CAPITULO III

ESTIMACIÓN DE LA DEMANDA HOTELERA.

CAPITULO III

ESTIMACIÓN DE LA DEMANDA HOTELERA.

1. Introducción.

Para una lavandería hotelera, la ocupación de los hoteles dicta la producción diaria de ropa que hay que tratar. Por tanto será bueno conocer como se comporta la ocupación o demanda hotelera para poder así diseñar y desarrollar de la mejor manera posible, las actuaciones ante cualquier eventualidad o circunstancia.

También el conocimiento del comportamiento de la ocupación será muy importante para el estudio de la viabilidad del proyecto, puesto que permitirá dimensionar el tamaño mínimo necesario de la futura lavandería mediante el diseño en planta de la misma, a su vez será necesario conocer el tipo y número de maquinarias a invertir. Ante todos estos condicionantes se hace necesario el estudio de la ocupación hotelera.

2. Predicción.

El conocimiento del comportamiento de la ocupación de los hoteles se basará en los sucesos pasados, datos históricos, de tal manera que permitan predecir los sucesos en el futuro. El proceso de predicción comienza con la definición de las variables que intervienen y las relaciones esperadas entre las variables.

La ocupación diaria en un instante depende de la ocupación en el estado anterior y la suma de la diferencia entre entradas y salidas:

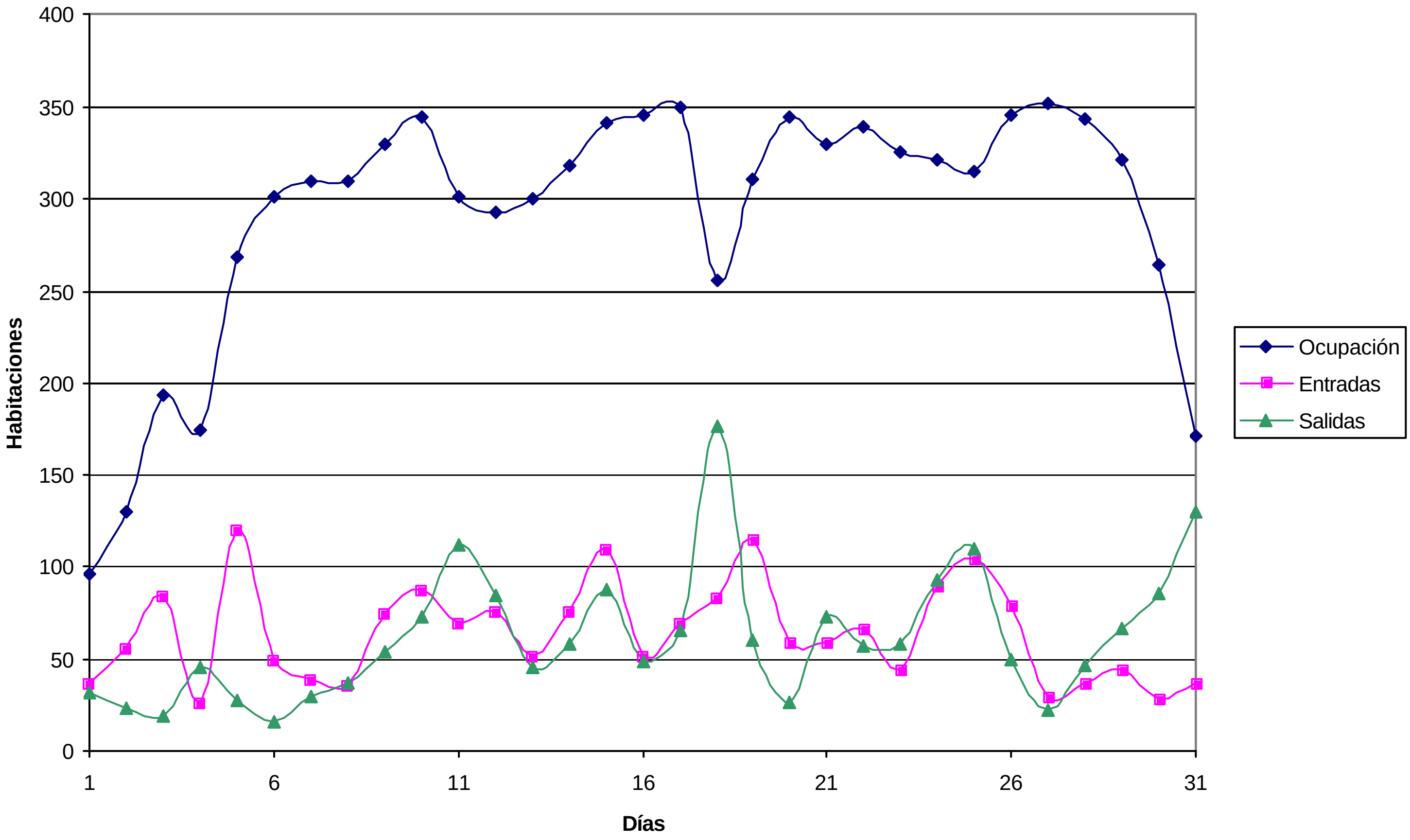
$$O(k)=O(k-1)+E(k)-S(k)$$

Siendo O(k) el número de habitaciones ocupadas en el instante de estudio, O(k-1) el número de habitaciones ocupadas en el instante anterior, E(k) y S(k) las entradas y salidas de habitaciones en el instante de estudio respectivamente.

A su vez O(k-1)=O(k-2)+E(k-1)-S(k-1) ; O(k-2)=O(k-3)+E(k-2)-S(k-2) … y así sucesivamente hasta llegar al instante inicial en el que se considera que el hotel está vacío. Por tanto:

$$O(k) = \sum_{i=1}^{i=k} (E(i) - S(i))$$

Así por ejemplo la ocupación diaria, con las entradas y salidas en un mes cualquiera tendría el siguiente aspecto:



Se puede observar en el gráfico como las entradas y salidas se comportan de manera parecida. Hay que hacer notar que los picos de las entradas y salidas se producen en los fines de semana.

