

Seminario Excel I - Estadística I

Grado en Economía

VNIVERSIDAD D SALAMANCA

Marzo 2023

Índice

1	Introducción a Excel	1
2	Estadística univariante	2
2.1	Variable discreta	2
2.2	Variable continua	10
2.3	Medidas de concentración. Curva de Lorenz, índice de Gini	15

1 Introducción a Excel

Microsoft Excel es un programa de hojas de cálculo desarrollado por Microsoft Corporation. Es una de las aplicaciones más populares y ampliamente utilizadas en todo el mundo para realizar tareas financieras, contables y de análisis de datos.

Excel ofrece una amplia variedad de funciones y herramientas de análisis de datos para trabajar con datos, realizar cálculos complejos, crear gráficos y tablas, y automatizar tareas repetitivas.

En este seminario mostraremos algunas de las herramientas de Estadística Descriptiva que incorpora Excel.

Abrir el programa e indentificar:

- Barra de título
- Barra de acceso rápido
- Barra de opciones. Menú archivo.
- Barra de fórmulas
- Barra de etiquetas

Algunas notas sobre Excel:

- Los documentos de Excel, se denominan libros, y se organizan en hojas. Cada hoja está formada por una matriz de celdas.
- Las celdas se identifican con una letra, relativa a la columna, y un número relativo a la fila (A1, B3, . . .)
- En cada celda es posible escribir: texto, valor constante, o una fórmula.
- Las fórmulas comienzan con el signo "=".
- Las fórmulas pueden incorporar funciones predefinidas en Excel (PROMEDIO, MAX, SUMA . . .), o se puede construir por el usuario con el operadores aritméticos (+,-,*,/) y funciones predefinidas
- En las fórmulas se puede utilizar referencias a otras celdas.
- Una fórmula se puede extender horizontalmente o verticalmente arrastrando la celda desde la posición inferior derecha.

- Cuando una fórmula se extiende, si la fórmula involucra referencias a otras celdas las referencias se modifican de forma relativa a como se realice la extensión.
- Para bloquear la referencia a una celda, cuando se extiende una fórmula, utiliza el carácter \$. Ejemplos:
 - `A$3` Bloquea la fila 3 (referencia absoluta en fila).
 - `$A3` Bloquea la columna A (referencia absoluta en columna).
 - `A3` Bloquea la fila 3, y columna A (referencia absoluta en fila - columna).

Ejercicio Trabajar con referencias relativas y absolutas de celdas.

2 Estadística univariante

Esta práctica utilizará como hilo conductor distintos ejemplos para presentar las funcionalidades de Excel. En este documento solo se discutirán las etapas fundamentales.

2.1 Variable discreta

Ejercicio 1. El número de convocatorias agotadas por un alumno hasta aprobar la asignatura de Estadística I en el grado de Economía fue el siguiente: con una convocatoria 40 alumnos, con dos convocatorias 32 alumnos, con tres 8, con cuatro 5, con cinco 0 y con seis 2. Se pide:

1. Construir la tabla de frecuencias en Excel.
2. Representarla gráficamente mediante un diagrama de sectores y de barras.
3. Calcular la moda y mediana.
4. Calcular media, varianza y coeficiente de variación
5. Calcular el coeficiente de asimetría y el de apuntamiento de Fisher.

Resolución

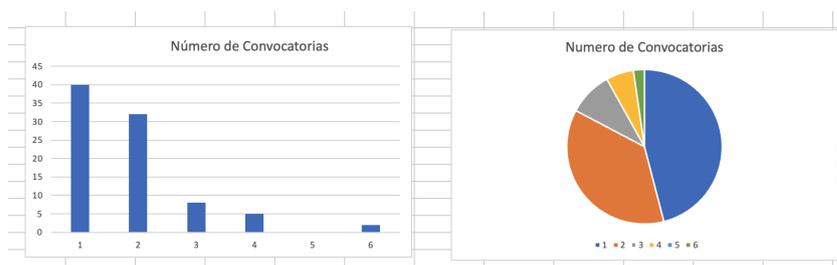
1. Construir una tabla en excel con la información de las características $\{1, 2, 3, \dots, 6\}$ y las frecuencias absolutas $\{n_1 = 40, n_2 = 32, \dots, n_6 = 2\}$ por columnas.

Generar las columnas f_i (frecuencias relativas), N_i (Frecuencias absolutas acumuladas) y F_i (frecuencias relativas acumuladas), para obtener:

	A	B	C	D	E	F
1						
2		Ejercicio 1				
3						
4		xi	ni	fi	Ni	Fi
5		1	40	0,46	40	0,46
6		2	32	0,37	72	0,83
7		3	8	0,09	80	0,92
8		4	5	0,06	85	0,98
9		5	0	0,00	85	0,98
10		6	2	0,02	87	1,00
11			87	1,00		

- Para calcular el número de datos N , se escribió en la celda C11: `=SUMA(C5:C10)`
- Para generar f_i , se escribió en la celda D5: `=C5/C$11` y se arrastró la fórmula.

- Podemos comprobar que la suma de todas las frecuencias relativas es 1, escribiendo en la celda D11: $\boxed{=SUMA(D5:D10)}$
 - Para generar frecuencias absolutas acumuladas, en la celda E1 se escribe $\boxed{=C5}$ ($N_1 = n_1$), y en la celda E2: $\boxed{=C6 + E5}$ ($N_2 = N_1 + n_1$), después se arrastra la fórmula hasta el final de la columna ($N_i = N_{i-1} + n_i$).
 - Las frecuencias relativas acumuladas F_i , se obtienen de forma semejante.
2. Seleccionamos la columna de frecuencias absolutas. Después en el menú buscamos la pestaña *Insertar* y en ella gráficos. Seleccionamos el diagrama de barras. Una vez obtenido el gráfico, nos situamos sobre él y pulsamos botón derecho del ratón, para configurar (seleccionar datos). Por ejemplo, es conveniente fijar el eje de categorías.
- Se procede de forma similar para el gráfico de sectores.



3. No existe (al menos en la versión actual) fórmulas en Excel que permitan calcular directamente la moda y mediana de una distribución de datos en forma de tabla de frecuencias. La moda se puede implementar a través de la función BUSCARX. La implementación de la mediana, no es directa, y se ha evitado (se calculara como se ha hecho en la pizarra).
- Para calcular la moda se puede utilizar el comando

$$\boxed{=BUSCARX(MAX(C5:C10);C5:C10;B5:B10)}$$

Busca en la ayuda la función BUSCARX, y trata de entender el resultado.

Para el cálculo de la mediana, calculamos $N/2$ y procedemos como hemos hecho en clase.

El resultado debería ser este:

Moda	1			
Mediana	2	N/2	Ni	i-->2
		43,5		72

4. Para el cálculo de media una nueva columnas a la tabla de frecuencias, con los valores: $x_i n_i$. Sumamos los resultados de esta columna en la celda H11 y el valor resultante lo dividimos por el tamaño de la muestra, para obtener la media: $\boxed{=H11/C11}$.

Calculada la media, procedemos de forma similar para la varianza. Añadimos una columna con los valores $(x_i - \bar{x})^2 n_i$, y los sumamos en la celda I11. Para obtener la varianza: $\boxed{=I11/C11}$.

La desviación típica se obtiene con la función RAIZ(). El coeficiente de variación requiere de la función ABS(), para determinar el valor absoluto.

El resultado debería ser:

xini	(xi - barx)^2 ni
40	28,16
64	0,83
24	10,78
20	23,35
0	0,00
12	34,63
160	97,75

Media	1,84
Varianza	1,12
Desv. Típica	1,06
Cv	0,58

5. Para calcular los coeficientes de asimetría y apuntamiento de Fisher, procedemos de forma semejante: en primer lugar calculamos columnas con las cantidades:

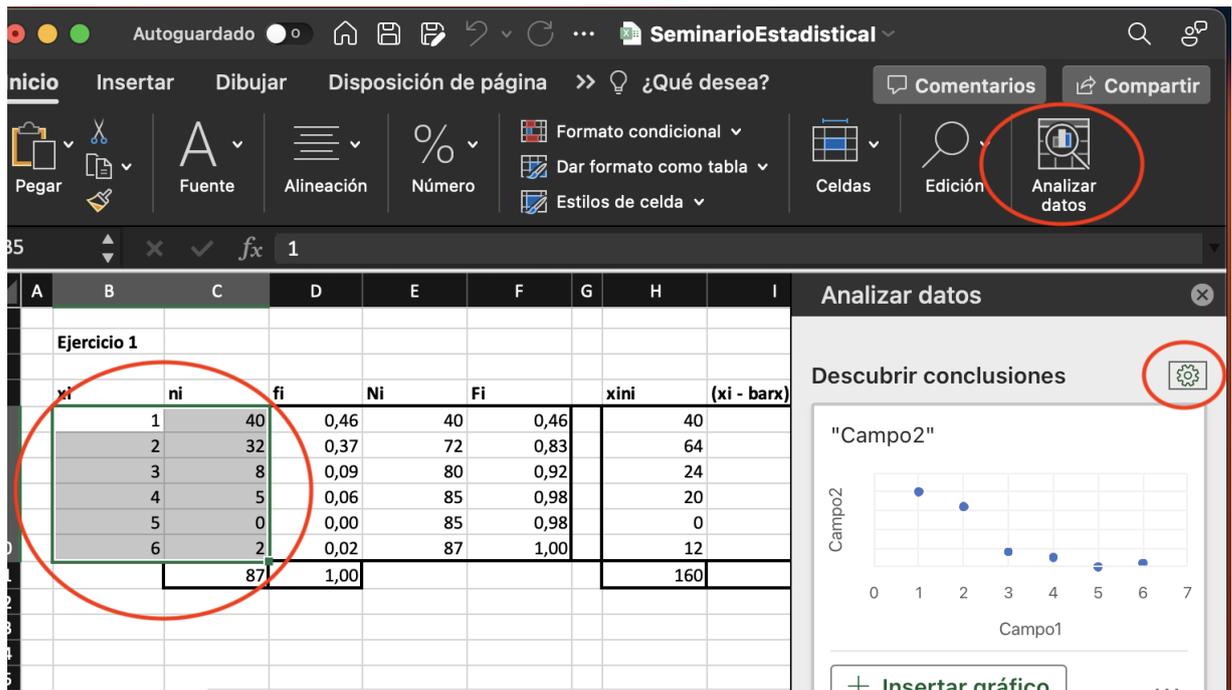
$$(x_i - \bar{x})^3 n_i \qquad (x_i - \bar{x})^4 n_i$$

Después la suma correspondiente, y dividiendo por el tamaño de la muestra los momentos centrales de orden 3 (m_3) y orden 4 (m_4) y finalmente los coeficientes de Fisher. Recordar que las potencias en Excel se indican con el carácter ^. El resultado debería ser el siguiente:

(xi - barx)^3 ni	(xi - barx)^4 ni
-23,63038589	19,82779506
256	512
52,85848755	99,18735615
320	1280
0	0
432	2592
1037,23	4503,02

m3	11,92216	g1	10,0110102
m4	51,75879	g2	41,0029084

Antes de continuar, es interesante añadir que Excel incorpora una herramienta que realiza un análisis automático. Para ello, seleccionamos los datos originales B5:C10 y pulsamos en la pestaña inicio **Analizar datos**. En la parte derecha de la pantalla, Excel, propone determinados análisis que pueden ser configurados.



En el siguiente ejercicio, trabajaremos con una colección de datos. Aquí Excel, incorpora herramientas que permiten una análisis estadístico directo (descriptivo). Para ello se precisa del módulo **Análisis de datos**. Para su instalación, pinchar en el siguiente enlace y proceder como se indica:

Windows Herramientas de análisis de datos

Mac Herramientas de análisis de datos

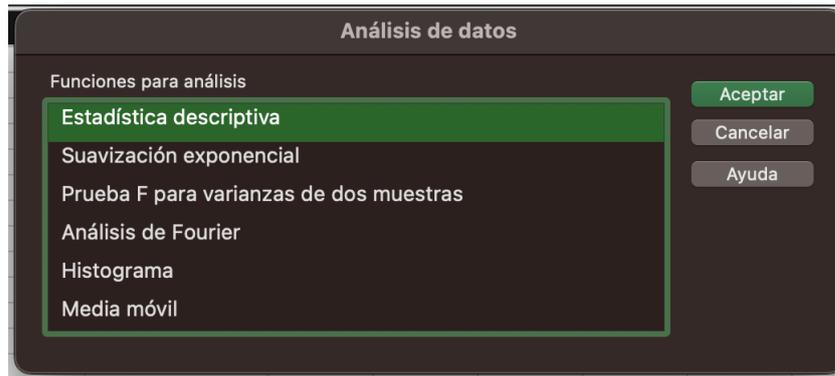
Ejercicio 2. En el Departamento de personal de una fábrica se ha realizado una investigación estadística en relación a los salarios, expresados en euros, que reciben anualmente sus trabajadores por todos los conceptos. Los resultados obtenidos fueron los siguientes:

19.000 24.000 21.000 21.000 28.000 19.000 21.000 26.000 29.000 27.000 22.000
 30.000 20.000 24.000 23.000 28.000 25.000 27.000 25.000 28.000 29.000 26.000
 22.000 26.000 25.000 26.000 24.000 27.000 26.000 20.000 22.000 22.000 19.000
 23.000 28.000 28.000 30.000 26.000 30.000 20.000 25.000 27.000 29.000 27.000
 21.000 29.000 24.000 25.000 27.000 23.000 29.000 27.000 26.000 26.000 23.000
 30.000 28.000 24.000 25.000 23.000

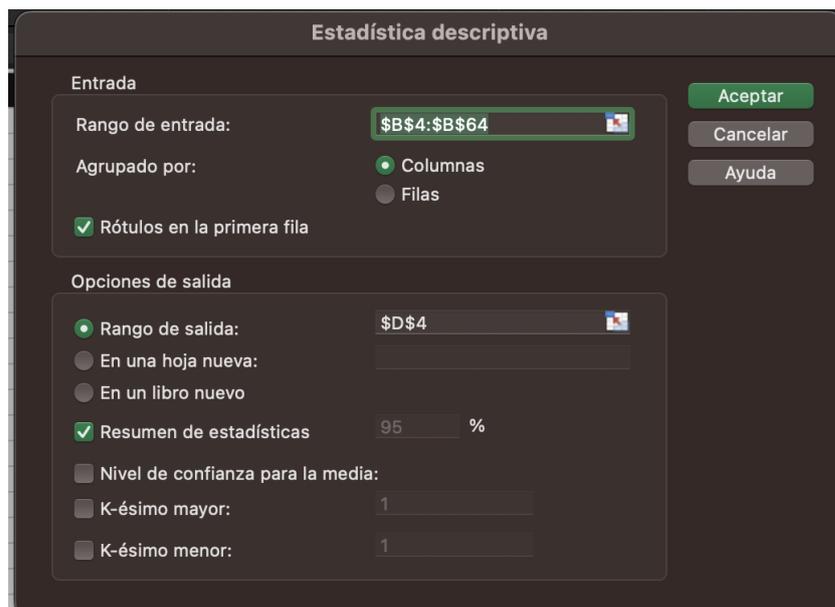
1. Análisis estadístico: medidas de posición, forma, apuntamiento.
2. Tabla de frecuencias de los salarios.
3. La representación gráfica de los datos como diagrama de barras y polígono de frecuencias.
4. Diagrama de caja.

Resolución

1. Seleccionamos el rango de datos B4:B16, y en la etiqueta de datos pulsamos **Análisis de datos**. En la ventana emergente, seleccionamos **Estadística descriptiva**



A continuación se configura esta ventana, tal y como aparece en la siguiente figura:



Tras pulsar en aceptar, se obtiene el siguiente resumen de variables estadísticas:

Salarios	Salarios	
19000		
21000	Media	25066,66667
20000	Error típico	403,9820622
25000	Mediana	25500
25000	Moda	26000
22000	Desviación estándar	3129,231598
30000	Varianza de la muestra	9792090,395
29000	Curtosis	-0,869126378
27000	Coefficiente de asimetría	-0,287001962
23000	Rango	11000
24000	Mínimo	19000
26000	Máximo	30000
24000	Suma	1504000
28000	Cuenta	60
26000		
22000		

Existen funciones en Excel, para realizar calcular estas medidas estadísticas de forma individual. Se trata de las funciones:

media	→	PROMEDIO(B5:B64)
mediana	→	MEDIANA(B5:B64)
moda	→	MODA(B5:B64)
varianza	→	VAR(B5:B64)
desviación típica	→	DESVEST(B5:B64)
coeficiente de asimetría	→	COEFICIENTE.ASIMETRIA(B5:B64)
coeficiente de curtosis	→	CURTOSIS(B5:B64)
cuartil k	→	CUARTIL.INC(B5:B64;k)
percentil xx	→	PERCENTIL.INC(B5:B64;0,xx)

Que evidentemente, proporcionan los mismos resultados.

Media	25066,66667
Mediana	25500
Moda	26000
Varianza	9792090,395
Desviación Estándar	3129,231598
Coefficiente de Asimetría	-0,287001962
Coefficiente de Curtosis	-0,869126378
Cuartil 1	23000
Cuartil 2	25500
Cuartil 3	27250
Percentil 65	27000

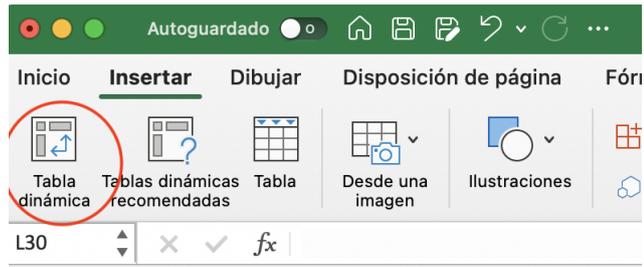
2. Para construir la tabla de frecuencias, primero se genera la columna de características de la variable (de 19000 a 31000), que suponemos que ocupa el rango D40:D51. Para calcular las frecuencias absolutas, utilizamos el comando CONTAR.SI. En concreto en la celda E40, colocamos la fórmula:

=CONTAR.SI(B5 : B64;D40)

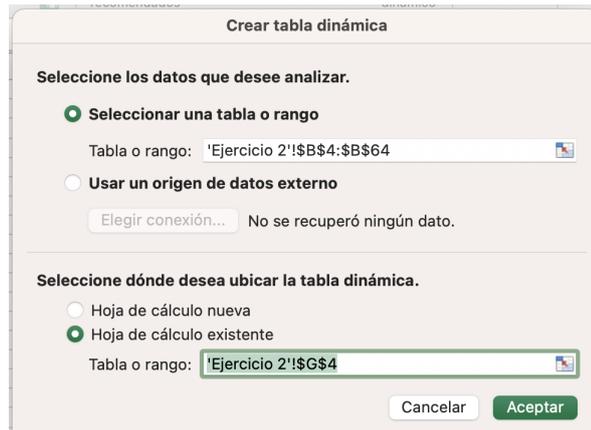
El primer argumento es el rango de los datos, donde Excel, busca cuantos datos son iguales a D40. Esta fórmula se extiende hasta E51, para obtener:

Tabla de frecuencias		
xi		ni
	19000	3
	20000	3
	21000	4
	22000	4
	23000	5
	24000	5
	25000	6
	26000	8
	27000	7
	28000	6
	29000	5
	30000	4
		60

