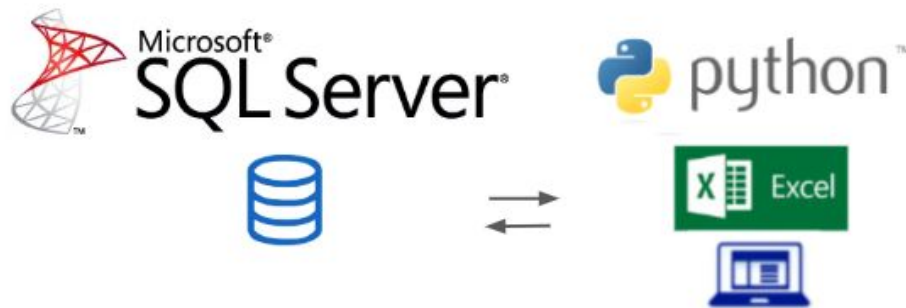


Distribución de Presupuesto Comercial

Este documento describe la implementación de un sistema de distribución de presupuesto comercial (SDPC) para Mastellone Hnos. S. A.. El objetivo es facilitar la ejecución de tareas recurrentes en el proceso de logística, y añadir nuevas herramientas que mejoren la eficacia y la precisión en la asignación del presupuesto.



El SDPC consiste en una interfaz basada en Microsoft Excel para acomodar la experiencia del usuario, mientras que las herramientas del lenguaje Python otorgan las herramientas de cómputo por detrás. El SDPC ofrece distintas etapas, que incluyen la automatización de algunas tareas sistemáticas, y la optimización numérica de la distribución comercial en un problema de investigación de operaciones. Esta optimización permite minimizar el margen de error con respecto a la distribución indicada por el usuario, contemplando las restricciones arbitrarias que sean requeridas.

Índice

1. Requisitos
2. Instalación
3. Instrucciones y funcionalidades
4. Ejemplo paso a paso
5. Explicación formal de la optimización
6. Resultados de ejemplo

Requisitos

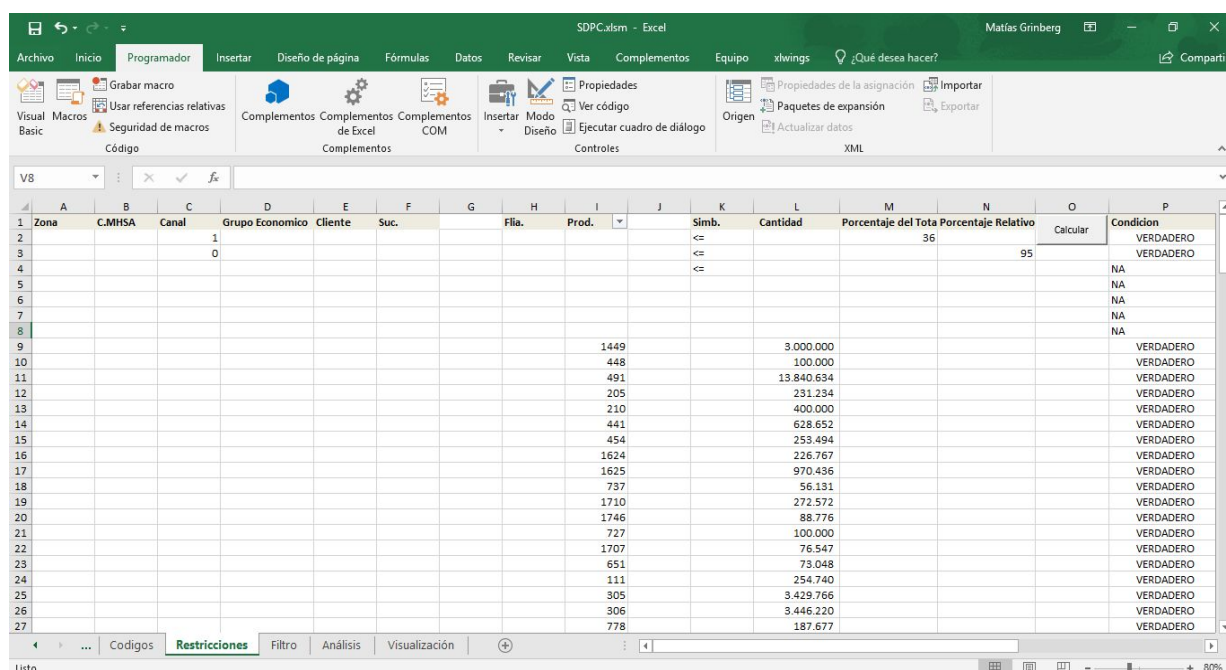
- Sistema operativo Windows >2000+
- RAM >=16gb preferiblemente
- Instalación Microsoft Excel >2007