La ocupación dada anteriormente es la ocupación diaria, si bien no es necesario para el estudio a realizar tal grado de detalle y afinamiento, por lo tanto bastará con la ocupación mensual. Esta se calculará como la suma de las habitaciones diarias ocupadas al cabo del mes.

Para elaborar pronósticos se pueden encontrar dos grandes clases de modelos: causales y de series de tiempo. Los primeros tratan de encontrar las relaciones de causalidad entre diferentes variables, de manera que conociendo o prediciendo alguna o algunas de ellas, se pueda encontrar el valor de otra. En el segundo caso no interesa encontrar esas relaciones, sino que se requiere solamente encontrar los posibles valores que asumirá una determinada variable. En todos los casos siempre se hace uso de la información histórica, ya sea para predecir el comportamiento futuro o para suponer que el comportamiento histórico se mantendrá hacia el futuro y sobre esta base hacer los estimativos.

Se debe tener presente que no existe ningún método de pronóstico infalible; lo que logran estos procedimientos es estimar un valor posible, pero siempre sujeto a errores. Si el fenómeno que se va a pronosticar fuera determinístico, solo bastaría utilizar la ley matemática que lo rige y predecir con exactitud el resultado; este sería el caso de fenómenos físicos, como por ejemplo la caída libre de un cuerpo.

Una vez decidido cual va a ser la variable a estudiar y como viene relacionada con otras variables es cuando comienzan los problemas, debido a la falta de datos históricos suficientes para poder construir un modelo de inferencia estadística para hacer los pronósticos. Ante esta situación es necesario aplicar el criterio, fruto de la experiencia, para predecir lo que ocurrirá con la variable. Por tanto lo que se intentará será determinar cual será el intervalo de confianza donde se encuentra la ocupación en ese instante.

En la siguiente tabla se muestran los datos históricos conocidos para cada uno de los hoteles.

	HOTEL 1 (389 HAB)		HOTEL 2 (335 HAB)		HOTEL 3 (389 HAB)	
	Habitaciones	**Ocupación**	**Habitaciones**	**Ocupación**	**Habitaciones**	**Ocupación**
ene-01	1832	15,19%				
feb-01	2054	18,86%				
mar-01	2125	17,62%				
abr-01	3221	27,60%				
may-01	2876	23,85%				
jun-01	5678	48,65%				
jul-01	7085	58,75%				
ago-01	8467	70,21%				
sep-01	5431	46,54%				
oct-01	3421	28,37%				
nov-01	1645	14,10%				
dic-01	2153	17,85%				
ene-02	1148	9,52%				
feb-02	1690	14,98%				
mar-02	3354	27,81%				
abr-02	1805	15,47%				
may-02	2197	18,22%	1885	18,15%		
jun-02	6086	52,15%	6256	62,25%		
jul-02	8582	71,17%	6959	67,01%	8719	72,30%
ago-02	8720	72,31%	7329	70,57%	9213	76,40%
sep-02	6342	54,34%	5874	50,33%	6512	55,80%
oct-02	4011	33,26%	3952	32,77%	4328	35,89%
nov-02	1632	13,98%			2692	23,07%
dic-02	2485	20,61%			3672	30,45%

En la siguiente tabla se muestran los datos históricos de ocupación, en tanto por ciento, del Instituto Nacional de Estadística para los años 1999, 2000 y 2001 de la Costa de Almería, que la componen los municipios de Adra, Almería, Carboneras, Cuevas del Almanzora, Garrucha, Mojácar, Níjar, Pulpí, Roquetas de Mar, Vera y El Ejido.

	1999	**2000**	**2001**
Enero	55,49	52,43	40,8
Febrero	58,22	66,29	53,51
Marzo	60,95	70,22	52,56
Abril	64,96	63,94	61,45
Mayo	65,15	69,29	53,64
Junio	71,27	66,34	62,24
Julio	78,88	71,75	68,43
Agosto	87,75	80,98	73,7
Setiembre	75,6	76,83	63,92
Octubre	64,03	66,63	54,73
Noviembre	50,25	56,73	37,5
Diciembre	46,02	43,06	33,08

Estos datos son puramente a título informativo ya que albergan el grado de ocupación de la Costa de Almería para cualquier categoría de hoteles en el conjunto de la costa. Además se puede observar un descenso significativo en el grado de ocupación al cabo del tiempo para un mes dado, si bien esto es debido al gran aumento del número de habitaciones en la zona en los últimos años.

3. Grado de estacionalidad.

Con estos datos se puede observar también como la Costa de Almería presenta una fuerte estacionalidad. La estacionalidad es un problema para cualquier complejo turístico, en la medida en que provoca que en determinados meses la capacidad hotelera esté infrautilizada (temporada baja), mientras que en otros meses esté utilizada prácticamente en su totalidad (temporada alta). Su importancia es tal que en la medida en que se logren paliar sus efectos, se podría lograr un turismo más o menos homogéneo, lo cual solucionaría problemas de rentabilidad de las infraestructuras, inestabilidad en el empleo, masificación, etc... Además, el fenómeno de la estacionalidad se puede presentar en dos vertientes: oferta (empleo) y demanda (turistas).

El grado de estacionalidad se puede medir mediante el índice de Gini:

$$I.G. = 1 + \frac{1}{N} - \frac{2}{N^2 \overline{Y}}\left(Y_1 + 2Y_2 + \ldots + NY_n\right)$$

Donde Y1, Y2,..., Yn son cada una de las observaciones en orden decreciente.

El valor que arroja el índice de Gini oscila entre 0 y 1. Cuanto más bajo sea el índice, será un indicativo de que la corriente de turistas se distribuye de forma más igualitaria a lo largo del año, es decir, la estacionalidad será menor. Por el contrario, cuanto más alto sea el índice, la estacionalidad será más elevada.

