers the prices applied to households resident in Spain). Therefore, it does not measure the evolution of prices paid by households, nor the official fee applied by hotel establishments, but rather the behaviour of the prices invoiced by the hotel establishments to different types of client, and through different distribution channels (households, companies, travel agencies and tour operators, whether traditional or online).

Indices and interannual variation rates are calculated and disseminated for the 17 Autonomous Communities, Ceuta and Melilla, in addition, indices and rates are published for the different categories on a national level.

2. Sample design

As the framework for the selection of the informant units, the survey uses the directories of the Tourist Offices of the Autonomous Communities and other sources, which include, in those that appear, amongst other data, the following information for each establishment: name, address, category, normal opening period, number of vacancies, number of rooms. number of holiday apartments and lots.

These directories are updated on a permanent basis.

The framework is divided into strata, the latter being defined by the the crossing of the variables category/province (or category/island in the case of the island provinces). Establishments are ordered according to number of places, from greatest to least, selecting the sample by means of a systematic procedure with random start in each stratum.

The survey is exhaustive in all provinces, except in some categories, for the provinces detailed hereunder (sample design 2014):

Pro-Cod	Provinces	Islands	3 gold		2 gold		1gold		2 and 3 silver		1 silver
03	Alicante		2/3		2/3		1/2				*See the following list*
06	Badajoz								1/2		
07	Balears Illes	Mallorca	1/3	1/2	2/5		1/3		1/3		
		Menorca	2/3								
		Ibiza-	1/2		2/3				1/2		
		Formentera	E		E		E		2/3		
08	Barcelona		1/2		2/5		2/5		1/5		
09	Burgos								**2/3**		
10	Cáceres								**1/2**		
11	Cádiz				2/3		2/3		1/2		
12	Castellón				2/3						
13	Ciudad Real								1/2		
15	A Coruña				1/2		2/3		1/3		
16	Cuenca								2/3		
17	Girona		1/2		2/5		2/5		1/3		
18	Granada		2/3		2/3		2/3		1/2		
20	Guipúzcoa						2/3		1/2		
22	Huesca				2/3		2/3		1/2		
24	León								1/2		
25	Lleida				2/3		2/3		1/2		
27	Lugo								1/2		
28	Madrid		1/2		2/3				1/5		
29	Málaga		1/2		1/2		2/3		1/2		
31	Navarra								1/3		
33	Asturias		1/2		3/10		1/2		1/2		
36	Pontevedra				1/2		2/5		2/5		
37	Salamanca								2/3		
38		Tenerife			1/2						

	S.C. Tenerife	La Palma				
		La Gomera				
		Hierro				
39	Cantabria			2/3	2/3	1/3
41	Sevilla			2/3		2/3
43	Tarragona			2/3	2/3	2/5
44	Teruel					1/2
45	Toledo					1/2
46	Valencia		2/3	2/3		
50	Zaragoza					1/2

The divisions of the Illes Balears and Santa Cruz de Tenerife are distinctions between summer (April - October) and winter (March -November), respectively.

Cell values indicate the stratum-sampling fraction. Blank values indicate that they are exhaustive.

For 1-silver-star category sample strata, rather than setting a sampling fraction, the number of items to be selected is determined:

Pro-Cod	Provinces	Islands	No. of elements	
01	Álava		12	
02	Albacete		24	
03	Alicante		30	
04	Almería		24	
05	Ávila		16	
06	Badajoz		24	
07	Baleares	Mallorca	16	24
		Menorca	8	8
		Ibiza	12	24
		Formentera	12	E
08	Barcelona		40	
09	Burgos		20	
10	Cáceres		24	
11	Cádiz		36	
12	Castellón		24	
13	Ciudad Real		20	
14	Córdoba		16	
15	A Coruña		60	
16	Cuenca		24	
17	Girona		24	
18	Granada		30	
19	Guadalajara		20	
20	Guipúzcoa		16	
21	Huelva		16	
22	Huesca		16	
23	Jaén		16	
24	León		36	