Guía de instalación

- Se debe contar con una instalación de Python >3.5 y la librería *xlwings* y *pandas* con sus dependencias.
- De no tener una instalación de Python:
 - o Descargar [Anaconda con Python >3.5](#)
 - o Instalar Anaconda con la opción de agregar el PATH como variable de entorno • luego de instalar ingresar en el Anaconda Prompt:
`python -c "import sys; print(sys.executable)"`
y copiar la dirección devuelta al lado de “Interpreter” en el archivo “xlwings.conf”
- En el archivo “xlwings.conf” reemplazar el valor para *PYTHONPATH* por la ruta a la carpeta con “SDPC.py”, y la ruta a “log.txt” en la variable *Log File*
- En la consola de Python o el Anaconda Prompt ingresar:

```
pip install xlwings, pandas
xlwings addin install
```

Instrucciones y funcionalidades



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
	Zona	C.MHSA	Canal	Grupo Economico	Cliente	Suc.	Flia.	Prod.	Simb.		Cantidad	Porcentaje del Tota	Porcentaje Relativo		Calcular	Condicion
1																
2				1					<=				36			VERDADERO
3				0					<=					95		VERDADERO
4									<=							NA
5																NA
6																NA
7																NA
8																NA
9								1449			3.000.000					VERDADERO
10								448			100.000					VERDADERO
11								491			13.840.634					VERDADERO
12								205			231.234					VERDADERO
13								210			400.000					VERDADERO
14								441			628.652					VERDADERO
15								454			253.494					VERDADERO
16								1624			226.767					VERDADERO
17								1625			970.436					VERDADERO
18								737			56.131					VERDADERO
19								1710			272.572					VERDADERO
20								1746			88.776					VERDADERO
21								727			100.000					VERDADERO
22								1707			76.547					VERDADERO
23								651			73.048					VERDADERO
24								111			254.740					VERDADERO
25								305			3.429.766					VERDADERO
26								306			3.446.220					VERDADERO
27								778			187.677					VERDADERO

ETL

En la hoja inicial se lee la base realizando una query a SQL Database o leyendo los archivos *lst* ubicados en la carpeta */tmp* , generando una base limpia *Historico.csv* y el [Maestro]. A los datos brutos se les aplica una limpieza descartando los registros de ventas negativas y eliminando los registros inválidos, y se calcula la columna de cantidad neta como la diferencia entre venta y devolución.

La segunda funcionalidad permite agregar **nuevos clientes** y **productos**, copiando la distribución de clientes y productos de referencia. Se modifica a *Historico.csv* aplicando las siguientes modificaciones: copia de distribución de nuevos productos y clientes, remoción de productos sin presupuesto en Restricciones y de clientes ausentes en el [Maestro].

Tablas

En estas hojas se ingresan las tablas maestras para clientes y para productos. Con estos datos se actualiza la base histórica, y se calcula la nueva en función de las restricciones. Se espera una fila por cliente y sucursal en la primera, y una fila por producto en la segunda, evitar filas repetidas o datos incompatibles.

Restricciones

La tabla de restricciones consiste en **3 grupos de columnas** delimitados por una columna vacía. Desde la izquierda, el primer grupo selecciona un subconjunto de clientes y sucursales, el segundo selecciona un producto o una familia de productos, y el tercero indica una igualdad, o un límite superior en caso de indicar "**<=**". La restricción de una fila particular significa entonces que la suma total de ese producto o familia para ese subconjunto de clientes y sucursales no puede superar ese límite superior, indicado en unidades o en un porcentaje relativo al total de ese producto o familia. El **presupuesto** se extrae de esta tabla. Solo se toman en cuenta valores numéricos positivos.

Análisis y salida

Se presentan recortes del resultado a distintos niveles de agregación para comparar la distribución original, la distribución directa, y el resultado calculado.

La hoja [Filtro] permite filtrar los resultados de esta agregación, para enfocar el análisis en un subconjunto dado.

Por último la hoja de [Visualización] muestra la distribución de diferencias proporcionales entre el resultado y el punto de partida junto con algunas métricas numéricas sobre el ajuste incluyendo el total de diferencia como la norma 2 de $|q-q_0|$, y el error con respecto a las restricciones como la norma de $Aq-b$.