En el caso que se estudia el I.G. para los dos primeros años es de 0,3051 y 0,3575 respectivamente, lo que indica que se está ante una fuerte estacionalidad.

Esta estacionalidad provoca que en el caso de una lavandería, el no estudiar correctamente como se distribuyen los kilogramos de ropa a tratar en el año, se pueda cometer el error de sobredimensionar e infrautilizar la lavandería con el perjuicio que esto provocaría. Además se podrá gestionar mas eficazmente ya que se podrá prever cual será la cantidad de materias primas y empleados necesarios para el funcionamiento de la misma al estimar como se comportará la demanda.

4. Intervalos de confianza.

Ante estas cuestiones que son la falta de datos históricos y la fuerte estacionalidad, se utilizarán criterios basados en la experiencia para predecir lo que ocurrirá con la variable. Por tanto lo que se intentará será determinar cual será el intervalo de confianza donde se encuentra la ocupación en ese instante.

Así aplicando el teorema central del límite:

Sea *X* una variable aleatoria cualquiera de media μ y desviación típica s, entonces:

Si el tamaño muestral *n* es “suficientemente grande” (en la práctica suele valer *n>30*), la distribución de las medias muestrales se aproxima a la de una normal, es decir:

$$\overline{X} \approx N(m, \frac{s}{\sqrt{n}})$$

La importancia del teorema central del límite radica en que sea cuál sea la distribución de la población original (variable aleatoria *X*), conforme el tamaño de las muestras (*n*) aumenta, la distribución de las medias se va aproximando a la de una normal (de la cual conocemos muchas propiedades).

Así, si la población tiene una distribución de probabilidad normal, entonces, para cualquier tamaño de muestra la distribución del muestreo de la media también tendrá una distribución normal. Si la distribución de la población es simétrica (pero no normal), se verá que surge la forma normal como lo establece el teorema central del límite aún con muestras de al menos 30 para observar las características de normalidad.

El teorema de Tchebyshev dice que:

Si X ~ N(μ,s) , entonces:

a) $P(\mu-s<X<\mu+s) = 0,68$
b) $P(\mu-2s<X<\mu+2s) = 0,95$
c) $P(\mu-3s<X<\mu+3s) = 0,997$

Es decir, el 68% (aproximadamente) de los valores que tome la variable aleatoria X estarán situados a una distancia de la media inferior a una desviación estándar. Análogamente, el 95% de los valores estarán situados a menos de 2 veces la desviación estándar, y un 99,7% de dichos valores se encontrarán dentro un radio de 3 sigma. Concretamente el 95% de los valores estarán situados a 1,96 veces la desviación estándar.

Por lo tanto, para una distribución normal, la mayor parte de todos los valores yacen a tres desviaciones estándar de la media.

Así pues, dada una población X, que sigue una distribución cualquiera con media μ y desviación estándar s.

1. Sabemos (por el teorema central del límite) que, para valores grandes de n, la media muestral $\overline{X}$ sigue una distribución aproximadamente normal con media $\mu_x = \mu$ y desviación estándar $s_x = s / \sqrt{n}$

2. Por otra parte, el Teorema de Tchebyshev nos dice que, en una distribución normal, aproximadamente un 95% de los datos estaban situados a una distancia inferior a 1,96 desviaciones estándar de la

media.

De lo anterior se deduce que:

$$P(m-1{,}96s_x < \overline{x} < m+1{,}96s_x) = 0{,}95$$
$$0{,}95 = P(\overline{x} < m+1{,}96s_x) - P(\overline{x} < m-1{,}96s_x) = P(m > \overline{x}-1{,}96s) - P(m > \overline{x}+1{,}96s)$$
$$P(\overline{x}-1{,}96s < m < \overline{x}+1{,}96s) = 0{,}95$$

Por tanto, ésta última fórmula nos da un intervalo de valores tal que la probabilidad de que la media de la población µ esté contenida en él es de 0,95.

Este tipo de intervalos se llaman intervalos de confianza de un parámetro poblacional. El nivel de confianza (1 - a) del intervalo es la probabilidad de que éste contenga al parámetro poblacional.

Como dice el teorema central del límite son necesarias al menos muestras de tamaño 30 para aproximar la variable aleatoria a una distribución normal. En nuestro caso no se dispone de muestras suficientemente grandes, pero aproximaremos la media muestral de la variable aleatoria $\overline{X}_i$ que representa el grado de ocupación en el mes i, mediante una distribución normal. Esta distribución normal nos servirá para calcular un intervalo de confianza con un nivel de confianza del 95%.

Así $\overline{X}_i \sim N(\mu_x=\mu, S_x=S/\sqrt{n})$, para estimar la desviación estándar poblacional s vamos a utilizar la desviación estándar muestral S de un mes cualquiera. Por ejemplo para el mes de agosto. Al no tener suficientes datos históricos se utilizarán, además de los datos que se tienen del grado de ocupación de cada hotel para el mes de agosto, los datos del Instituto Nacional de Estadística para la costa de Almería.

	Hotel 1 01	Hotel 1 02	Hotel 2	Hotel 3	INE 1999	INE 2000	INE 2001
Grado de ocupación	0,7021	0,7231	0,7057	0,764	0,8775	0,8098	0,737

La media muestral vale $\overline{X}_8$ = 0,7599 y la desviación estándar muestral S = 0,0639. S_x = 0,0241.

Utilizamos la t-Student porque la desviación estándar poblacional s es desconocida.

$$\overline{X} \pm t(n-1, a/2) \cdot \frac{S}{\sqrt{n}} = 0{,}7599 \pm t(6{,}0{,}05/2) \cdot 0{,}0241 = 0{,}7599 \pm 2{,}447 \cdot 0{,}0241 = 0{,}7599 \pm 0{,}0590 =$$
$$= (0{,}7009 \ ; \ 0{,}8189)$$

Con una confianza del 95% el grado de ocupación medio del mes de agosto estará entre un 70,09% y un 81,89%.