25	Lleida		24
26	La Rioja		24
27	Lugo		48
28	Madrid		56
29	Málaga		20
30	Murcia		12
31	Navarra		20
32	Ourense		30
33	Oviedo		30
34	Palencia		24
35	Las Palmas	Gran Canaria	8
		Lanzarote	4
		Fuerteventura	4
36	Pontevedra		32
37	Salamanca		30
38	S. C. de Tenerife	Tenerife	8
		La Palma	4
		La Gomera	4
		Hierro	4
39	Santander		36
40	Segovia		20
41	Sevilla		24
42	Soria		20
43	Tarragona		8
44	Teruel		8
45	Toledo		20
46	Valencia		24
47	Valladolid		20
48	Vizcaya		16
49	Zamora		16
50	Zaragoza		20
51	Ceuta		8
52	Melilla		4

3. Estimators

Computation of the HPI, base 2008

The improvements introduced do not alter the foundations of the methodology used in the computation of the HPI, that is:
- The formula used is a chain-linked Laspeyres index.
- Simple geometrical averages are used to aggregate the prices of the establishments.
- Different weightings are calculated for each month.

For the year 2009, the simple indices are obtained as the relationship between the average price of reference month m and the average price in the same month of the base year, for each province/category/client:

$$_{m(08)}I_{jkt}^{m(09)} = \frac{\overline{M}_{jkt}^{m(09)}}{\overline{M}_{jkt}^{m(08)}} \cdot 100 = \frac{average\ price\ of\ the\ client_t\ category_k\ province_j\ in\ month\ m\ of\ the\ year\ 2009}{average\ price\ of\ the\ client_t\ category_k\ province_j\ in\ month\ m\ of\ the\ year\ 2008} \cdot 100$$

In order to compute the HPI, base 2008, in month m, of year 2009, the weighted simple indices are aggregated.

The weightings are calculated on a level of province, establishment category and type of client, that is, at the same level of detail as the simple indices, and they represent the percentage of income received by the hotels for the rooms occupied, by type of client, in a category and in a province, over the total income:

$$_{m(08)}L_{jkt} = \frac{\overline{M}_{jkt}^{m(08)} \cdot \widehat{B}_{jkt}^{m(08)}}{\sum_{j,k,t} \overline{M}_{jkt}^{m(08)} \cdot \widehat{B}_{jkt}^{m(08)}}$$

where $\widehat{B}_{jkt}^{m(08)}$ is the estimation of the total number of rooms occupied of the establishments of category k and province j, sold to type of client t, in month m, of base year 2008.

$$\widehat{B}_{jkt}^{m(08)} = \left(\sum_{i=1}^{e_{jk}} B_{ijk}^{m(08)} \cdot A_{ijkt}^{m(08)} \right) \cdot f_{jk}^{m(08)}$$

where:

. $B_{ijk}^{m(08)}$, is the rooms occupied in establishment i, of category k, of province j, in reference period $m(08)$.

. $A_{ijkt}^{m(08)}$, is the percentage of rooms sold to type of client t, in establishment i, of category k, of province j, in period $m(08)$.

. $f_{jk}^{m(08)}$, is the elevation factor in stratum jk, calculated as the quotient of the population rooms available of stratum jk, among the rooms available of the sample in that same stratum, in period $m(08)$.

. e_{jk}, represents the group of establishments of the sample of stratum jk, which have responded to the questionnaire, in period $m(08)$.

These weights are aggregated subsequently (*W*) by category, type of client, province or Autonomous Community, according to the aggregated index desired.