El resultado final es el archivo *Q7.csv* donde se encuentra la cantidad correspondiente de cada producto para cada cliente y sucursal, y el **maestro actualizado**.

Comentarios generales de uso:

- El proceso completo demora ~10min. La primer función "Cargar Data" demora 3-7 minutos en la computadora de prueba para 2 archivos .lst. La cantidad de archivos máximos posibles depende de la memoria RAM a disposición. La función "Actualizar Base" demora ~50s. La optimización, "Calcular", demora entre 2 y 5 minutos.
- Las funciones dentro de cada hoja asumen la existencia de los archivos generados por las hojas previas.
- Excel estará bloqueado durante la ejecución de las funciones, alertando con una ventana al finalizar donde se deberá clicar "Aceptar"
- En *variables.py* pueden cambiarse las columnas a seleccionar (ej. Kg. o Cantidades) y cambiar la información de conexión a SQL Database.

Ejemplo de uso paso a paso:

- Abrir interfaz de Excel: **"SDPC.xlsm"**
Siguiendo el orden de las hojas:
- [ETL] Se cargan los **.lst** que se encuentren en la carpeta /tmp/, y se genera **Historico.csv** y el **Maestro** con el botón **leer base**.
- Ahora para continuar con el proceso, aplicando las modificaciones pertinentes, debemos:
 - [Restricciones] Ingresar el **presupuesto**, es decir las restricciones de forma producto \leq unidades. Este presupuesto se extrae de las restricciones **chequeando que haya valores solo en esas dos**

columnas. Para **eliminar un producto** simplemente se debe no incluirlo en el presupuesto.

- **Agregar la información de clientes y sucursales nuevos en [Maestro]**, indicados en la primera página. De no ingresarse el cliente en el maestro, se **eliminará** al no tenerse los datos para la restricción, Agregar clientes requiere 2 pasos: **[ETL]** “Distribuir como”, que copia todas las columnas del cliente y sucursal indicado como referencia, dando la distribución de productos para ese cliente, y **modificar el Maestro**, para corregir la información de ese cliente que no es igual al de referencia.
- Para cada **producto nuevo** (incluido en el **presupuesto**) indicamos en **[ETL]** que se distribuya entre los clientes igual que un producto de referencia.
- Para cada **cliente nuevo** copiamos la distribución de productos de un cliente de referencia. Se puede ingresar solo el cliente (donde se copian las sucursales que tenga) o un **cliente-sucursal** usando un separador no numérico. El resto de la información del cliente, se debe agregar en **[Maestro]**.
- Agregar las restricciones necesarias en **[Restricciones]**. Una vez ingresadas, el botón **calcular** ejecuta el algoritmo de distribución (NRAS + IterProj). El botón **consultar** permite recuperar los resultados y con ello apagar la máquina virtual de la API. El resultado se guarda como **Q1.csv** en la carpeta */tmp*.
- Una vez resuelta la optimización, usamos la pestaña **[Filtro]** para tomar un subconjunto de clientes y focalizar el análisis.
- Por último evaluamos el resultado en las pestañas **[Análisis]** y **[Visualización]**.

Optimización

El problema de distribución consiste en repartir un presupuesto dado entre todos los puntos de venta, respetando restricciones para grupos arbitrarios y minimizando la variación con respecto a la distribución anterior. Estos productos y puntos de venta están organizados jerárquicamente en distintos niveles como: productos, familia de productos, sucursales, clientes, zonas, canales,

circuitos y grupos económicos. El programa recibe restricciones a aplicar sobre estos niveles, ej: “el canal 1 no puede superar el 40% del total”, o “el presupuesto del producto 1 es de 100” y genera el problema de optimización correspondiente.

El problema admite distintas formulaciones. Se contemplaron diferentes enfoques, incluyendo la minimización de la divergencia KL entre las distribuciones mediante distintos *solvers* de optimización convexa, aproximaciones tipo *round robin*, y varios algoritmos específicamente confeccionados para la tarea. Se optó por un algoritmo ad hoc, relacionado al método RAS, que otorga un resultado de alto ajuste y en un tiempo acotado en comparación a otros enfoques que alcanzaban varias horas de demora y requerían mayor capacidades de cómputo. Un segundo método basado en álgebra vectorial puede opcionalmente “pulir” el resultado del primero.

