

Corrección del 2º parcial de Estadística administrativa I

Febrero de 2006

1. Los índices de precios de Paasche son índices tipo media ponderada que se usan para evaluar la evolución de los precios de un conjunto de bienes a los largo del tiempo. En concreto, en el índice de precios de Paasche las ponderaciones (o importancia que se le da a cada bien) son los gastos que se efectuarían en cada bien en el período actual si se hubiesen mantenido los precios del período base, es decir, son los precios unitarios del período base por las unidades consumidas en el período actual. Matemáticamente, sería $w_i = p_{i0} \times q_{it}$, con lo que tendríamos:

$$P_{p0}^t = \frac{\sum_{i=1}^N P_0^t(i) w_i}{\sum_{i=1}^N w_i} = \frac{\sum_{i=1}^N \frac{p_{it}}{p_{i0}} p_{i0} \times q_{it}}{\sum_{i=1}^N p_{i0} \times q_{it}} = \frac{\sum_{i=1}^N p_{it} \times q_{it}}{\sum_{i=1}^N p_{i0} \times q_{it}}$$

Para este índice podemos dar dos **interpretaciones**. Por un lado, al ser una media ponderada representa la variación media de los precios entre el período base y el actual teniendo en cuenta el valor que tendrían actualmente los bienes de haberse mantenido los precios. Por otro lado, según vemos en la fórmula final, compara el gasto actual con el que se hubiese tenido al comprar las cantidades que se consumen actualmente a precios del período base. Esta interpretación está más en consonancia con el concepto común de “inflación” (variación del poder adquisitivo) que la correspondiente al índice de Laspeyres. Sin embargo, como depende de las cantidades que se consumen en el período actual, es más complicado de aplicar en la práctica. Aunque en España no se utiliza como indicador de la inflación, en otros países, como Estados Unidos, sí se utiliza.

2. El **objetivo del apartado a)** es calcular los índices de calidad conjuntos base 2003 de cada año. Tenemos que comenzar planteando el problema para localizar la información que nos dan. **Planteamiento:** tenemos 3 bienes, que son los grupos de asignaturas. La importancia de cada bien es de 10, 8 y 7 respectivamente. La magnitud es la calidad. Los períodos son 2003 (que actuará como base), 2004 y 2005. Nos dicen que en el grupo 1 el índice de calidad disminuyó un 4 % en 2004 respecto a 2003, es decir, el índice estará 0.04 puntos por debajo de 1: $I_{03}^{04}(1) = 1 - 0,04 = 0,96$ puntos. En el grupo 2 el índice de calidad de 2004 respecto a 2003 aumentó un 2.5 %, es decir, el índice estará 0.025 puntos por encima de 1: $I_{03}^{04}(2) = 1 + 0,025 = 1,025$ puntos. En el grupo 3 ese índice se sitúa en 101.5 puntos porcentuales: $I_{03}^{04}(3) = 1,015$ puntos. Además, nos dicen que los índices de 2005 base 2004 son: $I_{04}^{05}(1) = 1,05$, $I_{04}^{05}(2) = 1,03$ y $I_{04}^{05}(3) = 0,995$. En conclusión, tenemos índices simples y ponderaciones (porque los grupos tienen distinta importancia), luego tendremos que calcular el índice conjunto media ponderada:

$$I_{03}^{04} = \frac{I_{03}^{04}(1)w_1 + I_{03}^{04}(2)w_2 + I_{03}^{04}(3)w_3}{w_1 + w_2 + w_3} = \frac{0,96 \times 10 + 1,025 \times 8 + 1,015 \times 7}{10 + 8 + 7} = 0,9962,$$

lo que significa que la calidad de la docencia en esos tres grupos de asignaturas bajó en media un $(1 - 0,9962) \times 100 = 0,38$ % en el año 2004 respecto a 2003 teniendo en cuenta la importancia de cada grupo.

Para hallar el índice del año 2005 base 2003 nos fijamos en que tenemos los índices simples de cada grupo del año 2005 base 2004, pero como también conocemos los índices de 2003 a 2004, podemos hacer un **enlace** para aproximar su valor en base 2003 (se podría hacer también el enlace con el índice conjunto): $I_{03}^{05}(1) = I_{03}^{04}(1) \times I_{04}^{05}(1) = 0,96 \times 1,05 = 1,008$, $I_{03}^{05}(2) = 1,025 \times 1,03 = 1,0558$, $I_{03}^{05}(3) = 1,015 \times 0,995 = 1,0099$. El índice conjunto sería entonces:

$$I_{03}^{04} = \frac{1,008 \times 10 + 1,0558 \times 8 + 1,0099 \times 7}{10 + 8 + 7} = 1,0238,$$

lo que significa que la calidad de la docencia en esos tres grupos de asignaturas subió en media un 2,383 % en el año 2005 respecto a 2003 teniendo en cuenta la importancia de cada grupo de asignaturas.