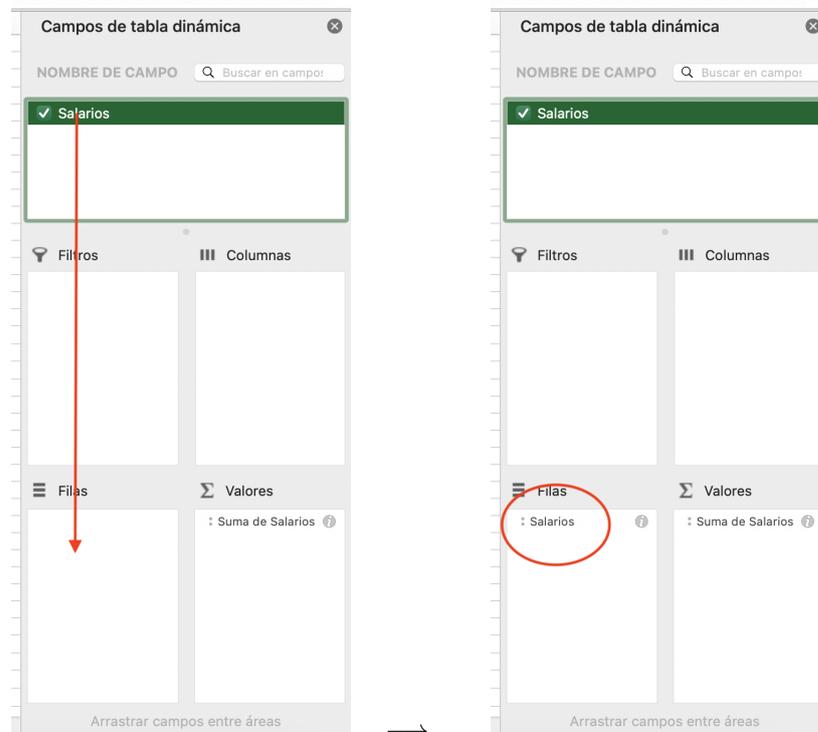
También es posible obtener la tabla de frecuencias, utilizando **tablas dinámicas**. Pulsamos la etiqueta insertar, y en ella el icono de tabla dinámica:



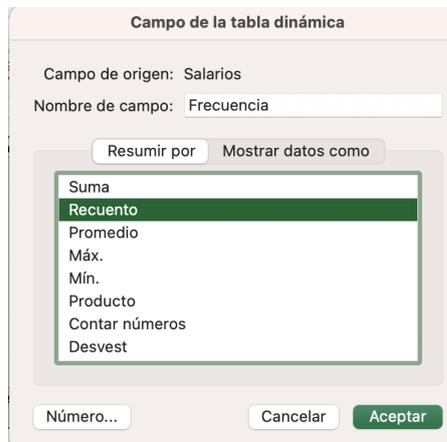
Después completamos la ventana que se despliega, indicando el rango de los datos y el lugar donde se incluirá la tabla:



En la parte izquierda de la pantalla aparecerá el menú de configuración de la tabla dinámica. Primero marcamos salarios, y después lo arrastramos hasta el recuadro columnas, tal y como se indica en el gráfico:



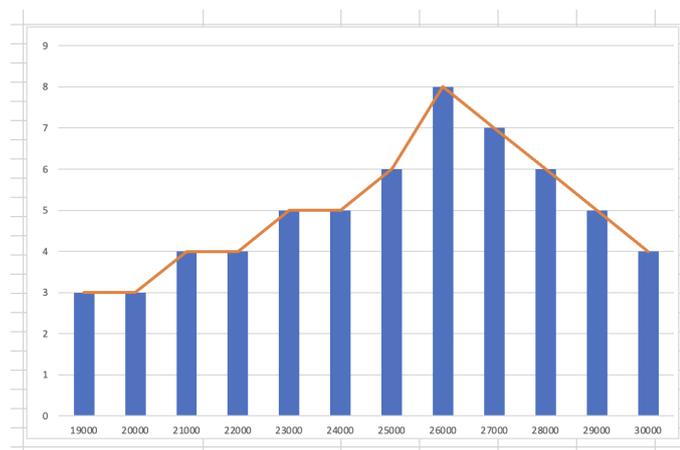
En el recuadro, valores, pulsamos en la *i* de suma de salarios, y en la ventana que se despliega escribimos Frecuencia como nombre de campo y seleccionamos (Resumir por) recuento:



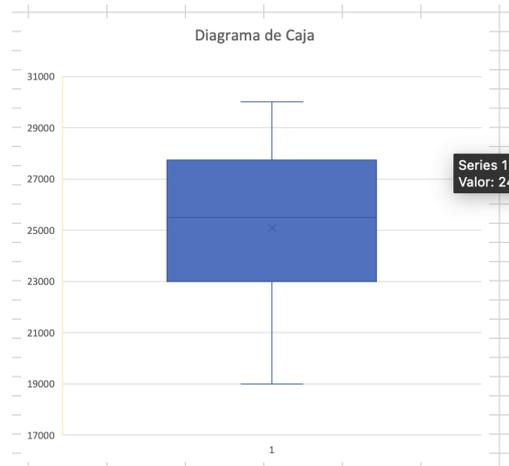
El resultado final es:

Salarios	Frecuencia
19000	3
20000	3
21000	4
22000	4
23000	5
24000	5
25000	6
26000	8
27000	7
28000	6
29000	5
30000	4
Total general	60

- Para dibujar el diagrama de barras junto con el polígono de frecuencias, se seleccionan los datos de la tabla de frecuencias anterior, y se selecciona un gráfico combinado. Después se ajustan las series de datos, hasta conseguir:



- Para dibujar el diagrama de caja, seleccionamos el rango de datos B5:B64 e insertamos el gráfico **Cajas y bigotes** (icono histograma). Tras ajustar ejes, deberíamos obtener:



2.2 Variable continua

Ejercicio 3. Ejercicio 2. En la hoja Ejercicio 3 se han introducido la tasa de paro de las provincias y ciudades autónomas españolas, del último cuatrimestre del año 2022 (Fuente: INE). Se pide:

1. Construir la tabla de frecuencias usando la fórmula de Sturges para determinar el número de intervalos. Esta expresión establece que, dado un conjunto de N datos, el número k de intervalos de clase puede aproximarse por $1 + \log N \log 2$.
2. Dibujar el histograma de frecuencias absolutas.
3. Calcular e interpretar los resultados para la media aritmética, la mediana y la moda (datos agrupados).
4. Calcular e interpretar los resultados para la desviación típica y el coeficiente de variación de Pearson.
5. Analiza la asimetría y la curtosis de la distribución.
6. ¿Cuál es la menor tasa de desempleo del 10% de las provincias que tienen mayor tasa de desempleo ?
7. Construir un diagrama de caja.

Resolución

1. Calculamos el máximo y mínimo de la distribución, para después calcular número de intervalos y amplitud:

Minimo	7,22
Máximo	30,56
N	52
m	7
A	3,334285714

Hemos utilizado los siguientes comandos:

- Mínimo → `=MIN(B4:B55)-0,01`
- Máximo → `=MAX(B4:B55)+0,01`
- N → `=CONTAR(B4:B55)`

- $m \rightarrow$ `=REDONDEAR(1+LOG10(F6)/LOG10(2);0)`
- $A \rightarrow$ `=CONTAR(B4:B55)`

Observar que hemos restado y sumado, respectivamente, una centésima al mínimo y máximo, para evitar problemas de redondeo.

A continuación, construimos la tabla de frecuencias:

Clase	ai	ai+1	ni	Ni	fi	Fi
1	7,22	10,55	24	24	0,46	0,46
2	10,55	13,89	10	34	0,19	0,65
3	13,89	17,22	8	42	0,15	0,81
4	17,22	20,56	5	47	0,10	0,90
5	20,56	23,89	3	50	0,06	0,96
6	23,89	27,23	1	51	0,02	0,98
7	27,23	30,56	1	52	0,02	1,00
			52			

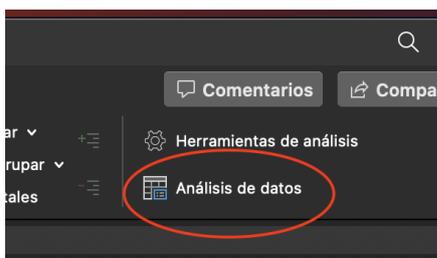
donde hemos actuado del siguiente modo:

- Donde a_0 es el mínimo, a_7 el máximo, y el resto se construyen sumando la amplitud A .
- Con las instrucciones que ofrece Excel, es más sencillo determinar primero la frecuencia acumulada absoluta con el comando CONTAR.SI. En este caso si asumimos que primera clase esta en la fila 12, y si frecuencia absoluta acumulada en la columna I, para calcular N_1 , escribiríamos: `=CONTAR.SI(B$4:B$55;"<"&G12)`. Esta instrucción cuenta todos los valores de la serie de datos menores que el extremo del superior del intervalo. Extendemos la fórmula y tenemos las frecuencias absolutas acumuladas.
- Para calcular las frecuencias absolutas, utilizamos:

$$n_1 = N_1, \quad n_2 = N_2 - N_1, \dots$$

- Las frecuencias relativas se calculan como en los ejercicios anteriores.

2. Para dibujar el histograma de frecuencias seleccionamos la lista de datos, y a continuación pulsamos la herramienta **análisis de datos** de la pestaña datos:

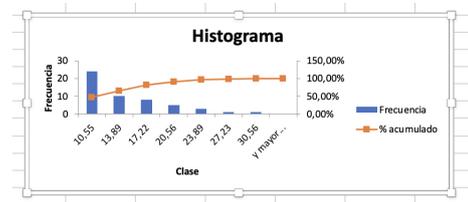


En el desplegable seleccionamos histograma, que configuramos del siguiente modo:

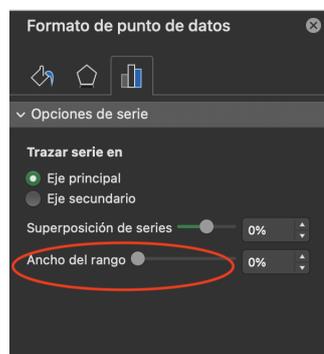


Notar que solo se introduce el extremo superior de los intervalos de clase. En la celda E22, aparece una tabla de frecuencias y un histograma:

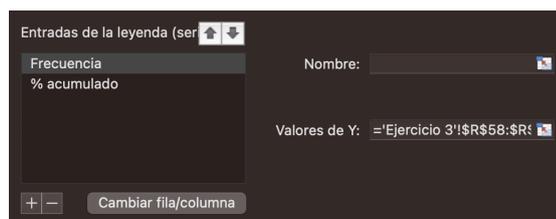
Clase	Frecuencia	% acumulado
10,55	24	46,15%
13,89	10	65,38%
17,22	8	80,77%
20,56	5	90,38%
23,89	3	96,15%
27,23	1	98,08%
30,56	1	100,00%
y mayor...	0	100,00%



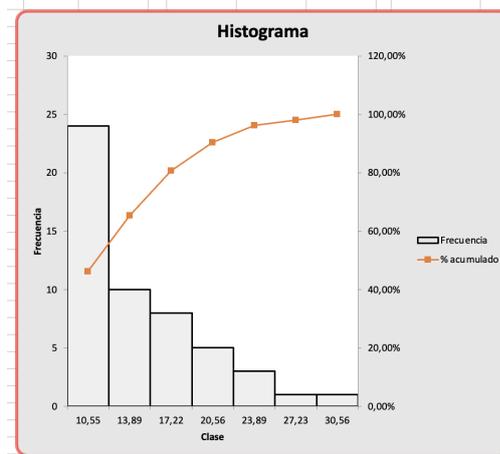
Ahora, configuramos el histograma. En primer lugar pulsando sobre las barras, aparece un menú en la parte izquierda de la pantalla (formato de serie de datos), donde seleccionamos **Opciones de serie de datos**, y fijamos el ancho de rango a cero



Con ellos conseguimos que las barras se superpongan. A continuación, vamos a eliminar la última columna, accedemos a la serie de datos (boton derecho del ratón situado sobre el gráfico), y modificamos los valores de la serie Y (eliminando la última fila).



Podemos modificar otras opciones del histograma hasta conseguir algo del tipo:



3-5 Para calcular la media (varianza, asimetría o curtosis) procedemos como en el ejemplo anterior, pero ahora con las marcas de clase, generando nuevas columnas en la tabla de frecuencias:

mi	mi ni	mi ² ni	ni (mi - barx) ³	ni (mi - barx) ⁴
8,88714286	213,291429	1895,5514	-1658,630197	6806,580887
12,2214286	122,214286	1493,63316	-4,555563895	3,505281142
15,5557143	124,445714	1935,84198	134,979671	346,2006066
18,89	94,45	1784,1605	1026,436035	6055,070248
22,2242857	66,6728571	1481,75663	2361,604323	21805,65293
25,5585714	25,5585714	653,240573	1985,027914	24947,22004
28,8928571	28,8928571	834,797194	4021,179378	63944,70609
	675,525714	10078,9814	7866,04156	123908,9361

Excel no tiene herramientas para calcular mediana y moda sobre tablas de frecuencia de datos agrupados, así que utilizamos las fórmulas que hemos visto en clase. Para las demás medidas estadísticas operamos igual que en ejercicios anteriores. El resultado debería ser:

Datos Agrupados				
Media		12,99087912		
Moda		9,325864662	Intervalo	1
Mediana		11,22114286	Intervalo	2
Varianza		25,0636256		
Desv Tip		5,006358517		
Cv		0,385374883		
Asimetría		1,205555063		
Curtosis		0,793250253		

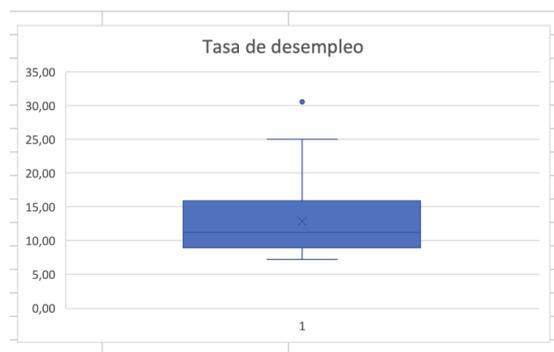
Podemos comparar el resultado, con el obtenido con la herramienta **Análisis de datos**→**Estadística Descriptiva** que trabaja con la serie de datos, como hemos visto en los ejercicios anteriores:

Columna1	Analisis de datos
Media	12,89134615
Error típico	0,700599293
Mediana	11,255
Moda	16,06
Desviación estándar	5,052093346
Varianza de la muestra	25,52364717
Curtosis	1,842387036
Coefficiente de asimetría	1,330407856
Rango	23,32
Mínimo	7,23
Máximo	30,55
Suma	670,35
Cuenta	52

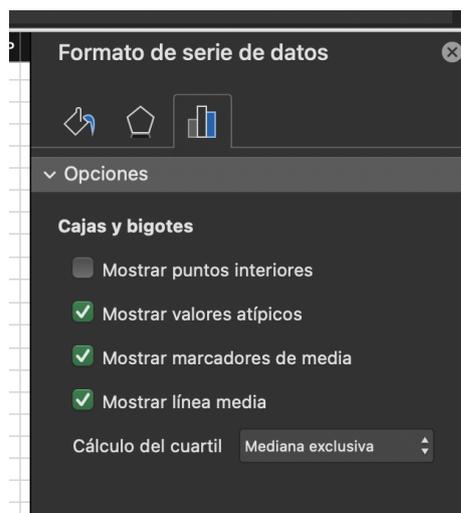
6. Buscamos en este caso el P_{90} que calculamos con la fórmula de Excel) PERCENTIL.INC que ya hemos visto, o con la fórmula para datos agrupados. La diferencia es mínima.

Sin agrupar	Percentil 90	19,62		
Agrupados	Percentil 90	20,42	Intervalo	4

7. Por último calculamos el diagrama de caja, que refleja la asimetría positiva de la distribución.



Hemos activado la visualización de puntos atípicos en la serie en formato de serie de datos (pulsar sobre el gráfico):



2.3 Medidas de concentración. Curva de Lorenz, índice de Gini

Ejercicio. Considerar la siguiente distribución de renta para una población de 170 individuos.

Renta	n_i
[500, 1000)	20
[1000, 1500)	30
[1500, 2000)	40
[2000, 3500)	30
[2500, 3000)	50

Calcular la curva de Lorenz y el índice de Gini.

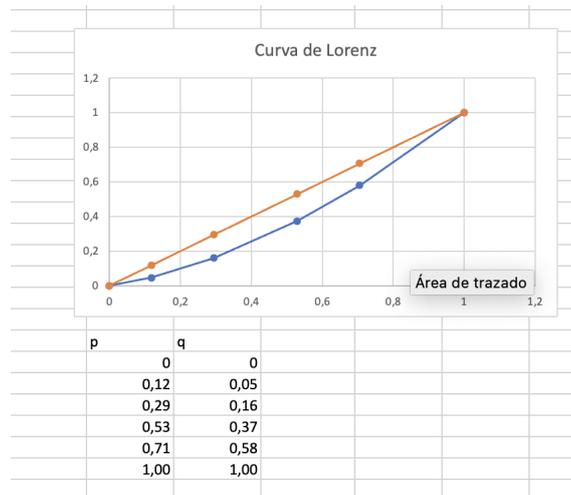
Resolución Utilizar Excel para construir la tabla:

Renta	x_i	n_i	N_i	p_i	$x_i n_i$	U_i	q_i	$p_i - q_i$
500	1000	20	20	0,12	15000	15000	0,05	0,07
1000	1500	30	50	0,29	37500	52500	0,16	0,13
1500	2000	40	90	0,53	70000	122500	0,37	0,16
2000	2500	30	120	0,71	67500	190000	0,58	0,13
2500	3000	50	170	1,00	137500	327500	1,00	0,00

donde se ha definido:

- x_i marca de clase
- N_i frecuencia absoluta acumulada
- $x_i n_i$ renta de la característica x_i
- U_i renta acumulada hasta la característica i -ésima
- p_i frecuencia relativa acumulada
- q_i proporción de renta acumulada
- $(p_i - q_i)$

Definir dos columnas con los valores de (p, q) para construir un **gráfico de dispersión**:



Utilizar la fórmula de Gini, para determinar el valor de índice con la función SUMA.

$$I_G = \frac{\sum_{i=1}^4 (p_i - q_i)}{\sum_{i=1}^4} = 0.2955$$

Seminario Excel II- Estadística I

Grado en Economía

VNIVERSIDAD D SALAMANCA

Mayo 2023

Índice

1	Introducción	1
2	Estadística bivalente	1
3	Regresión	13

1 Introducción

Dedicaremos este segundo Seminario de Estadística a la presentación de las principales herramientas que ofrece Excel para trabajar con estadística bivalente y regresión lineal

2 Estadística bivalente

En esta primera sección construiremos tablas de doble entrada, distribuciones marginales y condicionadas, para una serie de datos. También generaremos las principales representaciones gráficas.

Como en la práctica anterior, trabajaremos con ejemplos como hilo conductor.

Ejercicio 1. En la primera hoja del documento `SeminarioExcel_II.xlsx` figuran las calificaciones de un grupo de 100 alumnos de las asignaturas de Análisis Matemático y Estadística I. Se pide:

1. Construir la tabla de doble entrada asociada.
2. Hallar las distribuciones marginales con sus medias y varianzas.
3. Calcular la distribución de las notas de Análisis Matemático, condicionada a que se ha aprobado Estadística.
4. Construir un diagrama de barras tridimensional.
5. Construir un diagrama de dispersión.
6. Calcular la covarianza.
7. Calcular el coeficiente de correlación lineal.