	Hotel 1 01	Hotel 1 02	Hotel 2	Hotel 3	INE 1999	INE 2000	INE 2001
Enero	0,1519	0,0952			0,5549	0,5243	0,408
Febrero	0,1886	0,1498			0,5822	0,6629	0,5351
Marzo	0,1762	0,2781			0,6095	0,7022	0,5256
Abril	0,276	0,1547			0,6496	0,6394	0,6145
Mayo	0,2385	0,1822	0,1815		0,6515	0,6929	0,5364
Junio	0,4865	0,5215	0,6225		0,7127	0,6634	0,6224
Julio	0,5875	0,7117	0,6701	0,723	0,7888	0,7175	0,6843
Agosto	0,7021	0,7231	0,7057	0,764	0,8775	0,8098	0,737
Septiembre	0,4654	0,5434	0,5033	0,558	0,756	0,7683	0,6392
Octubre	0,2837	0,3326	0,3277	0,3589	0,6403	0,6663	0,5473
Noviembre	0,141	0,1398		0,2307	0,5025	0,5673	0,375
Diciembre	0,1785	0,2061		0,3045	0,4602	0,4306	0,3308

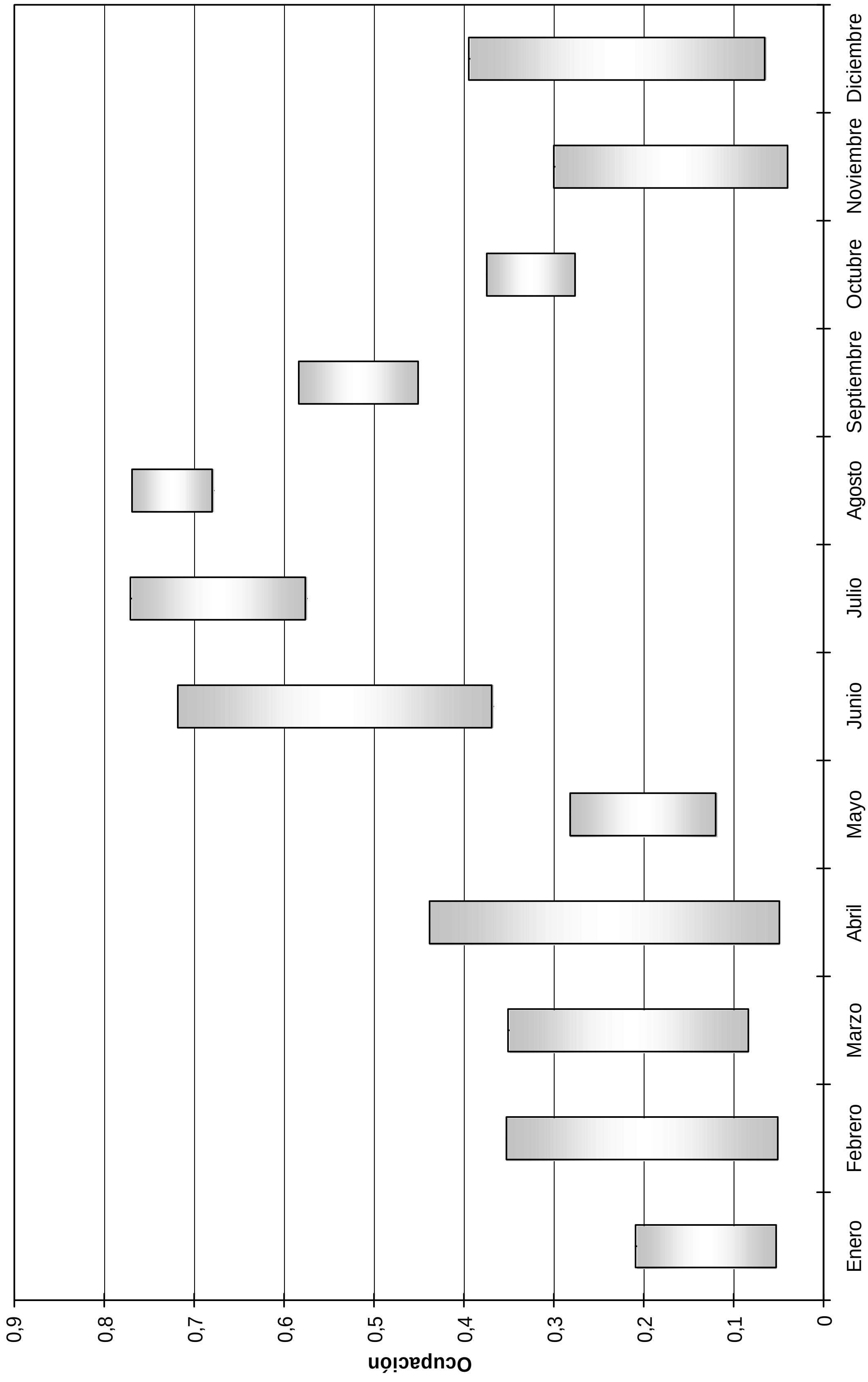
Si realizamos estas mismas operaciones con los restantes meses, obtendremos un intervalo de confianza del 95 % del grado de ocupación medio de cada mes. Uniendo ambas tablas, se obtienen los datos siguientes:

	Mínimo	Media	Máximo	Varianza
Enero	0,0836	0,3469	0,6101	0,1343
Febrero	0,1292	0,4237	0,7182	0,1503
Marzo	0,1814	0,4583	0,7352	0,1413
Abril	0,1764	0,4668	0,7573	0,1482
Mayo	0,1621	0,4138	0,6656	0,1285
Junio	0,515	0,6048	0,6947	0,0459
Julio	0,6408	0,6976	0,7543	0,0289
Agosto	0,7008	0,7599	0,819	0,0302
Septiembre	0,4938	0,6048	0,7158	0,0566
Octubre	0,3013	0,451	0,6007	0,0764
Noviembre	0,1327	0,326	0,5194	0,0987
Diciembre	0,1986	0,3185	0,4383	0,0611

Si bien los datos obtenidos para los meses de invierno están falseados, ya que no es lo que ocurre en la realidad, debido a los altos valores de ocupación que da el INE. Si despreciamos los datos del INE se obtienen los siguientes intervalos con el 95% de nivel de confianza.