El **objetivo del apartado b)** es calcular la variación del índice conjunto base 2003 al pasar de 2004 a 2005 a causa de las asignaturas de del grupo 1. Podríamos contestar tanto en términos absolutos (repercusión absoluta) como en términos relativos (repercusión en porcentaje). Nos fijamos en que no podríamos contestar con la participación, porque *no* nos preguntan por la *parte* que se debe al grupo 1, *sino* por la *cantidad* en sí. Como lo más habitual es medir dicha variación con la repercusión en porcentaje, es lo que haremos en este caso. Para

hallar la repercusión en porcentaje hay que hallar primero la absoluta:

$$R_{03}^{04 \rightarrow 05}(1) = \frac{(I_{03}^{05}(1) - I_{03}^{04}(1))w_1}{w_1 + w_2 + w_3} = \frac{(1,008 - 0,96) \times 10}{10 + 8 + 7} = 0,0192 \text{ puntos},$$

es decir, el índice conjunto de calidad base 2003 aumentó 0.0192 puntos al pasar de 2004 a 2005 a causa de las asignaturas del grupo 1. Para hallar la repercusión en porcentaje tenemos que dividir ese valor por el valor inicial del índice:

$$\%R_{03}^{04 \rightarrow 05}(1) = \frac{R_{03}^{04 \rightarrow 05}(1)}{I_{03}^{04}} = \frac{0,0192}{0,9962} = 0,0193,$$

lo que significa que los 0.0192 puntos representan un 1.93 % del valor inicial del índice, es decir, el índice conjunto de calidad base 2003 aumentó su valor un 1.93 % a causa de las asignaturas del grupo 1 al pasar de 2004 a 2005.

3. El objetivo del apartado a) se refiere a la regresión. Vamos a identificar en primer lugar todos los elementos del problema. **Planteamiento:** el experimento consiste en observar el número de vallas alquiladas (variable 1) y las ventas de refresco, en miles de litros, (variable 2) que se vende en cada ciudad (individuo). La población la constituirán entonces todas las ciudades de tamaño moderado de Europa y tenemos una muestra de tamaño desconocido. Las variables son cardinales (porque son respectivamente *número* de vallas y litros) y de razón (porque 0 significa no tener ninguna valla o no vender nada respectivamente, es decir, nulidad). El número de vallas alquiladas es discreta, porque no se puede alquilar 1 valla y media y las ventas es continua, porque pueden venderse 8953 litros o 8769.2 litros, etc.

En el apartado a) nos piden que expliquemos el mejor modelo de regresión. Para saber cuál es el mejor modelo, debemos mirar el que tiene mayor *coeficiente de determinación* (R^2 o Rsq), ya que ese valor mide la proporción de variación de Y que se explica por la curva correspondiente para ese modelo. Cuanto mayor sea R^2 mayor será esa proporción, es decir, más cerca estarán los datos de la curva correspondiente. En este caso, el mayor R^2 es el del modelo logarítmico $Y = b_0 + (b_1 \ln(t))$, con un valor de 0.87. Como queremos determinar las ventas *en función* del número de vallas que se alquilan, la *variable independiente* X es el número de vallas alquiladas (es la que controlaremos) y la *variable dependiente* Y son las ventas de refresco (es la que podremos predecir). La *fórmula de la regresión* sería entonces $y^* = -900,07 + 534,661 \ln(X)$, eso *significa* que, de entre los modelos planteados, esta es la curva que mejor se aproxima a los datos. Como es la mejor fórmula para aproximar Y a partir de X , la regresión se puede *utilizar* para predecir los valores de las ventas en una ciudad si sabemos cuantas vallas se han alquilado en esa ciudad. La *fiabilidad* de la regresión para hacer predicciones se mide a través de su coeficiente de determinación. En este caso $R^2 = 0,87$, lo que significa que el 87 % de la variabilidad de las ventas queda determinada por el número de vallas alquiladas a partir del modelo logarítmico. Como está bastante cerca del 100 %, eso significa que los datos se ajustan bastante bien a la curva, por lo que las predicciones serán bastante fiables.

El **objetivo del apartado b)** nos piden calcular valores de Y a partir de valores de X . Para ello debemos utilizar el mejor modelo de regresión. Ya sabemos del apartado anterior que, en este caso, el mejor modelo es $y^* = -900,07 + 534,661 \ln(X)$. Utilizando esa fórmula conseguiremos una *predicción*, no un valor exacto (nunca podemos asegurar que va a ser una cantidad exacta, siempre puede haber algún error) aunque sí bastante fiable, ya que el $R^2 = 0,87$ indica que los datos de la muestra se ajustan bastante bien a ese modelo. En primer lugar debemos asegurarnos que el número de vallas que nos dan están entre el mínimo y el máximo de la muestra, porque la regresión sirve para interpolar, pero no podemos decir si la curva es válida fuera del rango de la muestra (además las ciudades deberían ser del tipo de las de la muestra: de tamaño moderado y europeas). Nos dicen que en la muestra el número de vallas alquiladas variaba entre 10 y 150, luego podríamos hacer la predicción en el primer caso (valor de 100 vallas), pero no en el segundo (valor de 200 vallas). Sólo tenemos que sustituir X por 100 en la fórmula anterior: $y^* = -900,07 + 534,661 \ln(100) = 1562,1349$ miles de litros (recordamos que la Y en la muestra venía medida en miles de litros). En conclusión, si se alquilan 100 vallas, pronosticamos unas ventas de 1562134.9 litros en los 6 meses (recordamos que esta aproximación es fiable, es decir, estará cerca del valor real, pero no es exacta).