Resolución

1. La tabla de doble entrada se puede generar a partir de los datos desagregados utilizando la función `CONTAR.SI.CONJUNTO` o mediante el uso de tablas dinámicas. Veamos el primero de los casos. Decidimos escribir las calificaciones de Análisis Matemático por filas, y las de Estadística por columnas, y construimos la siguiente tabla en Excel:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1																	
2																	
3		Calificaciones															
4		Ánalysis Matemático	Estadística														
5		8	9			1) Tabla de doble entrada											
6		7	5														
7		7	8														
8		5	3				AM \ Est	1	2	3	4	5	6	7	8	9	10
9		3	3					1									
10		7	9					2									
11		7	8					3									
12		7	7					4									
13		7	8					5									
14		5	4					6									
15		8	8					7									
16		7	8					8									
17		6	5					9									
18		5	5					10									
19		6	9														
20		7	8														

A continuación en la celda para calcular la frecuencia n_{11} , en la celda G8 escribimos:

$$=CONTAR.SI.CONJUNTO(\$B\$5:\$B\$104;\$F8;\$C\$5:\$C\$104;G\$7)$$

Para entender el comando anterior:

- Busca en la ayuda información sobre la función CONTAR.SI.CONJUNTO
- Razona sobre la colocación de los \$

A continuación extiende la fórmula de la celda G8 horizontal y verticalmente hasta obtener:

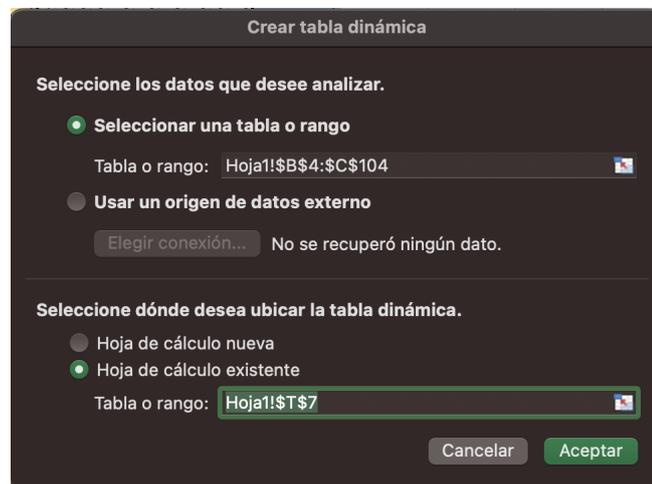
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1																	
2																	
3		Calificaciones															
4		Ánalysis Matemático	Estadística														
5		8	9			1) Tabla de doble entrada											
6		7	5														
7		7	8														
8		5	3				AM \ Est	1	2	3	4	5	6	7	8	9	10
9		3	3					1	0	0	0	0	0	0	0	0	0
10		7	9					2	0	0	0	0	0	0	0	0	0
11		7	8					3	0	2	1	0	0	0	0	0	0
12		7	7					4	0	1	2	2	3	0	0	0	0
13		7	8					5	0	2	4	9	3	2	0	0	0
14		5	4					6	0	0	0	3	9	9	8	2	1
15		8	8					7	0	0	0	4	4	6	9	2	1
16		7	8					8	0	0	0	0	1	2	2	4	0
17		6	5					9	0	0	0	0	0	0	0	0	1
18		5	5					10	0	0	0	0	0	0	0	0	0
19		6	9														
20		7	8														

Finalmente añade la distribuciones marginales, utilizando la función SUMA

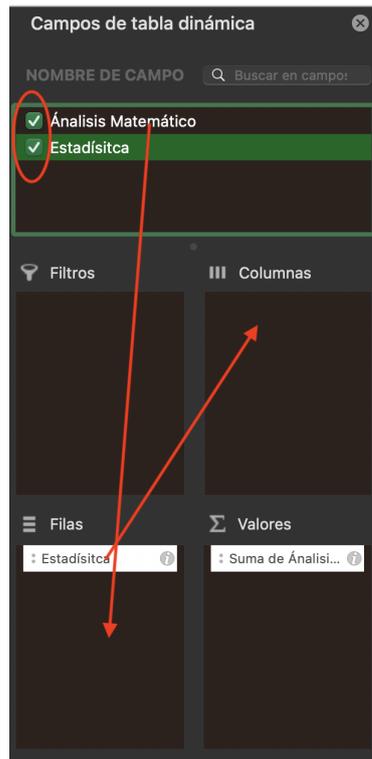
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1																	
2																	
3		Calificaciones															
4		Análisis Matemático	Estadística														
5		8	9			1) Tabla de doble entrada											
6		7	5														
7		7	8														
8		5	3														
9		3	3														
10		7	9														
11		7	8														
12		7	7														
13		7	8														
14		5	4														
15		8	8														
16		7	8														
17		6	5														
18		5	5														
19		6	9														
20		7	8														

AM \ Est	1	2	3	4	5	6	7	8	9	10	fi.
1	1	0	0	0	0	0	0	0	0	0	1
2	0	0	0	0	0	0	0	0	0	0	0
3	0	2	1	0	0	0	0	0	0	0	3
4	0	1	2	2	3	0	0	0	0	0	8
5	0	0	2	4	9	3	2	0	0	0	20
6	0	0	0	3	9	9	8	2	1	0	32
7	0	0	0	0	4	4	6	9	2	1	26
8	0	0	0	0	0	1	2	2	4	0	9
9	0	0	0	0	0	0	0	0	0	1	1
10	0	0	0	0	0	0	0	0	0	0	0
f.j	1	3	5	9	25	17	18	13	7	2	100

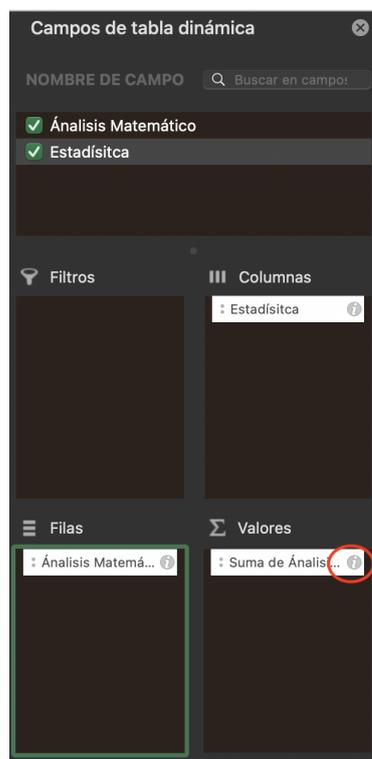
Las tablas de doble entrada también se pueden generar con un tabla dinámica. Pulsamos en la etiqueta insertar, y después en tabla dinámica. Aparece la siguiente ventana, que configuramos del siguiente modo:



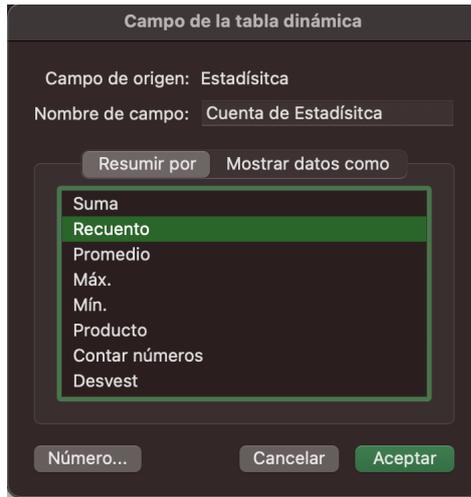
Observar que en el rango de datos hemos incluido el nombre de las series (Análisis Matemático, Estadística). A continuación nos aparece en la parte izquierda de la pantalla el menú de configuración de la tabla dinámica:



Marcamos los dos campos (Análisis Matemático, Estadística), y colocamos Estadística en columnas, y Análisis Matemático en Filas. Ahora, pulsamos sobre la *i*, de la variable suma de Análisis Matemático en la ventana valores:



se despliega una ventana , que configuramos con la función recuento:



De este modo obtenemos la tabla de doble entrada:

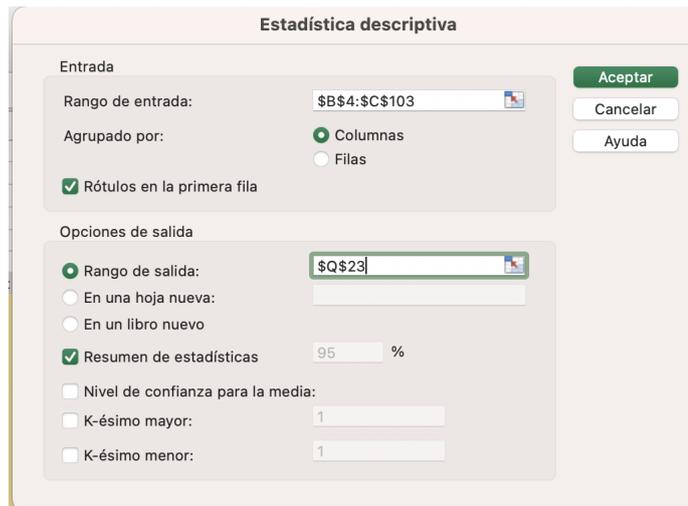
AM\Est	Etiquetas de columna										Total general
Etiquetas de fila	1	2	3	4	5	6	7	8	9	10	
1	1										1
3		2	1								3
4			1	2	2	3					8
5			2	4	9	3	2				20
6				3	9	9	8	2	1		32
7					4	4	6	9	2	1	26
8						1	2	2	4		9
9										1	1
Total general	1	3	5	9	25	17	18	13	7	2	100

Donde en la celda T7 hemos escrito "AM/EST". Notar, que en este caso, si una categoría tiene frecuencia cero, no aparece en la tabla (por ejemplo, las calificaciones 2 y 10 en Análisis Matemático).

- Las distribuciones marginales para las variables consideradas aparecen ya en las tablas que hemos construido en el apartado anterior. Podemos volver a copiarlas utilizando las herramientas de Excel, tal y como aparecen en la siguiente figura.

2) Distribuciones Marginales		Análisis Matemático		Estadística							
AM		Media	5,96969697	Media	6						
1	1	Error típico	0,13426894	Error típico	0,187201284						
2	0	Mediana	6	Mediana	6						
3	3	Moda	6	Moda	5						
4	8	Desviación e	1,33595907	Desviación e	1,862629259						
5	20	Varianza de	1,78478664	Varianza de	3,469387755						
6	32	Curtosis	1,15599285	Curtosis	-0,159499518						
7	26	Coefficiente c	-0,6512161	Coefficiente c	-0,164383527						
8	9	Rango	8	Rango	9						
9	1	Mínimo	1	Mínimo	1						
10	0	Máximo	9	Máximo	10						
	100	Suma	591	Suma	594						
		Cuenta	99	Cuenta	99						
Est											
	1	2	3	4	5	6	7	8	9	10	
	1	3	5	9	25	17	18	13	7	2	100

Para el cálculo de los estadísticos, que ya aparecen en la figura anterior, hemos utilizado el complemento Análisis de Datos, de la pestaña Datos, configurado como aparece en la siguiente imagen:



Notar que Excel, calcula los estadísticos de ambas variables. Por otro lado, se debe tener en cuenta que este complemento devuelve la cuasivarianza (y cuasi-desviación típica) en lugar de la varianza.

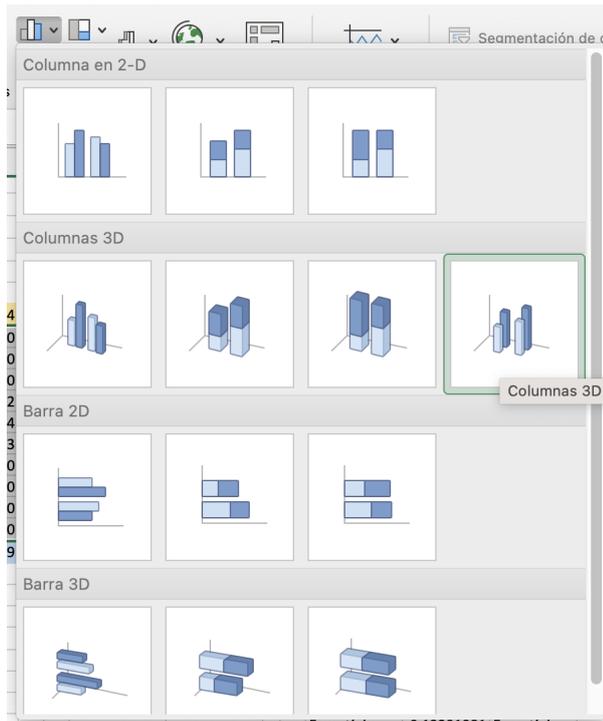
- Para construir la distribución de las notas de Análisis Matemático condicionadas a que se ha aprobado Estadística, podemos utilizar el filtro de la tabla dinámica. Para ello, en primer lugar copiamos la tabla dinámica a la celda F46, pulsamos en el triángulo de “Etiquetas de columna”, configuramos el filtro como mayor o igual que 5. La distribución buscada aparece en el recuadro azul de la siguiente figura.

) Distribución de AM, condicionada a que la nota de Estadística es superior

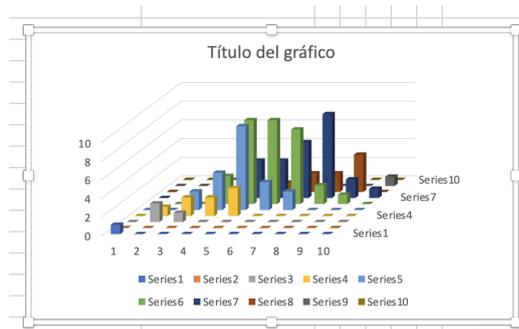
AM\Est	Etiquetas de columna	5	6	7	8	9	10	Total general
4			3					3
5		9	3	2				14
6		9	9	8	2	1		29
7		4	4	6	9	2	1	26
8			1	2	2	4		9
9							1	1
Total general		25	17	18	13	7	2	82

Recordar, que Excel elimina las categorías con frecuencia cero.

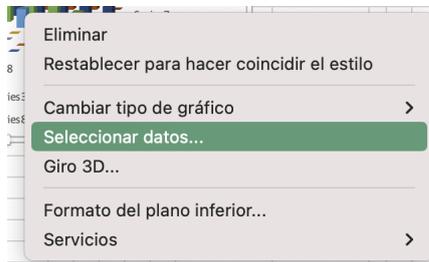
- Excel no incorpora una herramienta para dibujar directamente diagramas de barras 3D. Lo que haremos es dibujar diversas series, y luego modificar los nombres de esta series. En primer lugar seleccionamos la tabla de doble entrada que hemos generado en el primer apartado: G8:P17, e insertamos el gráfico Columnas3D:



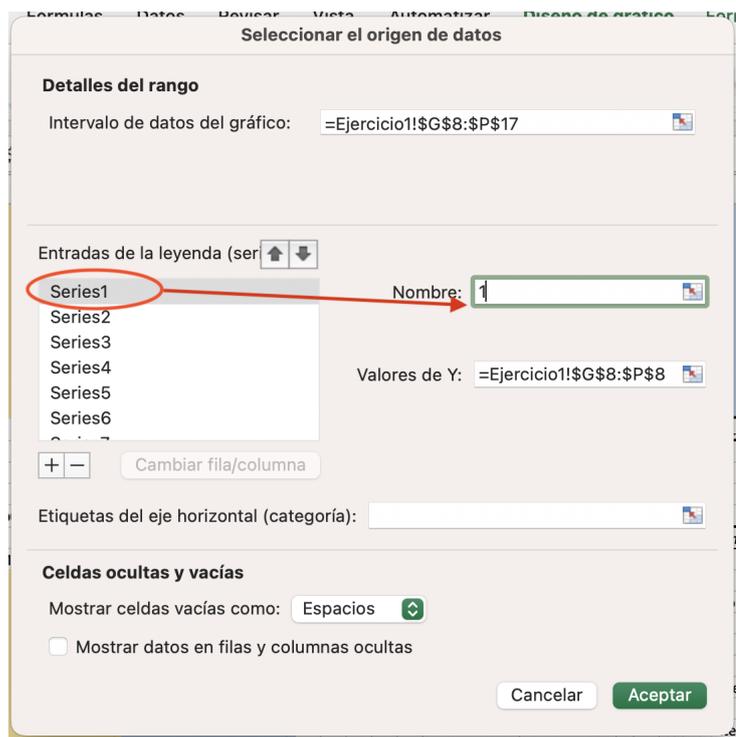
A continuación situamos el gráfico que tendrá un aspecto similar a:



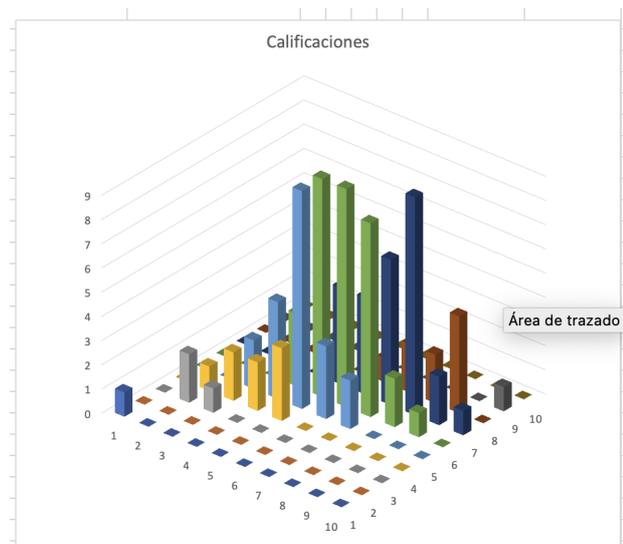
en la hoja de excel. Observamos que en lugar de tener el valor de la categoría, estamos obteniendo "Serie n". Para incluir el valor de la categoría colocamos el ratón sobre el gráfico, y pulsando el botón derecho accedemos a Seleccionar datos:



a continuación damos un nombre a cada una de las series, tal y como se muestra en la imagen:

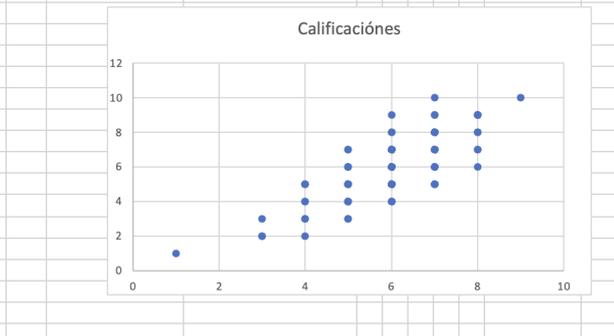


tras estos cambios, y girar el gráfico, podemos obtener algo similar a:

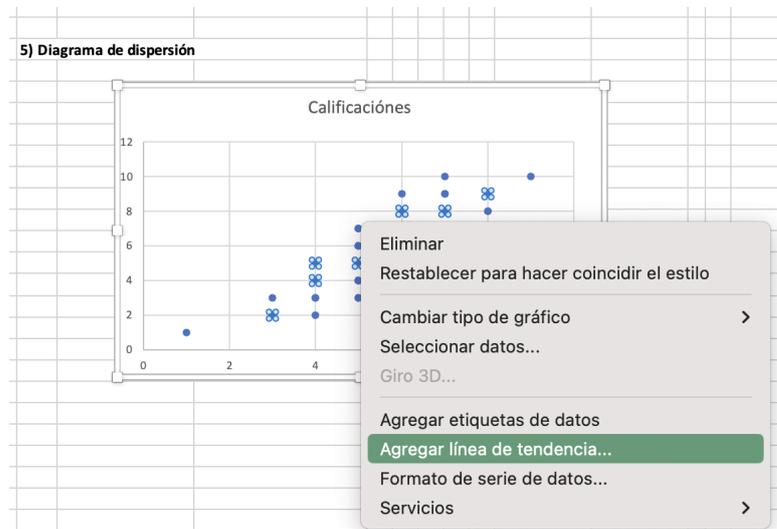


5. Para construir el diagrama de dispersión, seleccionamos los datos originales B5:C104, e insertamos un gráfico de dispersión, para obtener:

5) Diagrama de dispersión



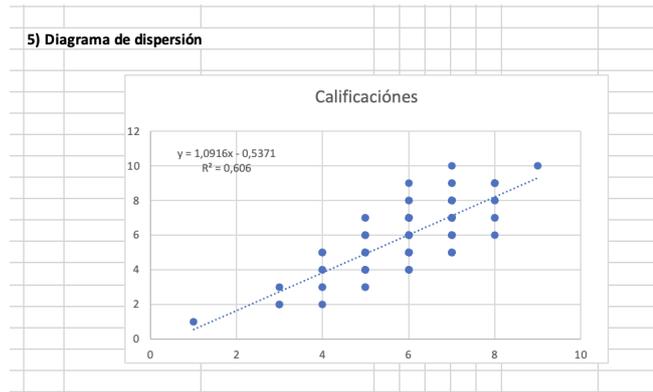
A continuación, situamos el ratón sobre este gráfico y pulsamos el botón derecho para agregar la línea de tendencia



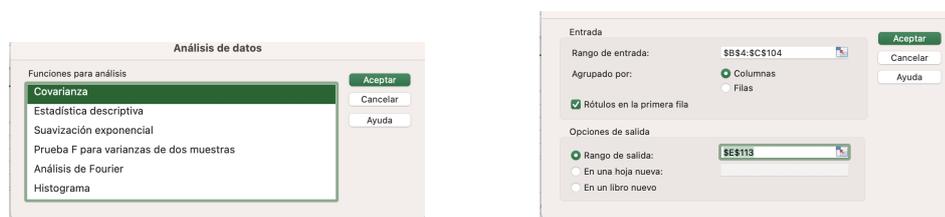
que configuramos del siguiente modo:



de este modo, obtenemos la recta de regresión, su ecuación y el coeficiente de determinación lineal:



6. Para calcular la covarianza utilizamos de nuevo el módulo de Análisis de Datos de la pestaña Datos, que configuramos como sigue:



Esta herramienta nos devuelve la matriz de varianzas-covarianzas. Las varianzas aparecen en la diagonal, y la covarianza, por debajo de la diagonal. En este caso la hemos señalado en amarillo:

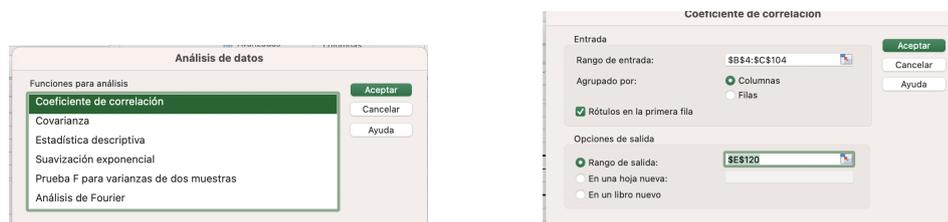
6) Covarianza

	Ánalysis Matemático	Estadística
Ánaly:	1,7491	
Estad:	1,9094	3,4396

Este resultado puede obtenerse directamente con el comando:

$$=COVARIANCE.P(B5:B104;C5:C104)$$

7. Por último, para calcular el índice de correlación lineal, utilizamos también el módulo Análisis de Datos:



De nuevo, este resultado se muestra en formato matricial. El coeficiente de correlación está marcado en amarillo:

7) Coeficiente de correlación		
<i>Análisis Matemático</i>	<i>Estadística</i>	
Ánali:	1	
Estad	0,778458792	1

Este resultado, también puede obtenerse con el comando:

$$=COEF.DE.CORREL(B5:B104;C5:C104)$$

Ejercicio 2. En la segunda hoja del documento *SeminarioExcel.II.xlsx* figuran la evolución semanal de los índices bursátiles:

IBEX NASDAQ NIKKEI

en el año 2022.

1. Generar un gráfico con la evolución de los índices en el año 2022.
2. Calcular las covarianzas:

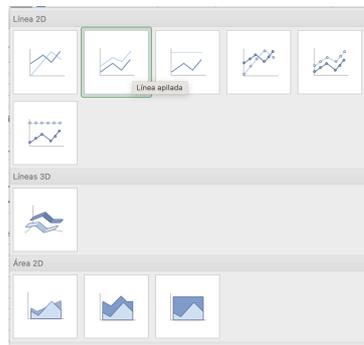
IBEX - NASDAQ IBEX - NIKKEI NASDAQ - NIKKEI

y decidir si están positivamente correlacionados.

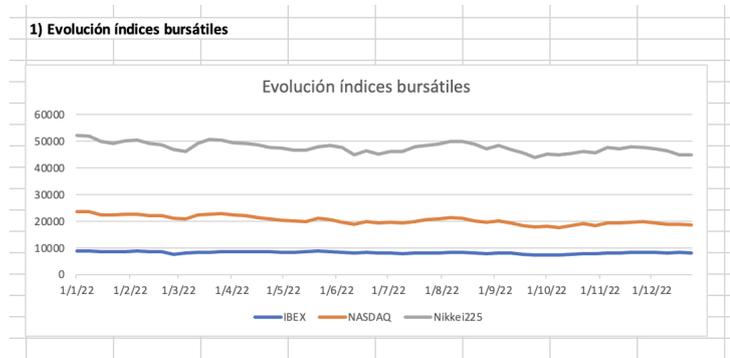
3. Calcular los coeficientes de correlación lineal, y comentarlos.
4. Dibujar los diagramas de dispersión, y relacionarlos con los resultados obtenidos en los apartados anteriores.

Resolución

1. Seleccionamos la tabla de datos A2:D53, e insertamos un gráfico de líneas 2D:



Una vez insertado el gráfico en la hoja de cálculo, accedemos a "Seleccionar datos" (pulsando el botón derecho del ratón) y modificamos el nombre de las series para obtener:



A partir de esta figura, no queda claro si los índices están correlacionados.

- Para calcular las covarianzas, podríamos trabajar dos a dos con cada una de las variables, o hacerlo de una vez, utilizando la herramienta Análisis de Datos. Comentaremos la segunda de las opciones. Para ello, seleccionamos la tabla de datos con títulos A1:D53, y en la etiqueta Datos, pulsamos Análisis de Datos, para seleccionar después Covarianza. El cuadro de dialogo lo configuramos del siguiente modo:

El resultado es la matriz de varianzas-covarianzas:

2) Covarianzas			
	IBEX	NASDAQ	Nikkei225
IBEX	137567,678		
NASDAQ	306897,882	1615700,24	
Nikkei225	62799,4517	246044,388	657952,327

De donde deducimos que los índices están correlacionados positivamente. Para cuantificar el grado de correlación calculamos el coeficiente de correlación de Pearson.

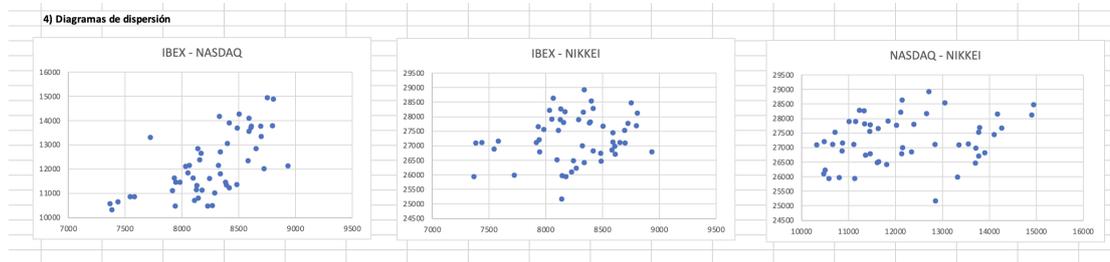
- Procedemos igual que en el caso anterior, seleccionamos la tabla de datos con títulos A1:D53, y en la etiqueta Datos, pulsamos Análisis de Datos, para seleccionar después Coeficiente de Correlación. El cuadro de dialogo lo configuramos del siguiente modo:

El resultado es la matriz de coeficientes de correlación

3) Correlación lineal			
	IBEX	NASDAQ	Nikkei225
IBEX	1		
NASDAQ	0,65096149	1	
Nikkei225	0,20873748	0,23863594	1

De donde deducimos que IBEX-NASDAQ tiene una correlación de cierta entidad, mientras que en los otros casos, no es demasiado relevante.

- En este último apartado, construimos los diagramas de dispersión tomando los datos 2 a 2, para obtener:



Para modificar el rango de los ejes, hemos pulsado sobre el eje x en el gráfico y modificado los límites en opciones del eje.

3 Regresión

En esta sección presentamos las herramientas que incorpora Excel para realizar regresión lineal.

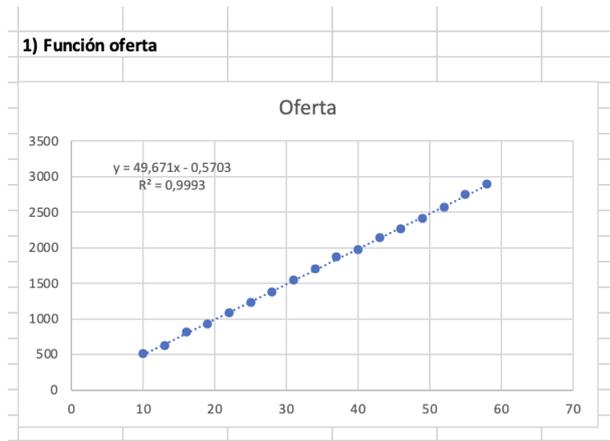
Ejercicio 3. Se dispone de la siguiente información sobre el precio de un determinado artículo, su oferta y demanda.

Precio	Oferta	Demanda
10	516	1749
13	622	1727
16	819	1722
19	928	1633
22	1083	1641
25	1233	1566
28	1377	1475
31	1546	1348
34	1705	1377
37	1875	1188
40	1968	1224
43	2141	1188
46	2261	1118
49	2409	1098
52	2571	1069
55	2753	961
58	2893	876

- Utilizar regresión lineal para estimar la recta de oferta.
- Utilizar regresión lineal para estimar la recta de demanda.
- Calcular el punto de equilibrio.

Resolución

- Seleccionamos las columnas de precio y oferta para crear un diagrama de dispersión. Una vez creado el diagrama de dispersión, seleccionamos los datos sobre el gráfico, pulsamos el botón derecho y seleccionamos "Agregar línea de tendencia", marcamos "Presentar ecuación en el gráfico", y "Presentar el valor R^2 en el gráfico, para obtener:



La herramienta, análisis de datos, permite realizar un análisis complementario. Para ello, pulsamos análisis de datos, y seleccionamos regresión para configurar la ventana emergente como sigue:

Regresión

Entrada

Rango Y de entrada: Ejercicio3!\$B\$2:\$B\$18

Rango X de entrada: \$A\$2:\$A\$18

Rótulos Constante igual a cero

Nivel de confianza: 95 %

Opciones de salida

Rango de salida: Ejercicio3!\$K\$5

En una hoja nueva:

En un libro nuevo

Residuales

Residuos Gráfico de residuales

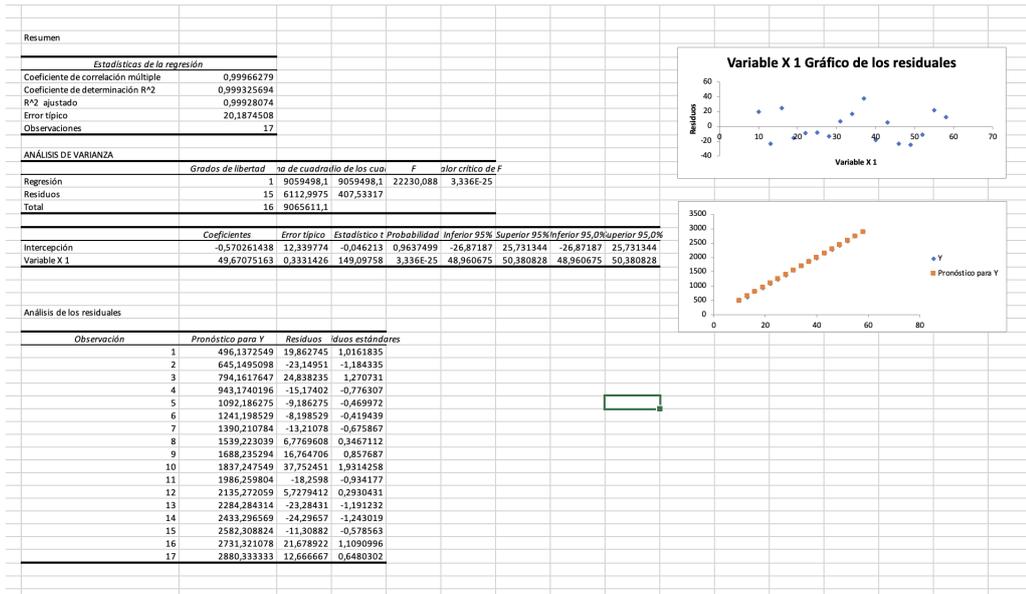
Residuos estándares Curva de regresión ajustada

Probabilidad normal

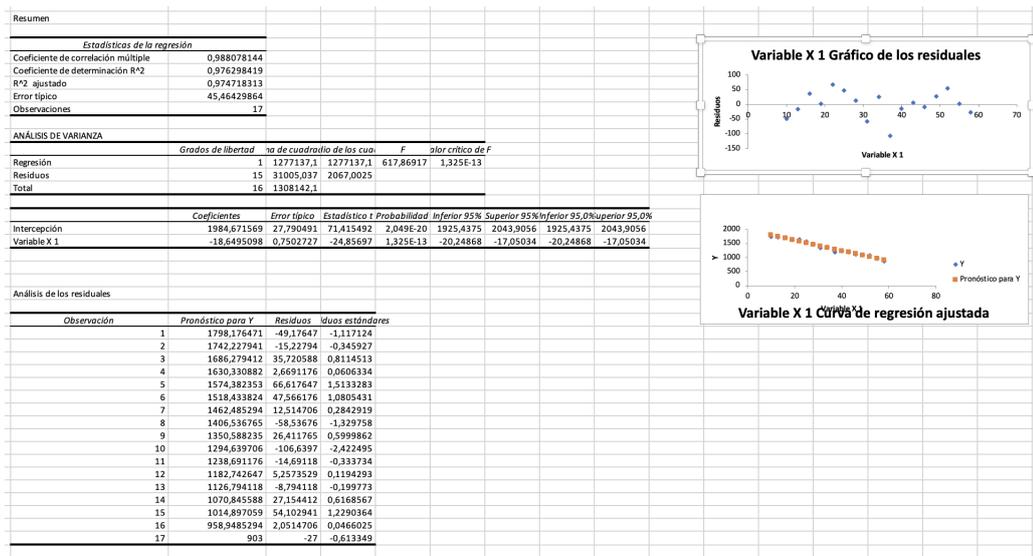
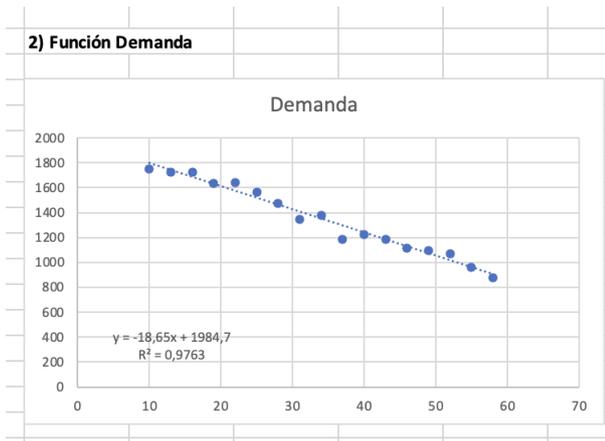
Gráfico de probabilidad normal

Aceptar Cancelar Ayuda

el resultado final es:



2. Repetimos el mismo análisis con los datos de demanda:



3. Para determinar el punto de equilibrio, basta con calcular la intersección:

$$\left. \begin{array}{l} S = 49.67p - 0.57 \\ D = 1984.7 - 18.65p \end{array} \right\} \Rightarrow p = 29.06$$

Ejercicio 4. Se dispone de la siguiente información sobre el precio de un determinado artículo, su oferta y demanda. Además, se conoce que las curvas de oferta y demanda, tiene la forma:

$$S = \alpha_S p_S^{\beta_S} \quad D = \alpha_D p_D^{\beta_D}$$

Precio	Oferta	Demanda
10	192	1493
13	282	1226
16	354	1044
19	492	907
22	658	842
25	776	735
28	943	701
31	1075	660
34	1271	604
37	1466	569
40	1620	529
43	1856	483
46	2056	481
49	2260	475
52	2505	467
55	2703	414
58	2941	379

Utilizar el método de mínimos cuadrados para estimar los coeficientes de las curvas de oferta y demanda.

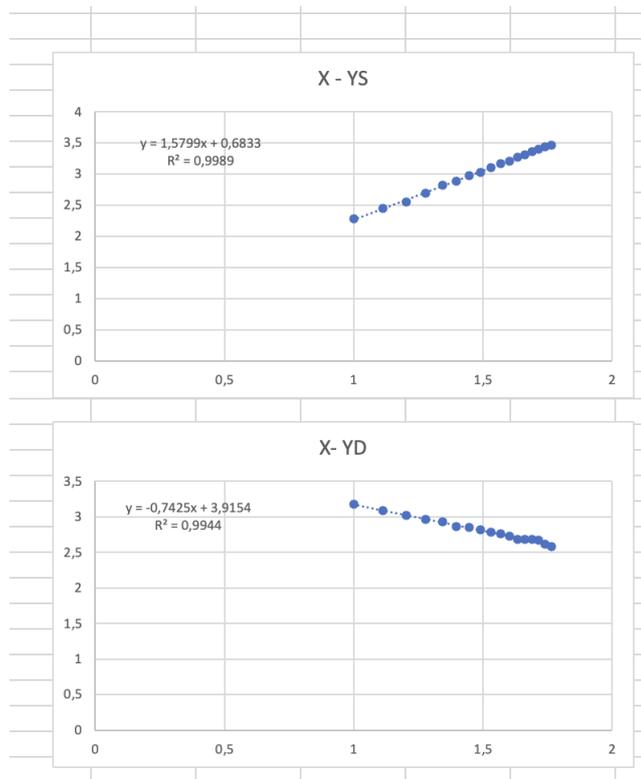
Resolución Reduciremos el problema anterior a un problema de regresión lineal, para ello, observar que si tomamos logaritmos:

$$\begin{aligned} S = \alpha_S p_S^{\beta_S} &\Rightarrow \log S = \log \alpha_S + \beta_S \log p \\ D = \alpha_D p_D^{\beta_D} &\Rightarrow \log D = \log \alpha_D + \beta_D \log p \end{aligned}$$

Es decir, existe una relación lineal entre $Y_S = \log S$ y $X = \log p$ y equivalentemente entre $Y_D = \log D$ y $X = \log p$.