El enunciado toma la siguiente forma: se tienen n incógnitas correspondientes a las cantidades para cada producto en cada sucursal, representadas por $\mathbf{q} \in \mathbf{R}^n$. Se tienen dos vectores de restricciones $\mathbf{b}, \mathbf{d} \in \mathbf{R}^n$. Cada uno de los niveles mencionados (ej. ‘canal’) corresponde simplemente a subconjuntos de estas variables, pudiendo representarse como un vector $\mathbf{r} \in \{0,1\}^n$ tal que $r_i = 1$ si n_i pertenece al canal 1, y siendo por lo tanto $\mathbf{r}^T \mathbf{q}$ el total de ese canal. De este modo se conforman las matrices $\mathbf{R}, \mathbf{M} \in \{0,1\}^n$ que “seleccionan” las variables relevantes para las ecuaciones e inecuaciones.

Se busca el vector \mathbf{q} “más cercano” \mathbf{q}_0 que satisfaga las restricciones, con la salvedad de que la distancia que se busca minimizar es la diferencia proporcional al valor absoluto previo para cada elemento. La distancia o norma elegida para medir la distancia entre \mathbf{q}_0 y \mathbf{q} es una decisión de negocios, contemplando una distribución de mayor varianza para pocos clientes (norma \mathbf{L}^∞) o menos varianza para más (norma \mathbf{L}^2). Teniendo en cuenta ambos objetivos, el problema queda formulado entonces como:

$$\begin{aligned} \underset{\mathbf{q}}{\operatorname{argmin}} & |(\mathbf{q} - \mathbf{q}_0) \oslash \mathbf{q}_0|^2, \quad |(\mathbf{q} - \mathbf{q}_0) \oslash \mathbf{q}_0|^\infty \\ \text{subject to : } & \mathbf{R}\mathbf{q} = \mathbf{b}, \\ & \mathbf{M}\mathbf{q} \leq \mathbf{d} \end{aligned}$$

Para resolver esto se calcula una primera solución mediante un algoritmo iterativo ad hoc (NRAS) que muestra una buena optimización de la norma \mathbf{L}^∞ . Luego se usa esta solución con el algoritmo ad hoc IterProj para distribuir las pocas unidades restantes minimizando la norma \mathbf{L}^2 . Inicialmente se asigna a las

desigualdades incumplidas una igualdad al valor máximo permitido juntando los dos sistemas de ecuaciones en uno. Luego:

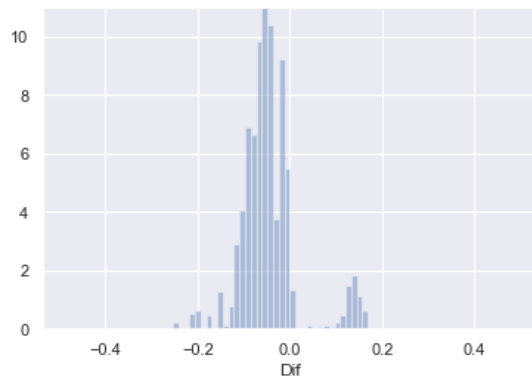
NRAS:

- Calcular el error proporcional para cada grupo para cada restricción, es decir $\mathbf{e} = \mathbf{b}_i / \mathbf{R}_i \mathbf{q}$
- Multiplicar cada fila \mathbf{R}_i por el elemento \mathbf{e}_i , $\mathbf{R}'_i: \mathbf{R}_{ij} := \mathbf{e}_i$
- Calcular promedio \mathbf{r} por columna de \mathbf{R}' , de dimensión igual a \mathbf{q}
- Multiplicar elemento a elemento a \mathbf{q} por \mathbf{r} llevándolo hacia la satisfacción de las restricciones, $\mathbf{q} := \mathbf{q} \circ \mathbf{r}$
- Se evalúa la mejoría repitiendo hasta un mínimo o convergencia

IterProj:

- Calcular solución particular mediante mínimos cuadrados para el sistema $\mathbf{Rq} = \mathbf{b}$
- Computar una base ortonormal del espacio de soluciones utilizando, en vez del producto escalar, un producto interno definido por la forma bilineal $\mathbf{q}^T \mathbf{W} \mathbf{q}_0$ donde \mathbf{W} es diagonal y $\text{diag}(\mathbf{W})_i = \mathbf{q}_{0i}^{-1}$
- Trasladar ese espacio (una variedad lineal) a un subespacio restando la solución encontrada por cuadrados mínimos
- Calcular la proyección ortogonal del vector \mathbf{q}_0 sobre este subespacio
- Trasladar la solución encontrada a la variedad lineal original
- Modificar la matriz diagonal \mathbf{W} aumentando el peso para los valores que resultaron alejados