	Mínimo	**Media**	**Máximo**	**Varianza**
Enero	0,0534	0,1313	0,2091	0,0397
Febrero	0,0519	0,2023	0,3527	0,0767
Marzo	0,0844	0,2175	0,3507	0,0679
Abril	0,0494	0,244	0,4385	0,0993
Mayo	0,1195	0,2007	0,282	0,0414
Junio	0,3681	0,5435	0,7189	0,0895
Julio	0,5753	0,6731	0,7708	0,0499
Agosto	0,6786	0,7237	0,7689	0,0230
Septiembre	0,4511	0,5175	0,5839	0,0339
Octubre	0,2761	0,3257	0,3754	0,0253
Noviembre	0,041	0,1705	0,3	0,0661
Diciembre	0,0652	0,2297	0,3942	0,0839

Si se comparan ambas tablas las variaciones de enero a mayo son muy significativas. Por tanto si tomamos los valores de esta última tabla y los representamos gráficamente, los intervalos de confianza de la ocupación quedan como sigue:


Enero
Febrero
Marzo
Abril
Mayo
Junio
Julio
Agosto
Septiembre
Octubre
Noviembre
Diciembre
0
0,1
0,2
0,3
0,4
0,5
0,6
0,7
0,8
0,9
Ocupación

5. Series temporales.

También se podría haber realizado todo esto mediante los modelos de series temporales.

Una serie temporal es un conjunto de observaciones ordenadas en el tiempo, que pueden representar la evolución de una variable (económica, física, etc.) a lo largo de él.

El objetivo del análisis de una serie temporal es el conocimiento de su patrón de comportamiento, para así prever su evolución futura, suponiendo que las condiciones no variarán.

Dado que no se trata de fenómenos deterministas, sino sujetos a una aleatoriedad, el estudio del comportamiento pasado ayuda a inferir la estructura que permita predecir su comportamiento futuro, pero es necesaria una gran cautela en la previsión debido a la inestabilidad del modelo.

La particular forma de la información disponible de una serie cronológica (se dispone de datos en periodos regulares de tiempo) hace que las técnicas habituales de inferencia estadística no sean válidas para estos casos, ya que nos encontramos ante n muestras de tamaño 1 procedentes de otras tantas poblaciones de características y distribución desconocidas.

Un método de pronóstico para analizar series de tiempo es el de descomposición. Un paso importante en el proceso de determinar el método de series de tiempo adecuado, es considerar los diferentes patrones que se encuentran en los datos. Se pueden identificar cuatro patrones típicos: horizontal o estacionaria, estacional, cíclico y de tendencia.

- Se presenta un patrón horizontal o estacionario (*H*) cuando los datos fluctúan alrededor de un valor promedio constante. Las ventas que no aumentan ni disminuyen con el tiempo, son un ejemplo de este tipo de comportamiento.

- Se presenta un patrón estacional (*E*) cuando los datos están afectados por factores que se repiten con cierta frecuencia (trimestral, mensual o en determinadas fechas, por ejemplo, verano, Navidad, Semana Santa, etc.).

- Un patrón cíclico (*C*) se presenta debido a efectos económicos de largo plazo y generalmente asociados con el ciclo económico. La construcción de vivienda puede ser un ejemplo de este tipo.

- Existe un patrón de tendencia (*T*) cuando existe un aumento o disminución temporal de los datos. Las ventas de la mayoría de las firmas presentan este comportamiento.

Los métodos de descomposición suponen que los datos contienen patrones estacionales, cíclicos y de tendencia; una función que representa esta relación puede ser la siguiente:

dato = patrón + error = *f(tendencia, estacionalidad, ciclo) + error.*

$X_t = f(T_{t,}, E_{t,}, C_{t,}, Er_t)$

Donde:

X_t es el dato al período t.

T_t es el componente de tendencia en el período t.

E_t es el componente o índice de estacionalidad del período t.

C_t es el componente cíclico del período t.

y Er_t es el error del período t.

Descomposición.

El procedimiento general para aislar los diversos componentes es el siguiente y se aplica a los diferentes métodos de descomposición.

1) Con los datos disponibles se calcula el promedio con un *N* igual a la longitud de la estacionalidad (12 meses, 4 trimestres, o 7 días, por ejemplo). Con esto se elimina la estacionalidad y el error, por lo tanto en el promedio móvil se encuentra sólo la tendencia y el ciclo.

2) Se separa el promedio móvil de los datos. Lo que queda es la estacionalidad y el error.

3) Se aíslan los factores estacionales que constituyen el período completo de estacionalidad (cada mes, semestre o trimestre, por ejemplo) promediándolos para cada período.

4) Se identifica la forma de la tendencia con los resultados de 1) (lineal, exponencial, cuadrático, etc.) y se calcula su valor para cada uno de los períodos para los cuales se tienen datos.

5) Se separa el resultado de 4) de los resultados de 1) para obtener el factor cíclico.

6) Se separa la estacionalidad, la tendencia y el ciclo de los datos para obtener el error.