Calculamos una nueva tabla con estos valores y realizamos un análisis como el del ejercicio anterior con las variables (X, Y_S, Y_D) para obtener:

Precio	Oferta = S	Demanda = S	log p = X	log S = YD	Log D = YD
10	192	1493	1	2,28330123	3,17405981
13	282	1226	1,11394335	2,45024911	3,08849047
16	354	1044	1,20411998	2,54900326	3,0187005
19	492	907	1,2787536	2,6919651	2,95760729
22	658	842	1,34242268	2,81822589	2,92531209
25	776	735	1,39794001	2,88986172	2,86628734
28	943	701	1,44715803	2,97451169	2,84571802
31	1075	660	1,49136169	3,03140846	2,81954394
34	1271	604	1,53147892	3,10414555	2,78103694
37	1466	569	1,56820172	3,16613397	2,75511227
40	1620	529	1,60205999	3,20951501	2,72345567
43	1856	483	1,63346846	3,26857797	2,68394713
46	2056	481	1,66275783	3,31302311	2,68214508
49	2260	475	1,69019608	3,35410844	2,67669361
52	2505	467	1,71600334	3,39880773	2,66931688
55	2703	414	1,74036269	3,43184605	2,61700034
58	2941	379	1,76342799	3,46849502	2,57863921

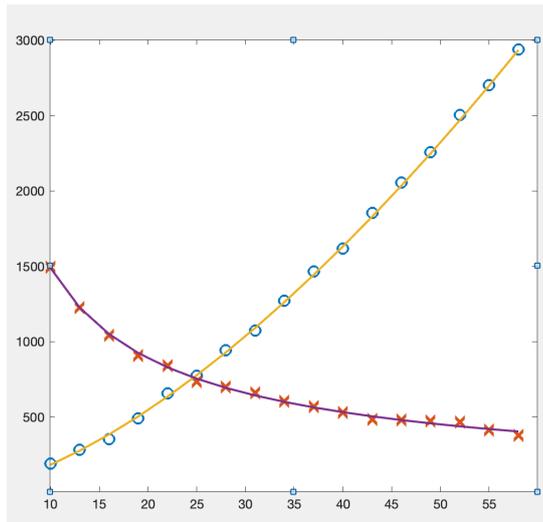


De donde deducimos:

$$\beta_S = 1.579, \quad \log \alpha_S = 0.6833 \Rightarrow \alpha_S = 4.823$$

$$\beta_D = -0.742, \quad \log \alpha_D = 3.915 \Rightarrow \alpha_D = 8222.426$$

El resultado final, del ajuste podemos verlo en el siguiente gráfico:



Seminario Excel III- Estadística I

Grado en Economía

VNIVERSIDAD D SALAMANCA

Mayo 2024

Índice

1	Introducción	1
2	Nota teórica	2
2.1	Ámbitos del indicador	2
2.2	Método general de cálculo	4
2.3	Tasas de variación y repercusiones	5
3	Uso de <i>Excel</i> para el cálculo del IPC en febrero del año 2024	6
3.1	Subclase 0113	7
3.2	IPC Clases	8
3.3	IPC Subgrupos	10
3.4	IPC Grupos	11
3.5	IPC	11
4	Excel: Tasas de variación y repersusiones	12
5	Python: Análisis resultados	13
5.1	Registro y acceso al entorno CoCal	14
5.2	Estructura de los datos	18
5.3	Función <code>plotByPeriod</code>	19
5.4	Función <code>plotBarChart</code>	20
5.5	Función <code>computeCumRi</code>	22
5.6	Función <code>computeMonRi</code>	23

1 Introducción

El índice de precios de consumo (IPC) es un concepto omnipresente en nuestro día a día y fundamental en economía. Este seminario tiene por objetivo desmitificar este indicador económico. Se pretende proporcionar una perspectiva práctica y accesible del concepto del IPC.

El seminario consta de tres partes:

- **Parte 1: Base teórica.** En esta primera parte, se revisan las principales características y aplicaciones del IPC. Se insiste especialmente en la composición de la cesta de bienes y servicios incluidos en el indicador, las técnicas de muestreo, los elementos y el algoritmo de cálculo (para más detalles ver [2]). Esta información sirve como complemento a la que se ha facilitado y presentado en las clases teóricas.
- **Parte 2: Simulación del cálculo.** A continuación, mediante el uso de *excel* se propone la simulación del cálculo del IPC a nivel nacional para el mes de febrero del año 2024. Esta segunda etapa requiere una procedimiento jerárquico que involucra diversos niveles de agregación (artículos, subclases, clases, subgrupos, grupos e índice general). La Figura 1 muestra los diferentes niveles de agregación considerados. Los datos empleados para la

realización del taller se han recuperado de la página web del Instituto Nacional de Estadística (INE), salvo los relativos a los artículos que han sido simulados ¹.

- **Parte 3: Análisis de resultados.** En la última parte, se han implementado diversas estrategias (*excel*, *python*) que permiten un análisis de la información proporcionada por el INE. En primer lugar, se utiliza *excel* para el cálculo de las tasas de variación (mensuales, acumuladas o interanuales) y las repercusiones asociadas de un año concreto. Después, mediante *Jupyter notebook*, se recupera la información sobre el índice general o por grupos, en diferentes estratos (nacional, autonómico o regional) para un amplio rango de fechas (2001-2024).

2 Nota teórica

El índice de precios de consumo, comúnmente conocido como IPC, es un indicador económico crucial que se utiliza para medir el nivel general de precios de una cesta de los bienes y servicios que adquieren habitualmente los hogares en España. Esta cesta incluye artículos como alimentos, ropa, vivienda, transporte, sanidad y educación. Los precios de estos artículos se ponderan según su importancia relativa en el presupuesto de un hogar medio. En España, el Instituto Nacional de Estadística (INE) es el organismo responsable de actualizar la información sobre el IPC.

En España, el IPC se originó en 1940 y tomó como año base el primer año de la Guerra Civil (1936) (ver [5], [1] y [6]). Desde 1940 hasta la fecha actual, se han implementado nueve sistemas de cálculo de índices de precios de consumo, incluyendo el actual (*índice de coste de la vida* hasta 1976). Estos sistemas han tenido períodos base en los años 1936, 1958, 1968, 1976, 1983, 1992, 2001, 2006, 2011, 2016 y 2021 (los períodos base son encadenados desde el año 2001). Actualmente, se utiliza el período base del año 2021, adoptado desde enero de 2022 (ver [3] para más detalles).

El INE actualiza el IPC mensualmente. A finales del mes anterior, se divulga lo que se denomina “dato adelantado”, y posteriormente, entre el 10 y el 15 del mes correspondiente, se publica el dato definitivo. Por lo general, el dato adelantado y el dato definitivo suelen coincidir. Cada mes, en la página web² del INE, se presenta la información mensual, anual y acumulativa del año del IPC.

2.1 Ámbitos del indicador

El campo de consumo del indicador engloba el conjunto de los bienes y servicios³ que los hogares del estrato de referencia⁴ destinan al consumo. Este conjunto está formado desde el año 2021 por 955 artículos (frente a los 977 de la base de 2016) que se encuentran clasificados en 199 subclases (219 en la base 2016), 92 clases, 41 subgrupos y 12 grandes grupos en el conjunto nacional⁵ (ver Tabla 1). El IPC se publica a nivel nacional, autonómico y provincial.

En cuanto al ámbito temporal del indicador, nos encontramos los siguientes periodos importantes (para obtener más información, consultar [2]):

- Periodo base o año de referencia. Aquel en el que el índice se hace igual a 100. Actualmente el año de referencia es el 2021.

¹El Instituto Nacional de Estadística (INE) no facilita los microdatos utilizados en el cálculo (ver [2])

²https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736176802&menu=ultiDatos&idp=1254735976607

³No se consideran los gastos en bienes de inversión, los autoconsumos y autosuministros, ni los alquileres imputados, ni los gastos subvencionados por las administraciones públicas. Tampoco forman parte del campo de consumo algunos impuestos.

⁴La población o estrato de referencia para el cálculo del índice está compuesta por todos los hogares residentes en viviendas familiares en España, tanto hogares urbanos como rurales e independientemente de sus ingresos.

⁵El ámbito geográfico lo constituye todo el territorio nacional.

- Periodo de referencia de los precios. Periodo con cuyos precios se comparan los precios corrientes. Varía cada año y es el mes de diciembre del año inmediatamente anterior al considerado.
- Periodo de referencia de las ponderaciones. Aquel al que están referidas las ponderaciones que sirven de estructura del sistema. A partir del año 2023 la fuente principal utilizada para el cálculo de las mismas pasó a ser la Contabilidad Nacional (CN), en lugar de la Encuesta de Presupuestos Familiares (EPF) como hasta ahora (ver [4]), adaptándose a los criterios establecidos por la reglamentación europea (ver Tabla 1).

Ponderaciones de grupos (tanto por cien)		
Grupo	2022	2023
01. Alimentos y Bebidas no alcohólicas	22,6	19,6
02. Bebidas alcohólicas y Tabaco	3,1	4,0
03. Vestido y calzado	6,0	3,9
04. Vivienda	14,2	12,7
05. Menaje	5,8	5,8
06. Medicina	4,4	6,0
07. Transporte	13,0	13,8
08. Comunicaciones	3,6	3,2
09. Ocio y Cultura	6,4	7,9
10. Enseñanza	1,6	2,0
11. Hoteles, cafés y restaurantes	13,0	13,2
12. Otros bienes y servicios	6,3	7,8

Tabla 1: Ponderaciones de los 12 grandes grupos del IPC para 2023 y su comparativa con las de 2022.

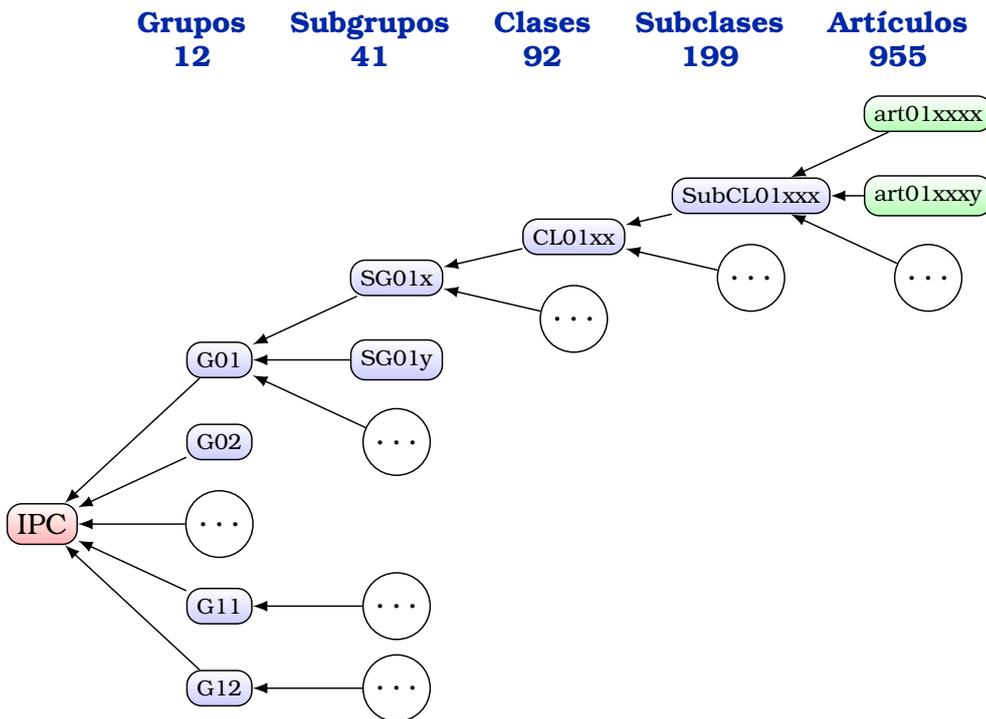


Figura 1: Fases del cálculo del IPC.

2.2 Método general de cálculo

La fórmula empleada para calcular los índices del IPC, base 2021, es un índice de precios complejo, en concreto es un índice de precios de Laspeyres encadenado [2]. Si como es habitual, utilizamos $\{p_k(i), q_k(i)\}$ para denotar el par precios, cantidades, correspondientes al periodo k del artículo o bien i , el índice de Laspeyres encadenado, Le_0^t , con t enlaces o etapas se expresa:

$$Le_0^t = \prod_{k=1}^t \left[\frac{\sum_{i=1}^n p_k(i) q_{k-1}(i)}{\sum_{j=1}^n p_{k-1}(j) q_{k-1}(j)} \right] 100 = \prod_{k=1}^t \left[\sum_{i=1}^n \frac{Ip_{k-1}^k(i)}{100} \omega_L(i; k) \right] 100$$

donde

$$Ip_{k-1}^k(i) = \frac{p_k(i)}{p_{k-1}(i)} \quad \text{y} \quad \omega_L(i; k) = \frac{p_k(j) q_k(j)}{\sum_{j=1}^n p_{k-1}(j) q_{k-1}(j)}$$

siendo $Ip_{k-1}^k(i)$ los índices simples del artículo o bien i -ésimo, y $\omega_L(i; k)$ la ponderación o peso asociado. Por recurrencia, el índice de precios de Laspeyres encadenado, se puede expresar en función del periodo anterior:

$$Le_0^t = \frac{Le_0^{t-1}}{100} \sum_{i=1}^n Ip_{t-1}^t(i) \omega_L(i; t)$$

Como hemos adelantado, en el IPC, el año base se cambia cada 5 años. La última actualización es del año 2021. El enlace es anual, los IPCs mensuales se calculan tomando como referencia los precios del mes de diciembre del año anterior, y las ponderaciones se determinan anualmente utilizando la Contabilidad Nacional.

Con todo ello, el IPC, $I_{21}^{m,t}(A)$, del mes m , año t , para el nivel de agregación A (subclase, clase, subgrupo, grupo, general), se calcula multiplicando el IPC de diciembre del año anterior (referido al periodo base) para ese nivel de agregación A , por un Laspeyres que relaciona los precios/componentes del mes actual con los precios/componentes del nivel de agregación A de diciembre del año anterior, es decir:

$$I_{21}^{m,t}(A) = \frac{I_{21}^{12,t-1}(A)}{100} \sum_{i \in A} I_{12,t-1}^{m,t}(i) \omega_A(i; t)$$

Uno de los objetivos del taller consiste en simular el cálculo del IPC partiendo de los artículos, y agregando de forma sistemática tal y como aparece en la Figura 1. Por tanto, las etapas del cálculo del IPC se resumen en las siguientes:

- **Etapla 1:** Cálculo de los índices elementales o simples de precios. Para el cálculo de la media de los precios $\{p_{m,t}(i; j)\}_{j=1, \dots, k}$ recogidos, se utiliza una media geométrica:

$$Ip_{12,t-1}^{m,t}(i) = \frac{\bar{p}_{m,t}(i)}{p_{12,t-1}(i)} 100 \quad \text{donde} \quad \bar{p}_{m,t}(i) = \sqrt[k]{\prod_{j=1}^k p_{m,t}(i; j)}$$

- **Etapla 2:** Cálculo para cada nivel de agregación,

$$A \in \{ \text{subclases, clases, subgrupo, grupo, general} \}$$

1. Recuperación de las ponderaciones y de los índices elementales o simples.

$$\{\omega_A(i; t), I_{12,t-1}^{m,t}(i)\} \quad \text{para } i \in A$$

2. Cálculo del último eslabón del índice de Laspeyres:

$$I_{12,t-1}^{m,t}(A) = \sum_{i \in A} I_{12,t-1}^{m,t}(i) \omega_A(i; t)$$

3. Cálculo del índice de Laspeyres encadenado (referido a la base):

$$I_{21}^{m,t}(A) = \frac{I_{21}^{12,t-1}(A)}{100} I_{12,t-1}^{m,t}(A)$$

Hay que tener en cuenta que la agregación de los datos publicados en la página web⁶ del INE están referidos al año base (2021) y no son agregables. Para realizar cualquier agregación es necesario recuperar el valor del índice referido al mes de diciembre del año anterior. Para ello, necesitamos los índices del mes actual y del mes de diciembre del año anterior referidos a la base, que pueden recuperarse mediante la siguiente expresión:

$$I_{12,t-1}^{m,t}(i; A) = \frac{I_{21}^{m,t}(i; A)}{I_{21}^{12,t-1}(i; A)} 100$$

En los documentos excel, que utilizamos, se ha precalculado el valor del índice (en los diferentes niveles de agregación) del mes de diciembre del año anterior. De este modo, el procedimiento descrito en esta sección puede ser implementado.

2.3 Tasas de variación y repercusiones

El INE también se encarga de calcular las tasas de variación del IPC a lo largo del tiempo y la repercusión o el impacto que la variación mensual de un artículo tiene sobre el IPC.

Las tasas de variación del IPC son indicadores que muestran cómo el IPC ha cambiado en relación con un período de tiempo anterior o en un período de tiempo específico en comparación con el mismo período del año anterior. En particular, el INE publica tres tipos de tasas de variación, que pueden expresarse en términos de índices absolutos (con respecto a la base), o índices relativos a diciembre del año anterior:

- Tasa de variación mensual:

$$VR_{m-1,t}^{m,t}(A) = \frac{I_{21}^{m,t}(A)}{I_{21}^{m-1,t}(A)} - 1 = \frac{I_{12,t-1}^{m,t}(A)}{I_{12,t-1}^{m-1,t}(A)} - 1$$

- Tasa de variación acumulada:

$$VR_{12,t}^{m,t}(A) = \frac{I_{21}^{m,t}(A)}{I_{21}^{12,t-1}(A)} - 1 = \frac{I_{12,t-1}^{m,t}(A)}{100} - 1$$

- Tasa de variación anual:

$$VR_{m,t-1}^{m,t}(A) = \frac{I_{21}^{m,t}(A)}{I_{21}^{m,t-1}(A)} - 1$$

⁶https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736176802&menu=ultiDatos&idp=1254735976607

El impacto o la repercusión de la variación mensual de un artículo o agregado sobre el índice general se define como la parte de la variación mensual del índice general que corresponde a ese artículo o agregado concreto. Esencialmente, la suma de todos los impactos mensuales de cada artículo de la cesta de la compra es igual a la variación mensual del índice general [2]. En términos más sencillos, el impacto o la repercusión de una variación mensual del precio de un artículo o agregado sobre la variación mensual del índice general es la variación que habría experimentado el índice si todos los demás precios de los artículos hubieran permanecido estables durante ese mes.

Formalmente, las repercusiones, $R_{m',t'}^{m,t}(i; A)$ de un componente i del nivel de agregación A proporciona la contribución (en puntos porcentuales) del bien i a la variación del índice, $VR_{m',t'}^{m,t}(A)$ en el periodo considerado.

- Repercusiones mensuales:

$$\begin{aligned} R_{m-1,t}^{m,t}(i; A) &= \frac{I_{12,t-1}^{m,t}(i; A) - I_{12,t-1}^{m-1,t}(i; A)}{I_{12,t-1}^{m-1,t}(A)} \omega_A(i) \\ &= V_{m-1,t}^{m,t}(i; A) \omega_A(i) \frac{I_{12,t-1}^{m-1,t}(i; A)}{I_{12,t-1}^{m-1,t}(A)} \end{aligned}$$

- Repercusiones acumuladas:

$$\begin{aligned} R_{12,t-1}^{m,t}(i; A) &= \frac{I_{12,t-1}^{m,t}(i; A) - I_{12,t-1}^{12,t-1}(i; A)}{I_{12,t-1}^{12,t-1}(A)} \omega_A(i) \\ &= \frac{I_{12,t-1}^{m,t}(i; A) - 100}{100} \omega_A(i) = V_{12,t-1}^{m,t}(i; A) \omega_A(i) \end{aligned}$$

3 Uso de *Excel* para el cálculo del IPC en febrero del año 2024

En esta sección simularemos la construcción o generación del IPC para un mes concreto (febrero de 2024) a nivel nacional. Pretendemos presentar tanto los elementos involucrados en el cálculo, como el carácter jerárquico en la determinación del mismo (subclases, clases, subgrupos, grupos, índice general).

El procedimiento se ha dividido en varias hojas del cálculo de *excel* que figuran en el documento *SeminarioIII_IPC.xlsx* que puede descargarse desde la página de *Stodium*. El nombre y el propósito de cada hoja es el siguiente:

- Subclase0113. En esta hoja se calculará el IPC de la subclase 0113 del mes de febrero de 2024 referido a diciembre del 2023. Se trabajará con datos ficticios.
- IPC Clases feb24. En esta hoja se calculará el IPC de las 92 clases del mes de febrero de 2024. Se proporcionará información sobre las 199 clases que ha sido recuperada y tratada desde la página del INE.
- IPC Subgrupos feb24. En esta hoja se calculará el IPC de los 41 subgrupos del mes de febrero de 2024 a partir del IPC de las clases. Las ponderaciones se han obtenido de la página del INE, pero el resto de datos se importará desde la hoja anterior.
- IPC Grupos feb24. En esta hoja se calculará el IPC de los 12 grupos del mes de febrero de 2024 a partir del IPC de los subgrupos. Las ponderaciones se han obtenido de la página del INE, pero el resto de datos se importará desde la hoja anterior.
- IPC feb24. En esta hoja se calculará el IPC del mes de febrero de 2024 a partir del IPC de los 12 grupos. Las ponderaciones se han obtenido de la página del INE, pero el resto de datos se importará desde la hoja anterior.