Thus, for example, in order to computer the HPI on a national level:

$$_{m(08)}IPH^{m(09)} = \sum_{j}\left(\sum_{k}\left(\sum_{t} {}_{m(08)}I_{jkt}^{m(09)} \cdot {}_{m(08)}W_{jkt} \right) {}_{m(08)}W_{jk} \right) \cdot {}_{m(08)}W_{j}$$

where

$$_{m(08)}W_{jkt} = \frac{\overline{M}_{jkt}^{m(08)} \cdot \widehat{B}_{jkt}^{m(08)}}{\sum_{t} \overline{M}_{jkt}^{m(08)} \cdot \widehat{B}_{jkt}^{m(08)}} \qquad _{m(08)}W_{jk} = \frac{\sum_{t} \overline{M}_{jkt}^{m(08)} \cdot \widehat{B}_{jkt}^{m(08)}}{\sum_{k,t} \overline{M}_{jkt}^{m(08)} \cdot \widehat{B}_{jkt}^{m(08)}} \qquad _{m(08)}W_{j} = \frac{\sum_{k,t} \overline{M}_{jkt}^{m(08)} \cdot \widehat{B}_{jkt}^{m(08)}}{\sum_{j,k,t} \overline{M}_{jkt}^{m(08)} \cdot \widehat{B}_{jkt}^{m(08)}}$$

The index expressed is a pure Laspeyres index, and as of 2010, a chain-linked Laspeyres index will be used in the computation of the HPI. This thereby guarantees the continuous updating of the structure used in the weighting of the indices, as the weightings are always calculated with data referring to the immediately preceding year. In order to ensure the comparability of the data obtained with different structures, a chain-linked index is used, which eliminates the need to calculate linking coefficients for each update carried out.

The use of a chain-linked Laspeyres index does have some inconveniences, such as:

. The lack of additivity: it is not possible to obtain a chain-linked index from any aggregate as the weighted average of the chain-linked indices that comprise it.
. Loss of inter-monthly comparison: this comparison is no longer possible, given that each month, different structures have been used for its computation. Only interannual variation rates can be calculated for the HPI to analyse the evolution of hotel prices.

As of January 2010, the HPI will be computed using the following mathematical expression (the national index as a category is used as an example), for any T greater than or equal to 2010:

$$_{m(08)}IPH_{k}^{mT} = 100 \cdot \prod_{b=2009}^{T} \frac{_{m(b-1)}IPH_{k}^{sb}}{100}$$

being the general expression for the calculation of the weightings

$$_{m(T-1)}L_{jkt} = \frac{\overline{M}_{jkt}^{m(T-1)} \cdot \widehat{B}_{jkt}^{m(T-1)}}{\sum_{i}\sum_{j}\sum_{t} \overline{M}_{jkt}^{m(T-1)} \cdot \widehat{B}_{jkt}^{m(T-1)}}$$

4. Base change 2008-2001

Base change. Linked series.

Due to this base change, it is necessary to calculate linking coefficients that avoid the break in the series already published in base 2001. These coefficients are calculated for the aggregates published, maintaining the interannual variation rates disseminated.

With the base change, all of the indices of the year 2008, base 2008, are equal to 100. By retroactively applying the interannual variation rates published month by month, the series is reconstructed to January 2001, with base 2008.

$$IPH_{m,07}^{(linked)Base\,2008} = \frac{IPH_{m,08}^{Base\,2008}}{1+\dfrac{TV_m^{07,08}}{100}} \qquad\qquad IPH_{m,a}^{(linked)Base\,2008} = \frac{IPH_{m,a+1}^{(linked)Base\,2008}}{1+\dfrac{TV_m^{a,a+1}}{100}}$$

This avoids the break caused by the new HPI system, and achieves the comparability of the series.

The complete series of the HPI in base 2001 (until December 2008) and in base 2008 are available to the user on the INE website (http://www.ine.es/en/).

5. Statistical Secrecy

It could be provide information about all those stratum (or geographical data groups) where the number of establishments with an incidence of 1 (open with movement) is equal to or greater than 5.