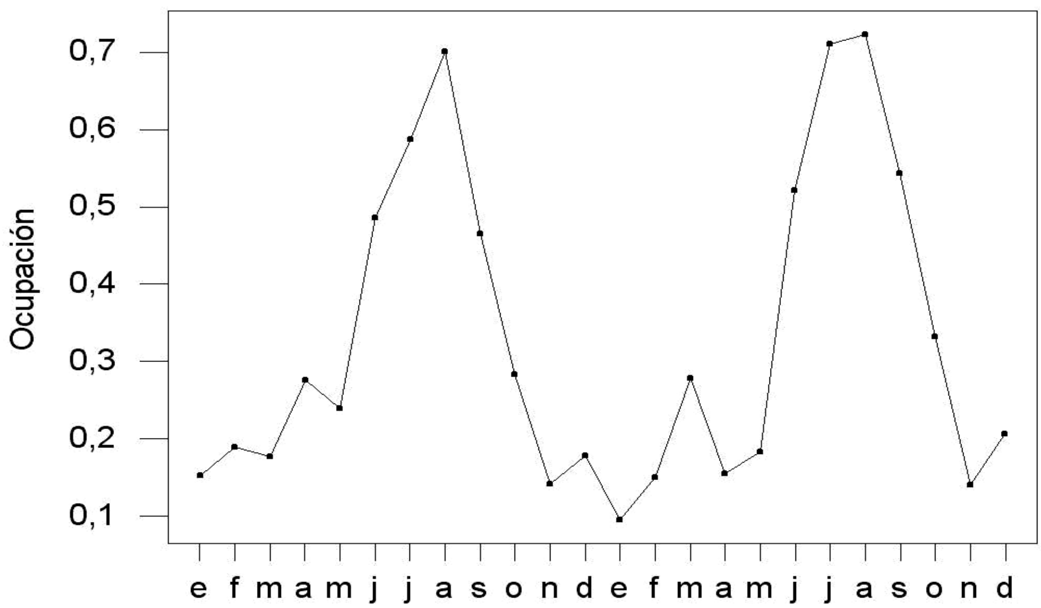
Para realizar los pasos anteriores habrá primero que definir un tipo de modelo a utilizar en la descomposición. Existen dos modelos el aditivo y el multiplicativo:

Modelo multiplicativo: X_t = Tendencia • Estacionalidad • Ciclo • Error

Modelo aditivo: X_t = Tendencia + Estacionalidad + Ciclo + Error

Las tres primeras componentes son determinísticas, llamadas señales, mientras el error, también llamado ruido, es una variable aleatoria. Para poder realizar una estimación apropiada debemos conocer que cantidad de cada componente hay en los datos. Así, para comprender y medir estos componentes, el proceso de estimación requiere inicialmente el quitar los efectos de cada componente de los datos (descomposición), para posteriormente unirlos para realizar la estimación.

Normalmente, la mejor forma de comenzar a analizar los datos de una serie temporal es representar las observaciones frente al tiempo a fin de detectar tendencias, patrones estacionarios, y outliers (datos que no concuerdan con la mayoría).



Se puede observar como los datos tienen una estacionalidad para los meses de verano, al igual que para fechas como la Semana Santa y Navidad. No existe ningún patrón claro estacionario y la tendencia parece que aumenta con el tiempo si bien muy poco a poco.

Para decidir que método tomamos el aditivo o el multiplicativo podemos utilizar el siguiente método:

La diferencia estacional se obtiene mediante la expresión $d = X_t – X_{t-L}$ y el coeficiente estacional mediante $c = X_t / X_{t-L}$ donde *L* representa el número de periodos dentro de un año cualquiera de la serie. Por ejemplo, si la serie fuera mensual, *L* sería igual a 12. Por otro lado, si la serie fuera trimestral, *L* sería 4. Finalmente, si la serie fuera anual, *L* sería igual a 1. En nuestro caso L = 12

Desde un punto de vista teórico, los pasos para la aplicación de este método son:

1) Cálculo de todos los cocientes y diferencias estacionales. Lógicamente perderíamos las observaciones correspondientes al primer año.

2) Obtener el coeficiente de variación (V), cociente entre la desviación típica y la media, para las diferencias y cocientes estacionales. Los llamaremos V(d) y V(c) respectivamente.

3) Aplicamos la siguiente regla de decisión:

   - Si V(d) < V(c), elegiríamos el modelo aditivo.

   - Si V(d) > V(c), elegiríamos el modelo multiplicativo.

La justificación del método es la siguiente: la obtención de las diferencias estacionales equivale a tomar la serie de incrementos interanuales. Los cocientes estacionales, sin embargo, tiene más que ver con la serie de crecimiento. Por lo tanto, si el crecimiento interanual para cada estación tiene menor variabilidad que en términos de incrementos, esto indicaría una asociación multiplicativa. Si sucediera lo contrario, tomaríamos la hipótesis aditiva. El utilizar el coeficiente de variación se justifica por ser una medida adimensional que facilita las comparaciones.

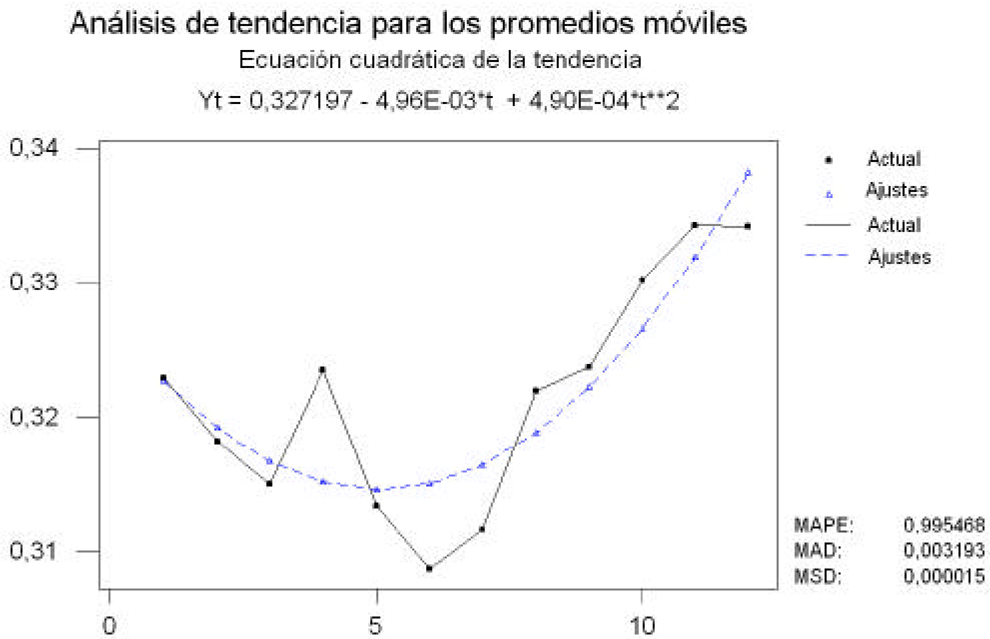
En nuestro caso V(d) = -0,3839 y V(c) = 0,1096, por lo que elegiremos el modelo aditivo.