3.1 Subclase 0113

En esta hoja se calculará el IPC de la subclase 0113 del mes de febrero de 2024 referido a diciembre del 2023, a la que denotaremos por:

$$I_{12,23}^{02,24}(S_{0113})$$

La subclase **01113 Pan** esta definida como todo tipo de pan (integral o no) de trigo, centeno, maíz o cualquier otro cereal, rallado o no. Incluye el pan de molde (integral o no), de hamburguesa, de perrito caliente; pan de especias, de pita y pan.

Esta subclase se han generado, **ficticiamente**. El tipo de artículos involucrados (10), los precios y las cantidades en el periodo de enlace (diciembre, 23) y los precios recogidos en el mes de febrero de 2024 han sido simulados para ejemplificar el procedimiento. Por otra parte, como se ha comentado anteriormente, los pesos se determinan anualmente por la Contabilidad Nacional (CN), pero en este caso también los hemos generado de forma ficticia.

A partir de esta información, determinamos:

1. Pesos de los artículos de la subclase. El peso del artículo i , al nivel de agregación de la subclase 01113, vendrá dado por:

$$w_A(i) = \frac{p_0(i)q_0(i)}{\sum_{k=1}^1 p_0(k)q_0(k)}$$

Para calcularlo basta con escribir en la celda E9, la fórmula:

$$=(C9*D9)/SUMAPRODUCTO(C$9:C$18;D$9:D$18)$$

y extender la fórmula hasta la celda E18. Discutir por qué es necesario utilizar referencias absolutas.

De esta manera hemos calculado la ponderación a nivel de la subclase, pero es habitual que los pesos se proporcionen en base 1000 para todos los artículos. Teniendo en cuenta que la clase pan pesa 10,512 sobre 1000, podemos calcular los pesos de cada artículo como:

$$\omega(i) = \omega_{0113}(i) \frac{10,512}{1000}$$

En *excel* basta escribir en la celda F9 $=E9*D5/F5$ y extender hasta la celda F18.

2. Precios medios. Calculamos los precios medios utilizando una media geométrica. Para ello en la celda R24 escribimos:

$$=MEDIA.GEOM(D24:Q24)$$

y extendemos hasta R33.

3. Índices simples de cada uno de los artículos. El índice (simple) de precios de cada artículo lo obtenemos a partir del precio medio, entre el precio a diciembre del 2023:

$$Ip_{12,23}^{02,24}(i) = \frac{p_{02,24}(i)}{p_{12,23}(i)} 100$$

Es decir, en la celda S24 introducimos:

$$=R24/C24*100$$

y extendemos.

4. IPC de la subclase 0113 del mes de febrero de 2024 referido a diciembre de 2023. Construimos el índice compuesto, utilizando los pesos que antes hemos calculado:

$$I_{12,23}^{02,24}(S_{0113}) = \sum_i I_{12,23}^{02,24}(i) \omega_{0113}(i)$$

Con *excel*, escribimos en la celda L9:

$$=SUMAPRODUCTO(S24:S33;E9:E18)$$

A continuación se muestra el resultado.

Ponderación índice general		10,512 sobre				1000	
Tipo	p0	q0	wA	w			
1 Barra de pan	0,800	45,000	0,27498	0,00289			
2 Hogaza de pan	1,800	10,000	0,11457	0,00120			
3 Pan de centeno	1,750	5,000	0,06683	0,00070			
4 Pan de maíz	2,150	8,000	0,13138	0,00139			
5 Pita	1,800	2,000	0,02750	0,00029			
6 Pan congelado	0,650	20,000	0,09930	0,00104			
7 Pan rallado	0,950	3,000	0,01260	0,00013			
8 Pan de molde	1,220	12,000	0,11182	0,00116			
9 Pan de hamburguesa	1,810	8,000	0,11060	0,00116			
10 Pan perritos	1,650	4,000	0,05041	0,00053			

Índice simple de precios		p (Precios observados - febrero 24)										pm (febrero) ip (dic23 - feb24)					
Tipo	p0 (dic 23)																
1 Barra de pan	0,800	0,81	0,82	0,85	0,85	0,74	0,95	0,79	0,81	0,90	0,90	0,73	0,84	0,82	0,77	0,824	102,977
2 Hogaza de pan	1,800	1,50	1,84	1,48	1,61	1,52	1,51	1,58	1,53	1,48	1,61	1,29	1,41	1,56	1,44	1,505	100,325
3 Pan de centeno	1,750	1,86	1,84	1,67	1,95	1,63	1,84	1,86	1,69	1,81	2,03	1,53	1,82	1,72	1,87	1,796	102,656
4 Pan de maíz	2,150	2,19	2,30	2,13	2,17	2,40	2,15	2,16	2,15	2,12	2,14	2,10	2,09	2,15	2,17	2,173	101,091
5 Pita	1,800	1,89	1,88	1,74	2,00	1,82	1,84	1,82	1,85	1,79	1,84	1,72	1,69	1,90	1,86	1,827	101,521
6 Pan congelado	0,650	0,83	0,81	0,46	0,90	0,56	0,82	0,73	0,53	0,63	0,61	0,53	0,71	0,72	0,65	0,644	99,006
7 Pan rallado	0,950	0,47	0,64	0,49	0,59	0,49	0,56	0,56	0,45	0,53	0,55	0,56	0,53	0,40	0,66	0,530	96,335
8 Pan de molde	1,220	1,14	1,30	0,96	0,80	1,07	1,11	1,19	1,20	1,11	1,11	0,98	1,58	1,43	1,19	1,141	93,526
9 Pan de hamburguesa	1,810	1,74	1,95	1,90	1,91	1,91	1,78	1,65	1,84	1,82	1,93	1,84	1,84	1,87	1,81	1,841	101,717
10 Pan perritos	1,650	1,73	1,66	1,85	1,67	1,68	1,98	1,64	1,62	1,78	1,72	1,51	1,67	1,62	1,69	1,655	100,924

Con esta primera hoja de cálculo hemos simulado el cálculo del IPC (referido al periodo de enlace), para una de las 199 clases que maneja el INE. El cálculo es más intensivo, pues se disponen de más datos (el INE recopila mensualmente más de 200000 precios), pero de la misma complejidad.

3.2 IPC Clases

En esta hoja de calculo obtendremos los IPC de las 92 clases, a partir de la información de las subclases. Como hemos visto, en la subsección anterior, este cálculo requiere para cada clase las siguientes operaciones:

1. Recuperación de ponderaciones e índices con respecto al enlace

$$\left\{ \omega_A(i; t), I_{12,23}^{02,24}(i) \right\} \quad \text{para } i \in A, A \in \text{Clase}$$

Hemos recuperado las ponderaciones que el INE ofrece para 2024, y están en la columna B de la hoja. Los índices referidos al enlace (diciembre de 2023), también se han recuperado del INE y están en la columna E de la hoja. Tan solo debemos incluir el que hemos calculado en la hoja anterior. Para ello, escribimos en la celda E11:

$$=Subclase01113!L9$$

En este paso, podemos calcular el IPC de cada subclase, utilizando la fórmula de Laspeyres encadenado. Basta multiplicar, las columnas E y F y dividir por 100:

$$\boxed{=E9*F9/100}$$

y extendemos hasta el final de la tabla.

2. Cálculo del último eslabón del índice de Laspeyres, es decir

$$I_{12,23}^{02,24}(A) = \sum_{i \in A} I_{12,23}^{02,24}(i) \omega_A(i; t) \quad A \in \text{Clases}$$

Nuestro objetivo, es implementar en *excel* una estrategia que permita realizar el cálculo anterior para todas las clases (sería muy tedioso hacerlo para las 92). Para ello, en primer lugar, vamos a incluir en la columna *D*, el código de la clase asociado a cada subclase. Según la codificación del INE, los 4 primeros dígitos de la subclase indican la clase. Procedemos a incluir en la celda *D9* la formula:

$$\boxed{=EXTRAE(A9;1;4)}$$

y extendemos la fórmula hasta el final de la tabla.

Las filas con la misma clase, son las que tenemos que agregar. A continuación, en la columna *I* vamos a colocar los códigos de las clases. Para hacerlo escribimos en la celda *I9*:

$$\boxed{=UNICOS(D9:D207)}$$

Ahora ya estamos en disposición de calcular el último eslabon de Laspeyres, para lo que necesitamos combinar las funciones SUMAPRODUCTO y SI. En la celda *J7* escribimos:

$$\boxed{=SUMAPRODUCTO(SI(D\$9:D\$207=I9;1;0);B\$9:B\$207;E\$9:E\$207)/SUMAPRODUCTO(SI(D\$9:D\$207=I9;1;0);B\$9:B\$207)}$$

Evidentemente, la fórmula puede parecernos compleja, pero como veremos no lo es tanto. Su extensión se debe a que pretendemos que nos sirva para todas las clases. Analicemos el numerador:

- Se están multiplicando 3 columnas:

$$\text{SI(D\$9:D\$207=I9;1;0)} \quad \text{B\$9:B\$207} \quad \text{E\$9:E\$207}$$

- La primera es una columna de ceros y unos. Es un cero si la subclase pertenece a la clase *I9*, y cero en otro caso.
- La segunda columna son las ponderaciones
- La tercera columna son los índices (respecto a los enlaces) de la subclases.

Al multiplicar los tres vectores podríamos pensar que estamos obteniendo el índice (respecto al enlace) de la clase *I9*. Sin embargo, debemos observar que las ponderaciones son globales. Por ello debemos relativizarlas con respecto a la clase que estamos agregando. Eso lo conseguimos con el denominador, que es el producto de dos columnas: la columna pertenencia a la clase *I9* por la columna ponderaciones.

Si extendemos la fórmula hacia abajo, tenemos los índices (respecto al enlace) para todas las clases.

3. Cálculo del índice de Laspeyres encadenado (referido a la base):

$$I_{21}^{02,24}(A) = \frac{I_{21}^{12,23}(A)}{100} I_{12,23}^{02,24}(A), \quad A \in \text{Clases}$$

Para lo cual basta con combinar las columnas J y K, es decir en la celda L9, debemos escribir:

$$=J9*K9/100$$

Los resultados de las primeras son:

Subclases --> Clases		Subclases			Clases				
feb-14		Código Clase	Índice (dic23-feb24)	Índice (21-dic23)	Índice (21-24feb)	Código	Índice (dic23-feb24)	Índice (21-23dic)	Índice (21-24feb)
Índice general	Ponderaciones								
01111 Arroz	1.000,000	01111	99,851	135,159	100,278	01111	100,278	127,912	128,267
01112 Harinas y otros cereales	0,435	01111	98,560	139,358	137,351	01112	99,672	122,456	122,055
01113 Pan	10,512	01111	100,583	120,846	121,555	01113	100,853	120,043	121,067
01114 Otros productos de panadería	8,919	01111	100,173	139,117	139,357	01114	100,151	132,519	132,719
01115 Pizza y quiche	1,630	01111	100,209	121,325	121,578	01115	107,185	206,004	220,806
01116 Pastas alimenticias y cuscús	1,701	01111	100,497	131,507	132,160	01116	99,054	120,454	119,314
01117 Cereales de desayuno	0,867	01111	99,214	124,736	123,755	01117	98,633	130,427	128,644
01118 Otros productos a base de cereales	0,667	01111	99,503	113,489	112,925	01118	103,107	128,661	132,659
01121 Carne de vacuno	6,017	01112	100,295	123,491	123,853	01119	100,824	128,488	129,547
01122 Carne de porcino	6,183	01112	100,229	127,779	128,072	01121	100,875	122,004	123,072
01123 Carne de ovino y caprino	1,327	01112	92,496	133,183	123,189	01122	102,446	123,532	126,553
01124 Carne de ave	8,367	01112	99,390	121,539	120,789	02111	103,969	109,342	113,682
01125 Otras carnes	0,562	01112	89,573	132,293	131,728	02112	103,530	111,012	114,931
01126 Despojos comestibles	0,661	01112	100,525	118,354	118,975	02113	101,677	120,971	123,000
01127 Carne seca, salada o ahumada	16,004	01112	99,837	118,186	117,963	0220	104,583	109,368	114,489
01128 Otros preparados de carne	3,485	01112	100,144	123,266	123,444	03112	86,730	113,279	98,247
01131 Pescado fresco o refrigerado	7,102	01113	101,865	120,138	122,403	03113	82,825	115,877	95,975

3.3 IPC Subgrupos

En esta hoja se calcula el IPC de los 41 subgrupos a partir de la información de las clases. El procedimiento es completamente análogo, y se deja al lector. Solo describimos, de nuevo, las etapas:

1. Recuperación de ponderaciones e índices con respecto al enlace

$$\{\omega_A(i; t), I_{12,23}^{02,24}(i)\} \quad \text{para } i \in A, A \in \text{Subgrupos}$$

Las ponderaciones se han recuperado de la página del INE, el índice $I_{12,23}^{02,24}(i)$ se debe recuperar de la hoja anterior.

2. Cálculo del último eslabón del índice de Laspeyres, es decir

$$I_{12,23}^{02,24}(A) = \sum_{i \in A} I_{12,23}^{02,24}(i) \omega_A(i; t) \quad A \in \text{Subgrupos}$$

Se deben calcular el código del subgrupo, que son los 3 primeros dígitos de la clase y proceder de modo semejante a la hoja anterior.

3. Cálculo del índice de Laspeyres encadenado (referido a la base):

$$I_{21}^{02,24}(A) = \frac{I_{21}^{12,23}(A)}{100} I_{12,23}^{02,24}(A), \quad A \in \text{Subgrupos}$$

El proceso es el mismo que en la hoja de cálculo anterior.

Con todo, obtenemos que el IPC del mes de febrero de 2024 es 113,807. Comprobar que este valor coincide con el publicado por el INE.

4 Excel: Tasas de variación y repersusiones

En esta sección utilizaremos *excel* para calcular las tasas de variación y repersusiones del año 2023.

En la hoja Var_Rep_IPC2023 hemos recuperado el indicador general del IPC de los 12 grupos del mes de diciembre de 2022 y del año 2023 completo (desde la fila 7 a la 21).

Comenzaremos calculando las tasas de variación mensual, para ello debemos calcular:

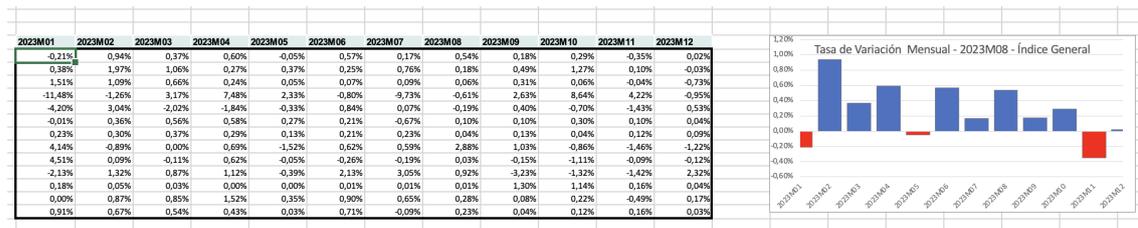
$$VR_{m-1,24}^{m,24}(A) = \frac{I_{21}^{m,24}(A)}{I_{21}^{m-1,24}(A)} - 1 = \frac{I_{12,23}^{m,t}(A)}{I_{12,23}^{23,t}(A)} - 1 \quad A \in \text{Grupos, Índice general}$$

Para implementarlo en *excel*, basta con escribir en la celda C26:

$$=C9/B9-1$$

y extender hacía la derecha y abajo en la tabla de la fila 25 a la 38. Observar que el formato de la tabla se ha fijado en porcentaje.

A la izquierda de la tabla se ha generado una tabla que resumen la evolución de la tasa mensual del índice general. Es importante interpretar los resultados obtenidos.



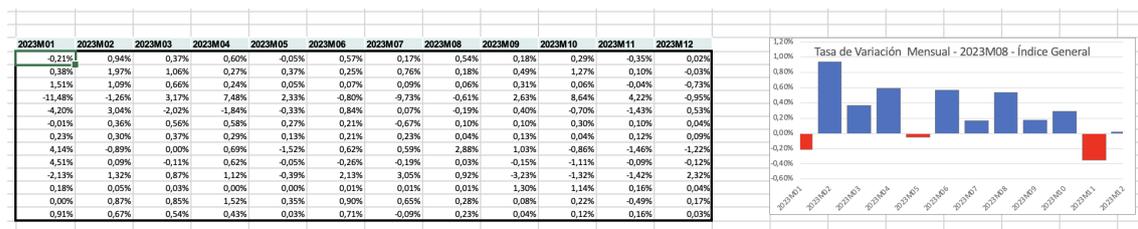
A continuación calculamos la tasa de variación acumulada (con respecto a diciembre de 2023):

$$VR_{12,22}^{m,23}(A) = \frac{I_{21}^{m,23}(A)}{I_{21}^{12,22}(A)} - 1 = \frac{I_{12,22}^{m,23}(A)}{100} - 1$$

Para ello, escribimos en la celda C42:

$$=C9/\$B9-1$$

Debemos darnos cuenta de que bloqueamos la fila y extendemos hacia la derecha y hacia abajo.



En la siguiente tabla calculamos las repercusiones mensuales del año 2023:

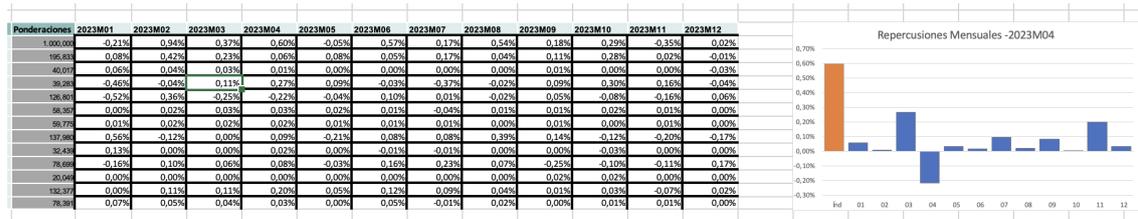
$$R_{m-1,23}^{m,23}(i; A) = V_{m-1,23}^{m,23}(i; A) \omega_A(i) \frac{I_{12,22}^{m-1,23}(i; A)}{I_{12,22}^{m-1,24}(A)}$$

Para lo que escribimos en la celda C58:

$$=C26*\$B58/\$B\$58*B9/B\$9$$

y extendemos a izquierda y derecha (Notar que los pesos están sobre 1000).

A la izquierda de la tabla, figura un gráfico con las repercusiones mensuales del mes de abril. Interpretar los resultados.



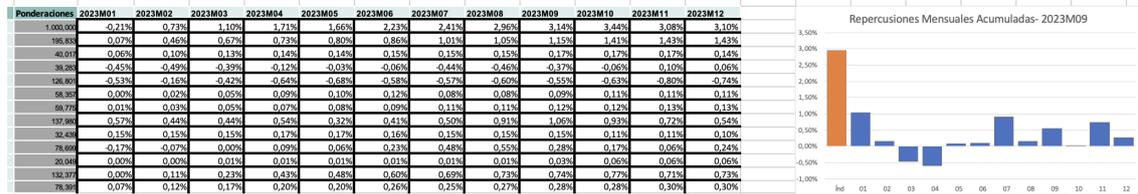
Finalmente calculamos las repercusiones acumuladas en el año 2023:

$$R_{12,23}^{m,24}(i; A) = V_{12,24}^{m,24}(i; A) \omega_A(i)$$

Para lo que escribimos en la celda C74:

$$=C42*\$B74/\$B\$74$$

para obtener:



La gráfica en esta ocasión se ha generado para el mes de septiembre y debemos recordar que hay que interpretar los resultados.