6. Variation quotients

In order to calculate a variance estimation of the Hotel Price Index we apply the JACKKNIFE technique. The elementary indexes are calculated as quotients of geometric averages, and then they are added by using weightings.
We have 8 price fees (t) per client or establishment (i), category (k) and province (j). Out of them 8 elementary indexes are built this way:

$$_{m(14)}I_{jkt}^{m(15)} = 100 \times \frac{\sqrt[n_{jk}]{p_{jkt1}^{m(15)} \times p_{jkt2}^{m(15)} \times ... \times p_{jktn_{jk}}^{m(15)}}}{\sqrt[n_{jk}]{p_{jkt1}^{m(14)} \times p_{jkt2}^{m(14)} \times ... \times p_{jktn_{jk}}^{m(14)}}}$$

Where n_{jk} corresponds to the number of clients or establishments of a certain t fee of the k category and j province.
For each fee and each stratum made up by province and category the Jackknife variance is calculated as follows:

$$\hat{V}(\hat{\theta}) = \sum_{h:1}^{L} \frac{n_h - 1}{n_h}(1 - f_h) \times \sum_{k:1}^{n_h} (\hat{\theta}_{h(k)} - \hat{\theta}_{h(.)})^2$$

where the $\hat{\theta}_{h(k)}$ subindex indicates that unit k is left out and $\hat{\theta}_{h(.)} = \dfrac{\sum_{k:1}^{n_h} \hat{\theta}_{h(k)}}{n_h}$

The variation quotient given by the following expression is published:

$$\hat{C}V(\hat{\theta}) = \frac{\sqrt{\hat{V}(\hat{\theta})}}{\hat{\theta}} \times 100$$

When analised the results, which are provided as %, it is important to bear in mind the following things:
1. It is a rough calculation of the sampling error;
2. As it is a survey analysing exhaustively a great part of the population, sampling errors are inexistent or almost negligible. In this type of surveys non-sampling errors prevail;
3. The calculation of the variance is carried out in strata with 5 or more sample units.

7. Data collection. XML tourism

Information will be collected by means of the Provincial INE delegations specialising in tourist occupancy surveys. These are those of Pontevedra, Cantabria, Álava, Huesca, Girona. Illes Balears, Castellón, Málaga, Cáceres, Toledo, S. C. de Tenerife, Asturias and Valencia.

The regular data collection procedure employed has been by post, with telephone and fax support. The informant may also complete and submit the survey online (ARCE System). Alternatively, and so long as the informant requests or approves it, electronic questionnaires may also be used, submitted and/or received by email.

For the HOS, a new optional telematic information coleection via XML files system has also been implemented, basically consisting of hotel establishment case-handling software extracting from hotel database records all information required in the questionnaire with regard to guests, overnight stays, occupied rooms and days open.

All the information about this new system of data compilation is available in the following link, where also the establishment upload the files:

https://arce.ine.es/ARCE/jsp/encuestaXml.jsp

The sample selected is monthly. Nevertheless, for the purposes of the questionnaires used, the sample is distributed weekly. The variables relating to prices and to occupied rooms are included in the weekly questionnaire, the question being as follows:

6. Tarifa Promedio Diaria (ADR) por tipo de cliente (en euros, no incluye IVA)

Indique el ADR (Average Daily Rate o Tarifa Promedio Diaria) para cada tipo de cliente por día incluyendo sólo alojamiento –sin desayuno, media pensión o pensión completa- **de una habitación doble con baño.**
Señale también el **porcentaje aproximado de habitaciones ocupadas por cada tipo de cliente** sobre el total **– el mismo del apartado 4.3-** en la semana de referencia (este porcentaje puede ser cero en algún caso).

	ADR en euros	%[2]
Tour operador tradicional		
Agencia de viajes tradicional[1] (incluye bonos y talones)		
Empresas		
Particulares (incluye tarifa normal y tarifa fin de semana)		
Grupos		
Contratación directa en la web del hotel o en la cadena hotelera		
Tour operador on-line		
Agencia de viajes on-line		
Otros		

[1] *Deben considerar en este tipo de cliente los huéspedes que han entregado* **bonos y/o talones de hoteles** *como contraprestación por los servicios prestados.*   **100 %[2]**
[2] *Este porcentaje será cero si no ha habido habitaciones ocupadas durante la semana de referencia.*

The HOS use two weekly questionnaire models (see annex), the fundamental difference being the list of countries of origin of guests and overnight stays. For hotels in higher categories (in other words, 3, 4 and 5 gold stars) the list includes more countries than for the remainder of categories. This does not influence the question of prices, which is identical in both cases.