Aplicando el proceso de descomposición al modelo aditivo:

Promedios	**E(t)+Er(t)**
0,322991667	-0,227791667
0,318266667	-0,168466667
0,315033333	-0,036933333
0,323525	-0,168825
0,313416667	-0,131216667
0,308725	0,212775
0,311641667	0,400058333
0,321991667	0,401108333
0,323741667	0,219658333
0,330241667	0,002358333
0,334316667	-0,194516667
0,334216667	-0,128116667

Realizando el paso 1 para N=12 se obtiene sólo la tendencia y el ciclo, además realizando el paso 2 queda la estacionalidad y el error:

Dibujando los promedios móviles obtendré la tendencia y se puede observar que el método que mejor se aproxima es el cuadrático. Sus errores son muy pequeños.



Siendo MAPE el error absoluto medio porcentual, MAD la desviación media absoluta y MSD la desviación cuadrática media:

- $$\text{MAPE} = \frac{\Sigma\left|(y_t - \hat{y}_t)/y_t\right|}{n} \times 100 \qquad (y_t \neq 0)$$

- $$\text{MAD} = \frac{\Sigma\left|y_t - \hat{y}_t\right|}{n}$$

- $$\text{MSD} = \frac{\Sigma(y_t - \hat{y}_t)^2}{n}$$

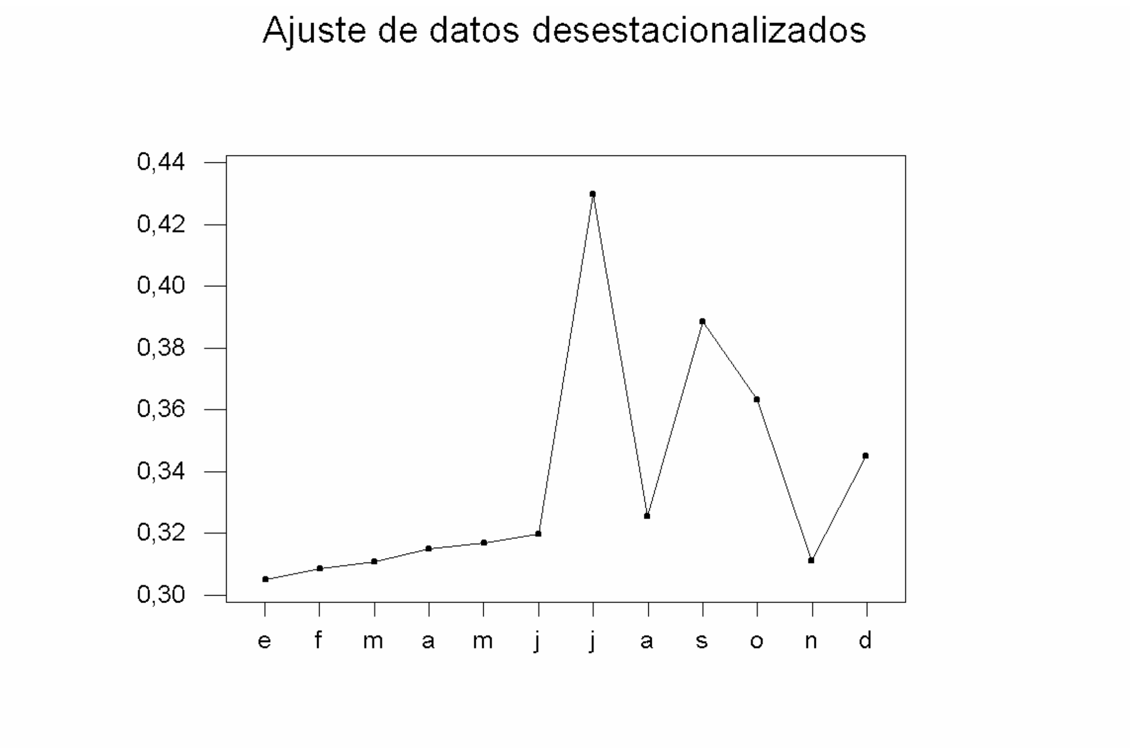
El siguiente paso es eliminar el error de los valores obtenidos de la estacionalidad. Los modelos clásicos de descomposición utilizan el enfoque del promedio medial. Para calcular el promedio medial se toman todos los datos de promedio móvil para cada período (mes, trimestre, etc.) y se eliminan los valores extremos, con los datos restantes se calcula el promedio.

Posteriormente se eliminan la tendencia de los promedios con lo que me queda la componente del ciclo. A su vez elimino la estacionalidad con lo que obtengo el ruido o error o residuo.

Con estas operaciones se obtendría una tabla como la que sigue:

Periodo	T(t)	C(t)	E(t)	Er(t)
1	0,322727	0,000265	-0,20976	-0,01803
2	0,319237	-0,00097	-0,15846	-0,01001
3	0,316727	-0,00169	-0,03244	-0,00449
4	0,315197	0,008328	-0,16017	-0,00866
5	0,314647	-0,00123	-0,1346	0,003383
6	0,315077	-0,00635	0,201805	0,01097
7	0,316487	-0,00485	0,28175	0,118308
8	0,318877	0,003115	0,39769	0,003418
9	0,322247	0,001495	0,154822	0,064836
10	0,326597	0,003645	-0,03051	0,032868
11	0,331927	0,00239	-0,17114	-0,02338
12	0,338237	-0,00402	-0,13899	0,010873

Representando gráficamente los datos desestacionalizados:



Por tanto la estimación para cada periodo sería la suma (modelo aditivo) de la tendencia, la estacionalidad y la ciclicidad:

Y(t) = T(t) + C(t) + E(t)

Siendo Y(t) los valores estimados, X(t) los datos históricos y Er(t) el error cometido en la estimación para cada periodo.