5 Python: Análisis resultados

En la última parte del seminario utilizaremos *python* para el análisis del IPC en un amplio rango de fechas (2001-2014) y en diferentes ámbitos geográficos (nacional, autonómico o regional). Para ello haremos uso del entorno *Jupyter notebook*. El alumno no necesita implementar o interpretar ningún código, tan solo utilizará *python* para analizar y comparar resultados, fomentando su análisis crítico.

Se han desarrollado una serie de funciones que permiten, a partir de la información proporcionada por el INE en formato csv, representar para un año del rango 2001-2023 y distintos ámbitos geográficos (regional, autonómico, nacional) la siguiente información:

- **Índices:** general o grupos.
- **Tasas de variación:** mensuales, acumuladas, interanuales.
- **Repercusiones:** mensuales, acumuladas, interanuales.

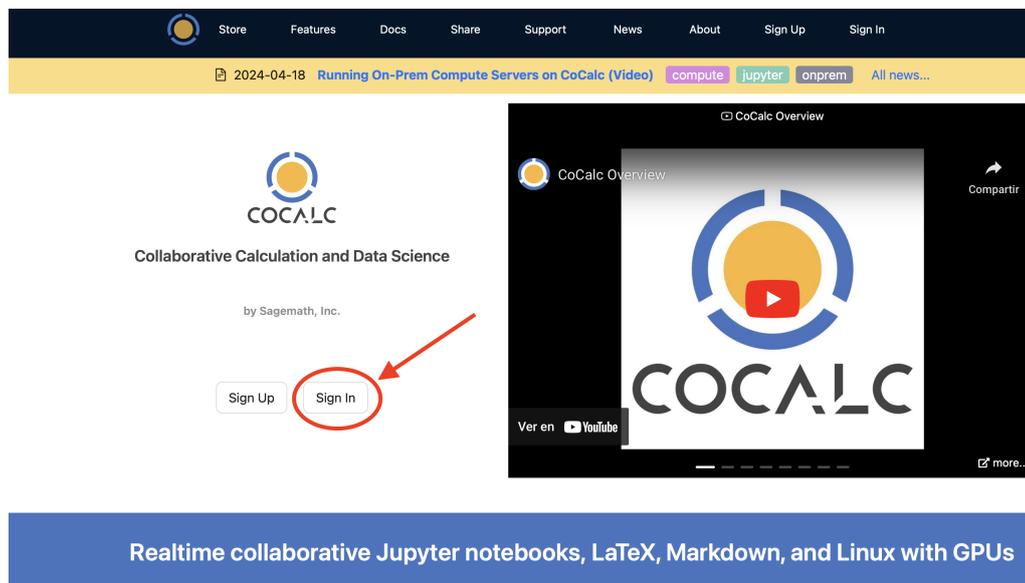
El INE no proporciona las repercusiones, y por lo tanto serán calculadas por nuestra rutina. Para este cálculo son necesarias las ponderaciones, que se actualizan anualmente. En la web del INE, solo se ofrecen las ponderaciones (nacionales, regionales, provinciales) de los tres últimos años (22-24). Si se solicitan repercusiones de años anteriores, las ponderaciones serán aproximadas mediante optimización cuadrática. Dado que los datos se proporcionan con escasa precisión (dos decimales), el resultado proporcionado por nuestras rutinas podría ser poco preciso.

5.1 Registro y acceso al entorno CoCal

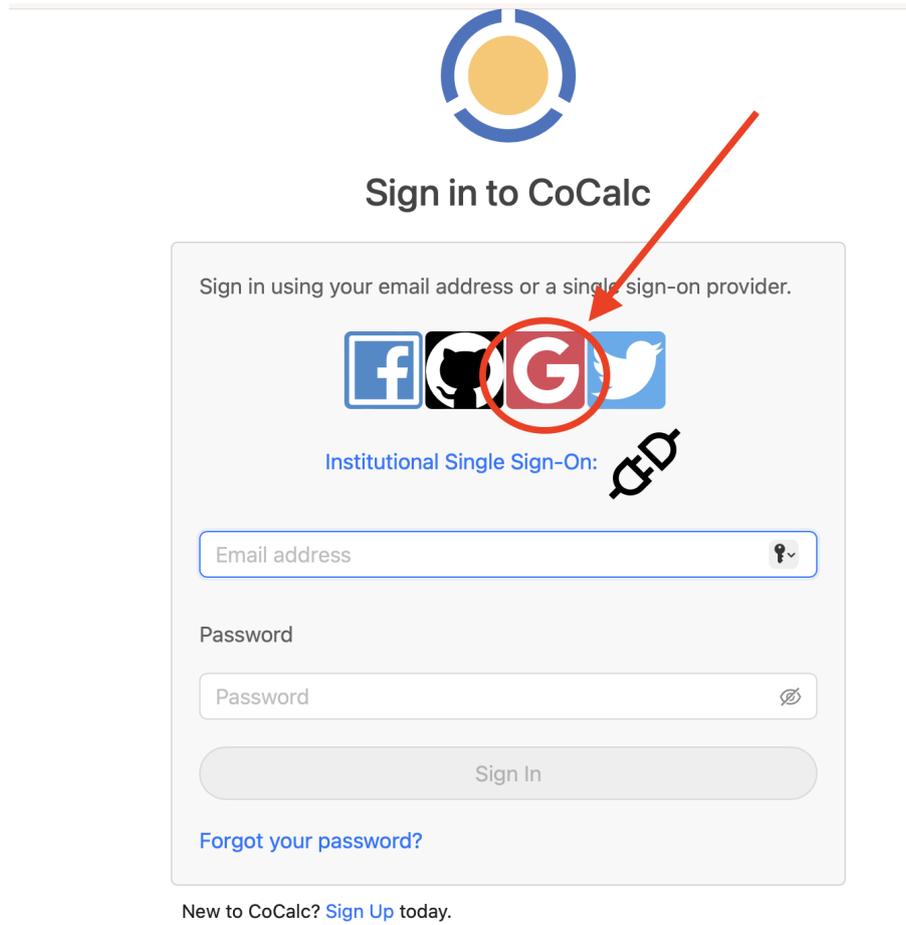
Para trabajar con el entorno *Jupyter*, aconsejamos registrarse en:

<https://cocalc.com/>

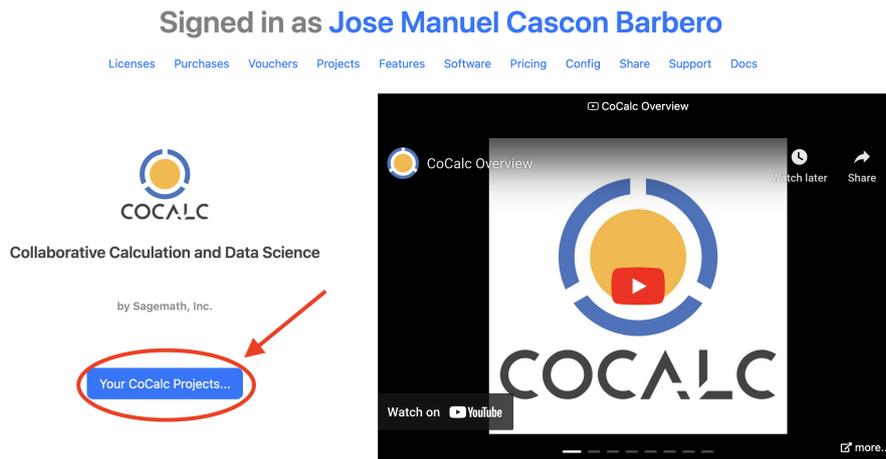
En esta sección indicaremos el proceso. Antes de acceder a la web anterior, aconsejamos acceder a *Stadium* y descargar el archivo `python.zip`, que descomprimiremos en nuestro ordenador.



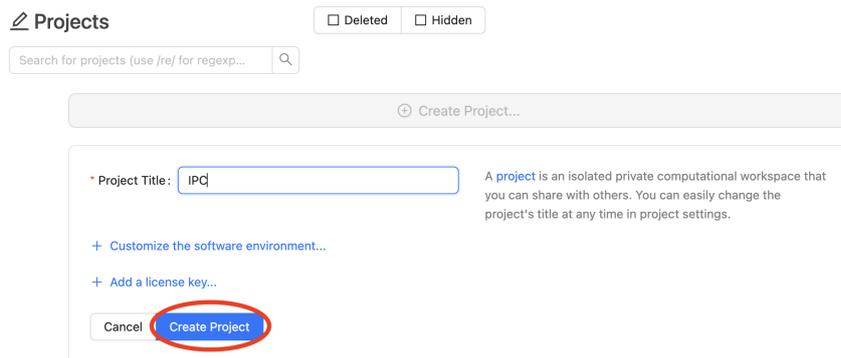
El registro puede completarse con perfiles de *Google*, *Facebook* o *Github* (la cuenta de la Usal puede utilizarse):



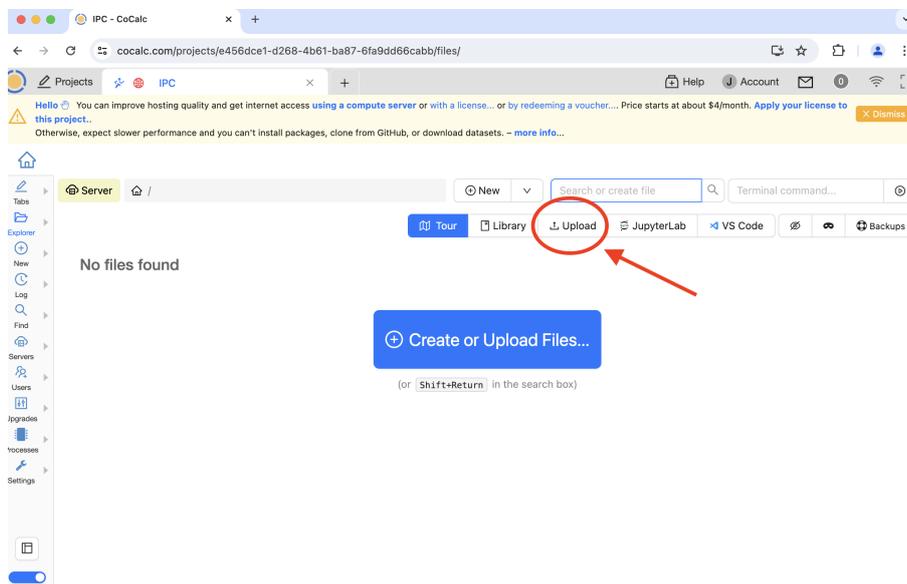
Tras el registro accedemos a nuestro perfil en Cocal:



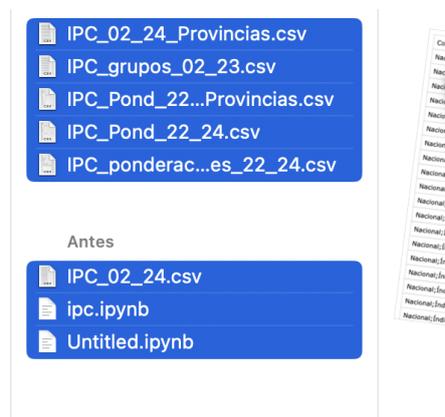
y creamos un nuevo proyecto (p.e. IPC):



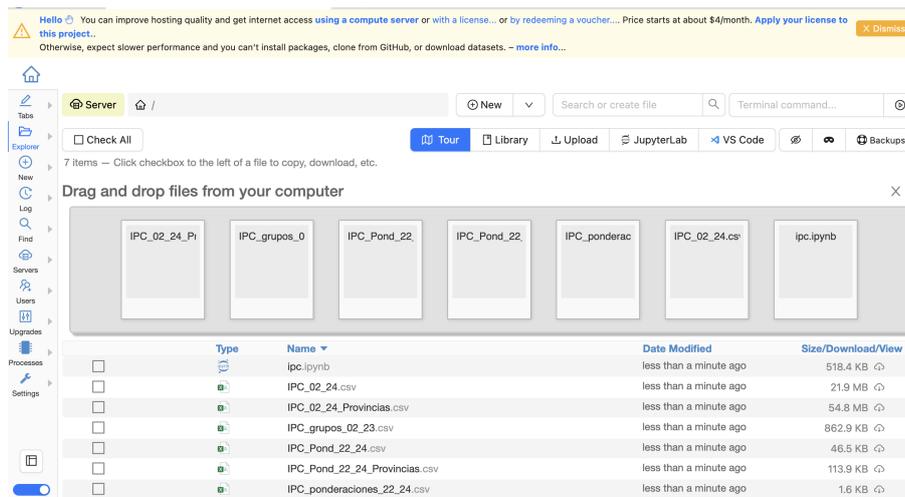
A continuación incluimos en el proyecto los archivos que necesitamos para trabajar y que descargamos de *Stadium* al inicio:



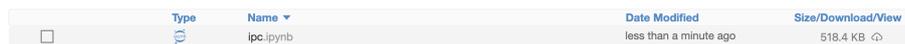
Navegamos para seleccionar los archivos



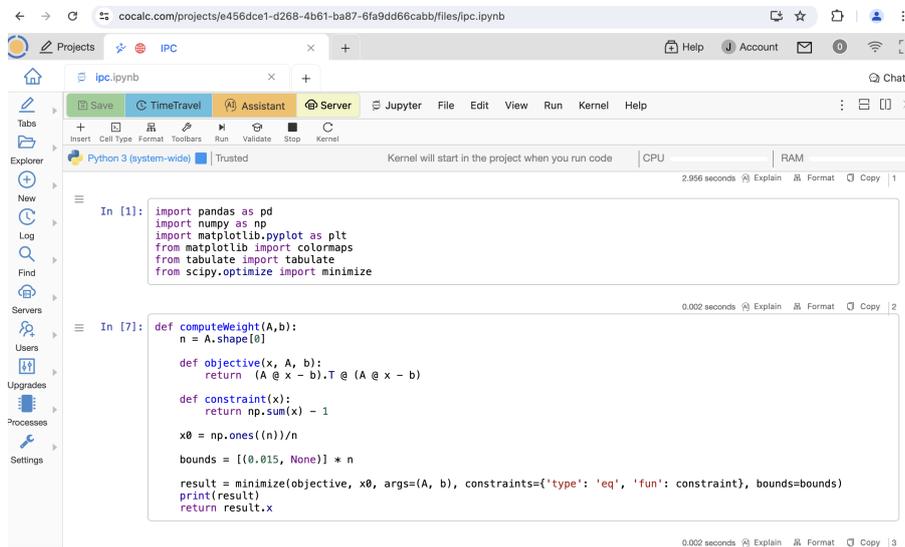
y esperamos a que suban a la plataforma



Por último, hacemos doble clic sobre el archivo con extensión ipynb para iniciar *Jupyter notebook*:



y obtendremos:

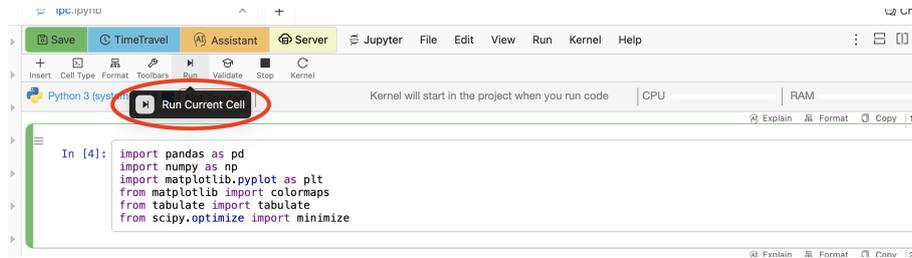


Jupyter notebook tiene aspecto similar a un documento de Mathematica. La información se encuentra organizada en celdas. Para evaluar una celda, nos situamos sobre ella y pulsamos

May+Enter

al igual que en Mathematica. También podemos evaluar la celda pulsando

Run



Evaluamos las 7 primeras celdas, que contienen la importación de las librerías necesarias y las definiciones de las funciones que vamos a utilizar para nuestro análisis.

5.2 Estructura de los datos

Se han descargado los siguientes archivos de la web del INE:

- **IPC_02_24.csv.** Archivo con información mensual desde el año 2002 hasta el año 2024 a nivel autonómico con los siguientes campos, que admiten los siguientes valores:
 - Comunidades y ciudades autónomas: “Nacional”, “01 Andalucía”, ..., “19 Melilla”
 - Grupos ECOICOP: “Índice general”, “01 Alimentos”, ..., “12 Otros bienes y servicios”.
 - Tipo de dato: “Índice”, “Variación mensual”, “Variación anual”, “Variación en lo que va de año”.
 - Periodo: “2002M01”, ..., “2024M12”.
 - Total: Valor del índice o tasa.
- **IPC_Pond_22_24.csv.** Archivo con las ponderaciones de los grupos del año 2022 al año 2024 a nivel autonómico con los siguientes campos, que admiten los siguientes valores:
 - Comunidades y ciudades autónomas: “Nacional”, “01 Andalucía”, ..., “19 Melilla”
 - Grupos ECOICOP: “Índice general”, “01 Alimentos”, ..., “12 Otros bienes y servicios”.
 - Periodo: “2022”, “2023”, “2024”.
 - Total: Ponderación (sobre 1000).
- **IPC_02_24_Provincias.csv.** Archivo con información mensual desde el 2002 hasta el 2024 a nivel provincial con los siguientes campos, que admiten los siguientes valores:
 - Comunidades y ciudades autónomas: “Nacional”, “02 Albacete”, ..., “52 Melilla”.
 - Grupos ECOICOP: “Índice general”, “01 Alimentos”, ..., “12 Otros bienes y servicios”.

- Tipo de dato: “Índice”, “Variación mensual”, “Variación anual”, “Variación en lo que va de año”.
- Periodo: “2002M01”, ..., “2024M12”.
- Total: Valor del índice o tasa.
- `IPC_Pond_22_24_Provincias.csv`. Archivo con las ponderaciones de los grupos del año 2022 al año 2024 a nivel provincial con los siguientes campos, que admiten los siguientes valores:
 - Comunidades y ciudades autónomas: “Nacional”, “02 Albacete”, ..., “52 Melilla”.
 - Grupos ECOICOP: “Índice general”, “01 Alimentos”, ..., “12 Otros bienes y servicios”.
 - Periodo: “2022”, “2023”, “2024”.
 - Total: Ponderación (sobre 1000).