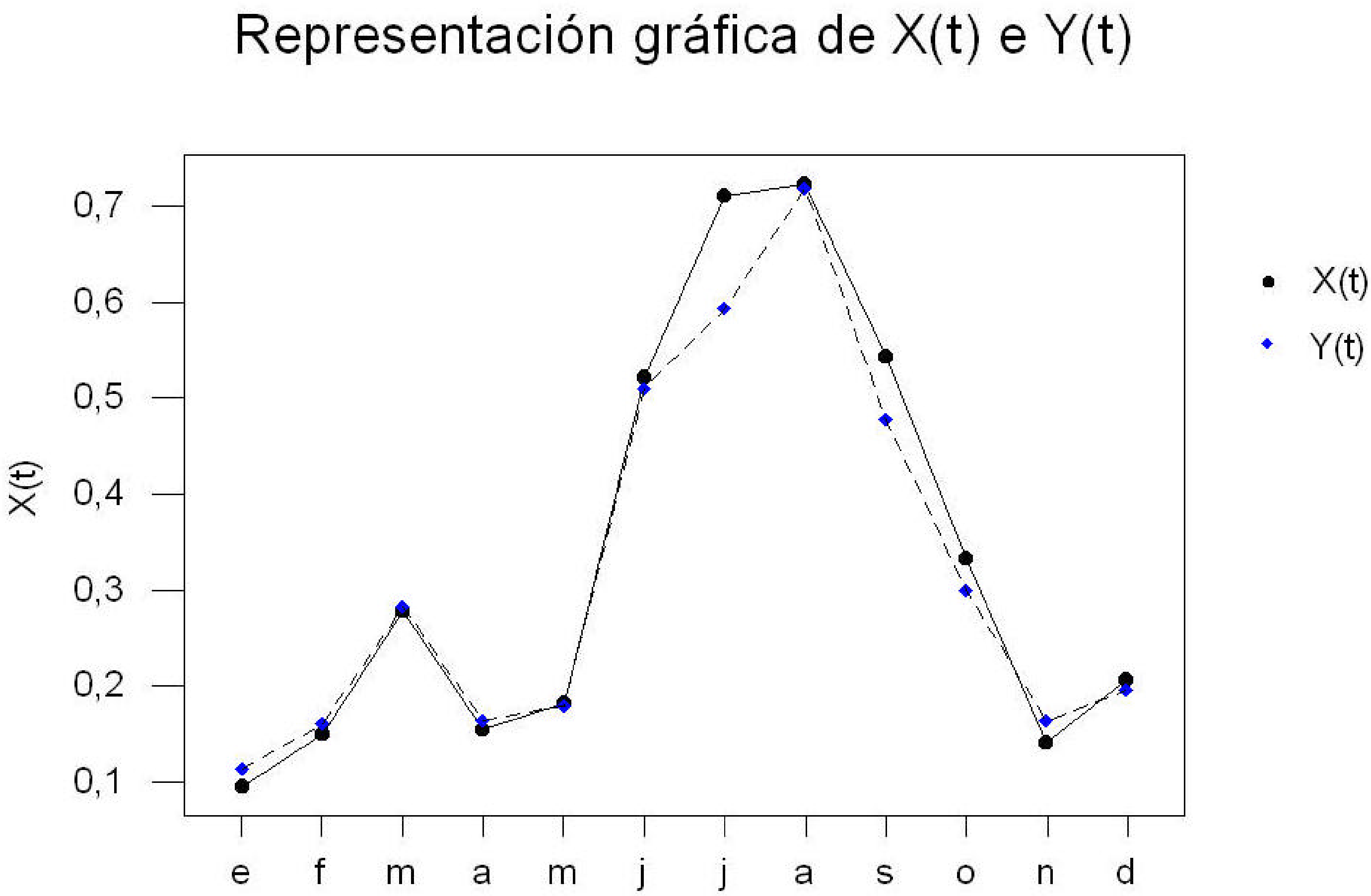
MAPE = 8,141578

MAD = 0,025768

MSE = 0,001717

Periodo	**X(t)**	**Y(t)**	**Er(t)**
1	0,0952	0,113232	-0,01803
2	0,1498	0,159807	-0,01001
3	0,2781	0,282593	-0,00449
4	0,1547	0,163355	-0,00866
5	0,1822	0,178817	0,003383
6	0,5215	0,51053	0,01097
7	0,7117	0,593392	0,118308
8	0,7231	0,719682	0,003418
9	0,5434	0,478564	0,064836
10	0,3326	0,299732	0,032868
11	0,1398	0,163177	-0,02338
12	0,2061	0,195227	0,010873

Además si se representan gráficamente se puede observar la bondad del ajuste de los datos estimados:



Comparando los datos estimados con los hoteles 2 y 3 se obtiene los siguientes errores:

Y(t)	Hotel 2	Hotel 3	Error 2	Error 3
0,113232				
0,159807				
0,282593				
0,163355				
0,178817	0,1815		0,002683	
0,51053	0,6225		0,11197	
0,593392	0,6701	0,723	0,076708	0,129608
0,719682	0,7057	0,764	0,013982	0,044318
0,478564	0,5033	0,558	0,024736	0,079436
0,299732	0,3277	0,3589	0,027968	0,059168
0,163177		0,2307		0,067523
0,195227		0,3045		0,109273

MAD_2 =	0,043008	MAD_3 =	0,081555
MSE_2 =	0,003336	MSE_3 =	0,007512
$MAPE_2$ =	7,723947	$MAPE_3$ =	19,93404

Se puede observar también como la ocupación estimada está dentro de los intervalos de confianza calculados anteriormente.