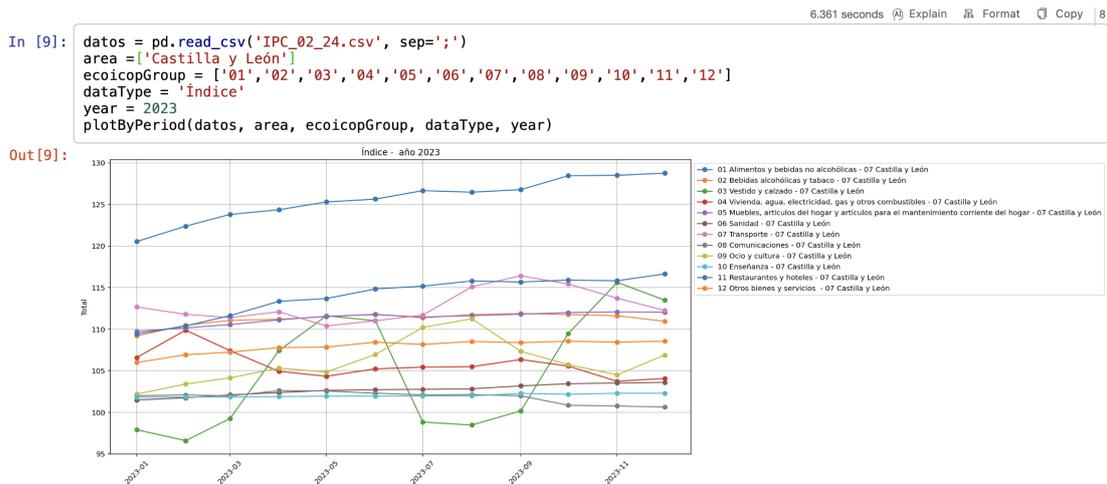
5.3 Función `plotByPeriod`

Esta función dibuja la evolución anual de un grupo ECOICOP en un ámbito geográfico de un tipo de dato. Este tipo de gráficos son recomendables para representar números índice. Los argumentos de la función son los siguientes:

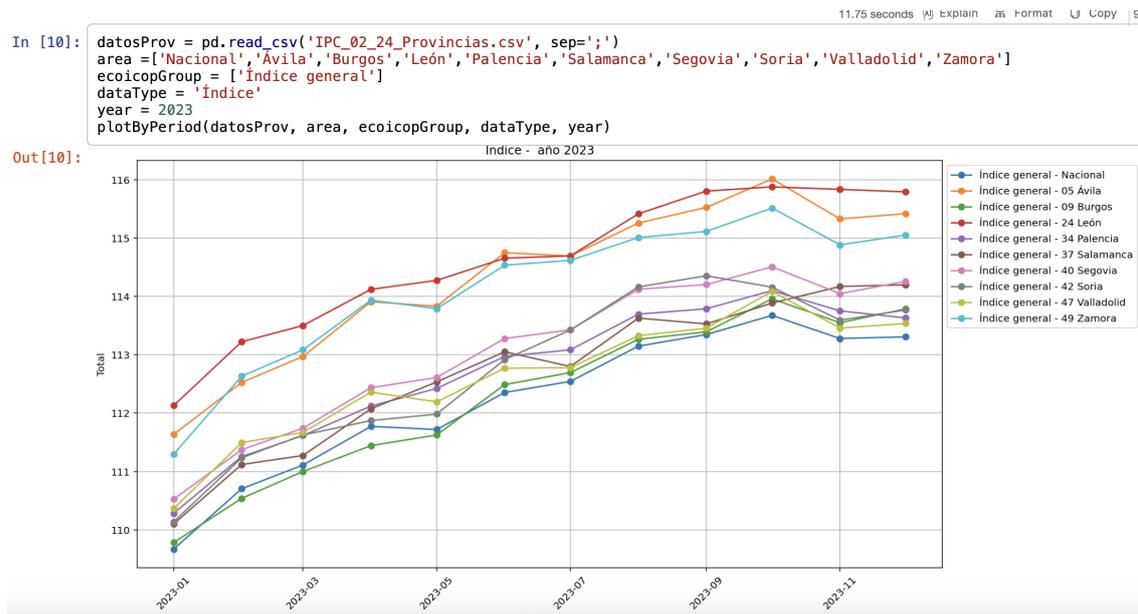
- `dataframe`: tipo de dato (dataframe) que usa `python` para leer archivos `.csv`.
- `regions`.
- `ecoicopGroup`.
- `dataType`.
- `year`.

Aconsejamos completar cada uno de estos campos antes de llamar a la función. A continuación figuran dos ejemplos de uso.

En el primer ejemplo se muestra el IPC del año 2023 para los 12 grupos a nivel de Castilla y León. Observar que no es necesario completar de forma precisa el grupo o la comunidad autónoma, puesto que el código buscará el registro que contenga la cadena de caracteres especificada.



En el segundo ejemplo se muestra el IPC del año 2023, en el ámbito regional (notar que ahora se ha recuperado la información de las provincias).



Sería recomendable modificar los ejemplos anteriores.

5.4 Función plotBarChart

Esta función dibuja un diagrama de barras de un grupo ECOICOP en un ámbito geográfico de un tipo de dato. Este tipo de gráfico es recomendable utilizarlo para representar las variaciones mensuales, anuales y las variaciones en lo que va de año. La función también imprime en pantalla los datos solicitados.

Los argumentos de la función son los siguientes:

- dataframe: tipo de dato (dataframe) que usa *python* para leer archivos *.csv*.
- regions.
- ecoicopGroup.
- dataType.
- year.

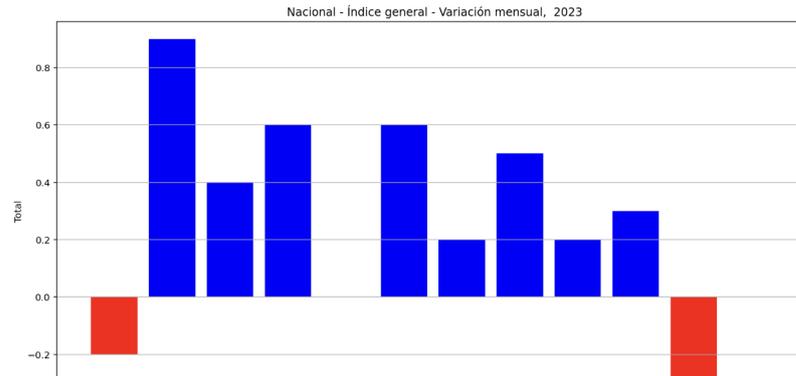
Aconsejamos completar cada uno de estos campos antes de llamar a la función. A continuación figuran dos ejemplos de uso.

En el primer caso, se representa la tasa de variación mensual del IPC (“Índice general”) a nivel nacional.

```
In [10]: ecoicopGroup = 'Índice general'
dataType = 'Variación mensual'
region = 'Nacional'
year = 2023
plotBarChart(datos, region, ecoicopGroup, dataType, year)

Out[10]:
```

	Comunidades y Ciudades Autónomas	Grupos ECOICOP	Tipo de dato	Periodo	Total
0	Nacional	Índice general	Variación mensual	2023M01	-0.2
1	Nacional	Índice general	Variación mensual	2023M02	0.9
2	Nacional	Índice general	Variación mensual	2023M03	0.4
3	Nacional	Índice general	Variación mensual	2023M04	0.6
4	Nacional	Índice general	Variación mensual	2023M05	0
5	Nacional	Índice general	Variación mensual	2023M06	0.6
6	Nacional	Índice general	Variación mensual	2023M07	0.2
7	Nacional	Índice general	Variación mensual	2023M08	0.5
8	Nacional	Índice general	Variación mensual	2023M09	0.2
9	Nacional	Índice general	Variación mensual	2023M10	0.3
10	Nacional	Índice general	Variación mensual	2023M11	-0.3
11	Nacional	Índice general	Variación mensual	2023M12	0

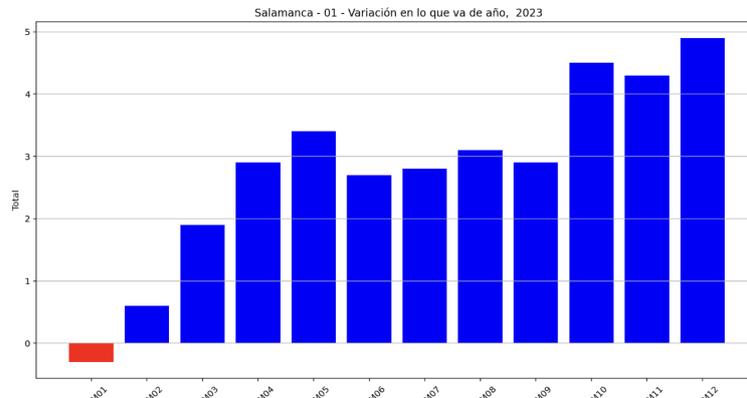


A continuación se representa tasa de variación acumulada (“Variación en lo que va de año”), en la provincia de Salamanca en el año 2023 para el grupo 01.

```
In [14]: ecoicopGroup = '01'
dataType = 'Variación en lo que va de año'
region = 'Salamanca'
year = 2023
plotBarChart(datosProv, region, ecoicopGroup, dataType, year)

Out[14]:
```

	Provincias	Grupos ECOICOP	Tipo de dato	Periodo	Total
0	37 Salamanca	01 Alimentos y bebidas no alcohólicas	Variación en lo que va de año	2023M01	-0.3
1	37 Salamanca	01 Alimentos y bebidas no alcohólicas	Variación en lo que va de año	2023M02	0.6
2	37 Salamanca	01 Alimentos y bebidas no alcohólicas	Variación en lo que va de año	2023M03	1.9
3	37 Salamanca	01 Alimentos y bebidas no alcohólicas	Variación en lo que va de año	2023M04	2.9
4	37 Salamanca	01 Alimentos y bebidas no alcohólicas	Variación en lo que va de año	2023M05	3.4
5	37 Salamanca	01 Alimentos y bebidas no alcohólicas	Variación en lo que va de año	2023M06	2.7
6	37 Salamanca	01 Alimentos y bebidas no alcohólicas	Variación en lo que va de año	2023M07	2.8
7	37 Salamanca	01 Alimentos y bebidas no alcohólicas	Variación en lo que va de año	2023M08	3.1
8	37 Salamanca	01 Alimentos y bebidas no alcohólicas	Variación en lo que va de año	2023M09	2.9
9	37 Salamanca	01 Alimentos y bebidas no alcohólicas	Variación en lo que va de año	2023M10	4.5
10	37 Salamanca	01 Alimentos y bebidas no alcohólicas	Variación en lo que va de año	2023M11	4.3
11	37 Salamanca	01 Alimentos y bebidas no alcohólicas	Variación en lo que va de año	2023M12	4.9



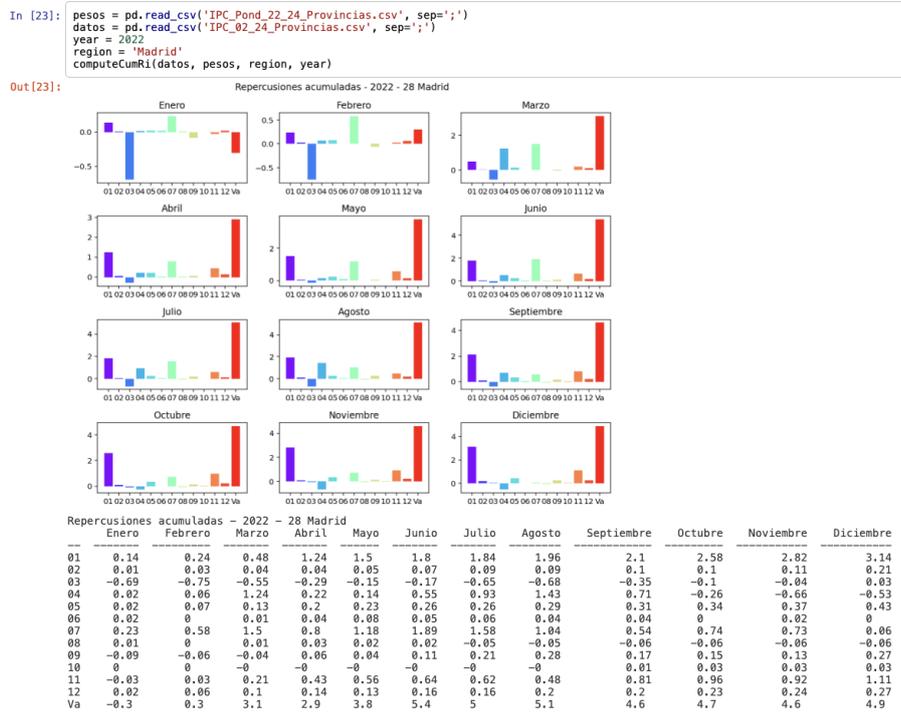
5.5 Función computeCumRi

Esta función calcula las repercusiones acumuladas de los grupos al IPC (índice general) para un ámbito geográfico los 12 meses del año.

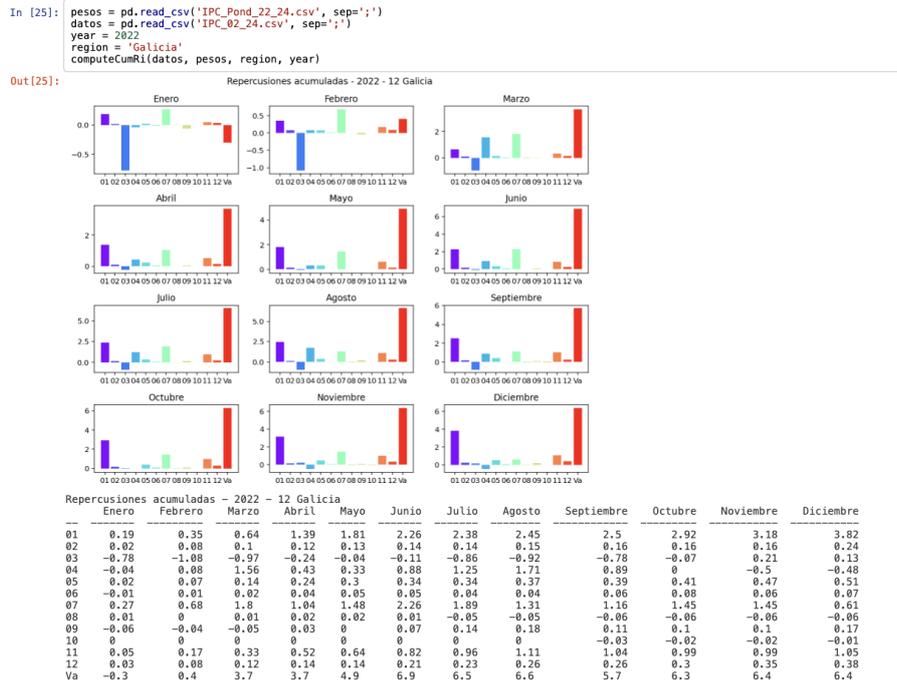
Los argumentos de la función son los siguientes:

- dataframe: tipo de dato (dataframe) que usa *python* para leer archivos *.csv*.
- weight: ponderaciones. Si las ponderaciones no se proporcionan se aproximan (peligroso).
- region.
- year.

En el ejemplo siguiente se muestran las repercusiones acumuladas del año 2022 para Madrid. Notar que en este caso es necesario leer tanto los datos, como las ponderaciones.



A continuación figuran las ponderaciones acumuladas para la comunidad autónoma de Galicia del año 2022. Notar que en este caso se han recuperado las ponderaciones en el ámbito nacional.



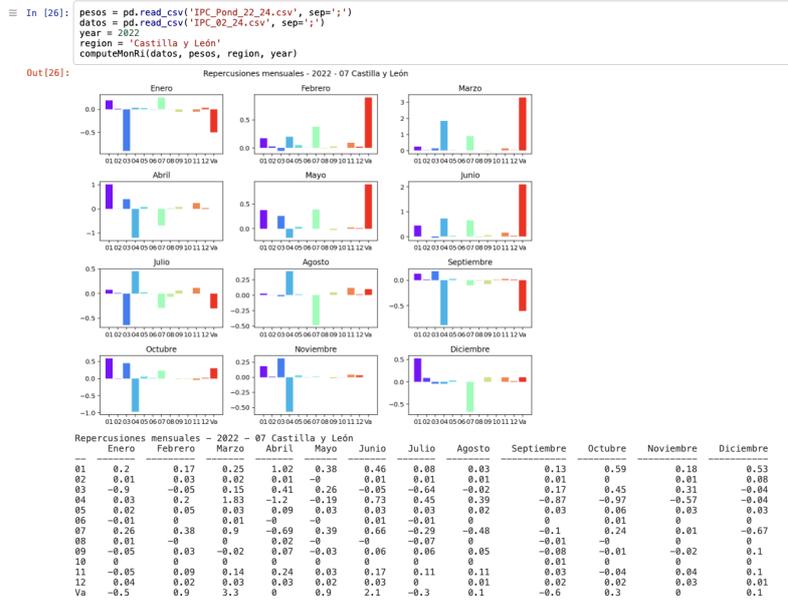
5.6 Función computeMonRi

Esta última función calcula las repercusiones mensuales de los grupos al IPC (índice general) para un ámbito geográfico, los 12 meses del año.

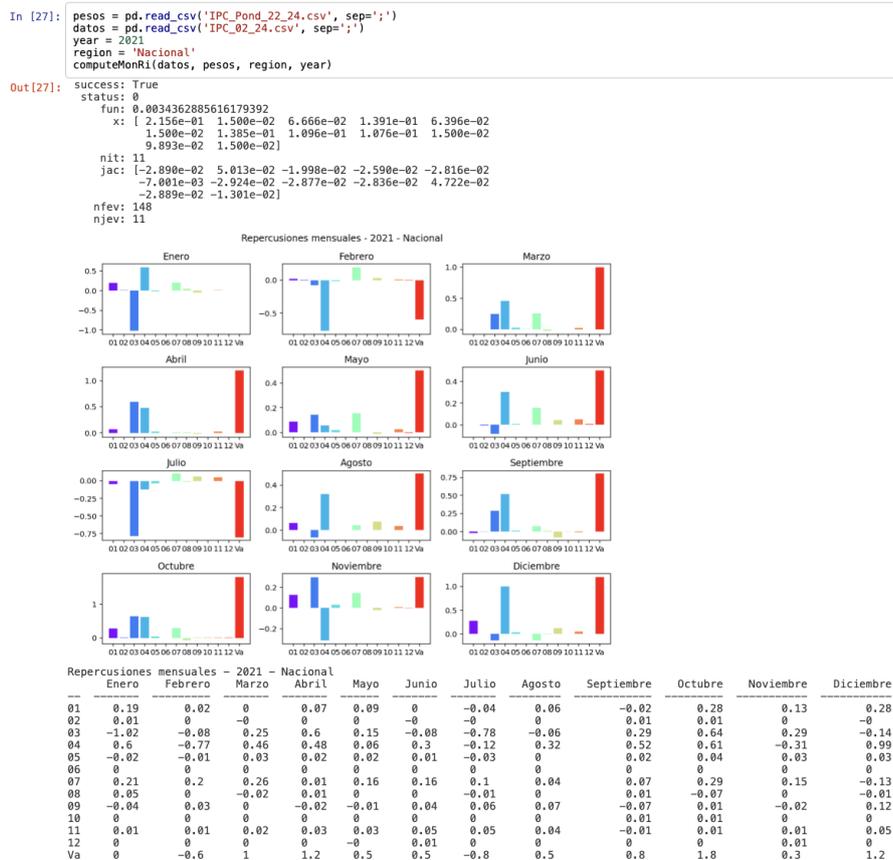
Los argumentos de la función son los mismos que en el caso anterior

- dataframe: tipo de dato (dataframe) que usa *python* para leer archivos *.csv*.
- weight: ponderaciones. Si las ponderaciones no se proporcionan se aproximan (peligroso).
- region.
- year.

En el siguiente gráfico figuran las repercusiones mensuales de los grupos de Castilla León para el año 2022.



Por último, representamos las repercusiones mensuales de los 12 grupos a nivel nacional para el año 2021. Notar que el código avisa, de que las ponderaciones han sido aproximadas.



REFERENCIAS BIBLIOGRÁFICAS

- [1] R. Escuder Vallés and J. S. Murgui Izquierdo. *Números Índices. Estadística Aplicada Economía y Ciencias Sociales*. 2011.
- [2] INE. Subdirección General de Estadísticas de Precios y Presupuesto Familiares. *Índice de Precios de Consumo. Base 2016. Metodología*, 2017.
- [3] INE. Subdirección General de Estadísticas de Precios y Presupuesto Familiares. *Principales novedades metodológicas del Índice de Precios de Consumo Base 2021, 2022*.
- [4] INE. Subdirección General de Estadísticas de Precios y Presupuesto Familiares. *Novedades metodológicas en el IPC*, 2023.
- [5] F. Jareño Cebrián. Cambio de metodología en la elaboración del IPC y su impacto sobre la respuesta sectorial ante anuncios de inflación. *ICE, Revista De Economía*, 1(836):127–140, 2007.
- [6] Jordi Maluquer de Motes. La inflación en España. Un índice de precios de consumo, 1830-2012. *Estudios de Historia Económica. Banco de España*, (64):1–143, 2013.
- [7] OIT, FMI, OCDE, EUROSTAT, ONU, and BM. *Manual del índice de precios al consumidor. Teoría y práctica*. 2006.