Ejemplo I



Resultados NRAS:

Norma $q-q_0$ 106.37

Max 0.16

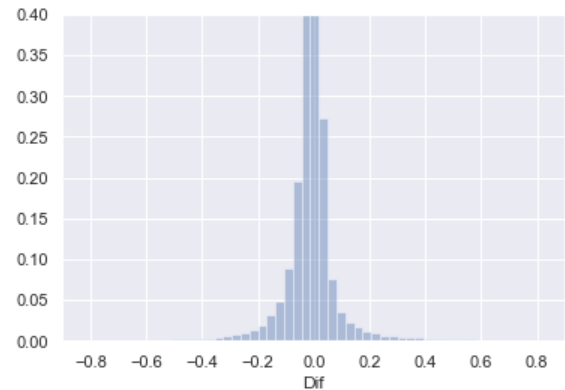
Min -0.43

Media -0.046

Mayores a 1.00 0

Mayores a 0.30 360

Norm Aq-b 85.91



Resultados IterProj:

Norma $q-q_0$ 34.87

Max 0.79

Min -0.76

Media -0.0016

Mayores a 1.00 0

Mayores a 0.30 3110

Norm Aq-b 3.06

El método NRAS resulta en menores extremos que IterProj, minimizando más la norma L^∞ . Posiblemente esto sea preferible para reducir los cambios drásticos en los clientes. Con IterProj se consigue menos diferencia total, focalizando los cambios en un subconjunto reducido.

NRAS + IterProj

Norma $q-q_0$ 106.37

Max 0.16

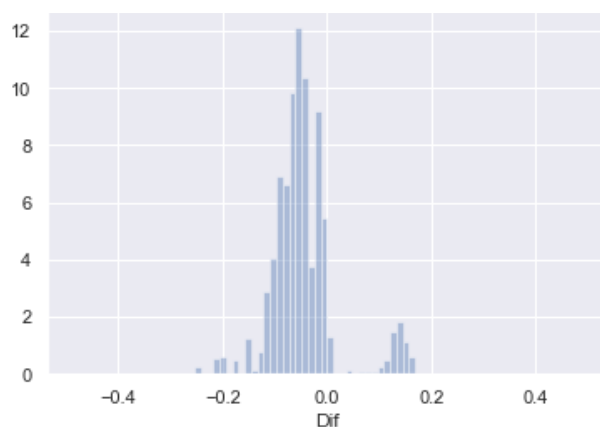
Min -0.43

Media -0.046

Mayores a 1.00 0

Mayores a 0.30 360

Norm Aq-b 3.069



La combinación de ambos algoritmos permite aproximar con mayor exactitud las restricciones (indicado por **Aq-b**), al mismo tiempo que se logran menores valores extremos. Vemos que en el caso de IterProj, la distribución conseguida aproxima a una gaussiana, mientras que NRAS muestra una distribución asimétrica y bimodal, representando un análogo a los métodos paramétricos y no paramétricos.

Ciente	Suc.	Vta. Neta	Vta. Neta_w	Vta. Neta_1	Vta. %	Canal	Fila.	Vta. Neta	Vta. Neta_w	Vta. Neta_1	Vta. %
390535	3	80	68,543	85,809	25,19	1	5	51881,25	47094,516	40077,862	-14,9
399396	1	22,5	19,278	24,134	25,19	1	0	344683,345	373986,248	319558,098	-14,55
477709	1	10	8,568	10,726	25,19	1	14	1041,375	887,088	759,329	-14,4
494486	1	75	63,407	79,378	25,19	1	7	151700,01	145608,119	125117,425	-14,07
484011	1	84	68,562	85,828	25,18	1	18	100183,5	90621,339	77965,816	-13,97
494434	1	122,5	84,77	105,575	24,54	1	15	325613,1	280309,213	241349,58	-13,9
456008	1	405,78	350,984	436,517	24,37	1	2	208027	170210,451	146750,06	-13,78
250893	1	92	78,305	97,336	24,3	1	16	129666,705	137246,682	118869,218	-13,39
475480	1	32,1	27,371	34,014	24,27	1	13	31364,335	30239,216	26602,841	-12,03
244325	2	86,75	73,434	91,083	24,03	1	17	97912,22	92330,105	81469,74	-11,76
416368	1	96,2	81,668	101,287	24,02	1	3	4638420,15	3706672,775	3273721,183	-11,68
483748	1	120,01	102,971	127,515	23,84	1	19	124971,965	101979,91	90181,448	-11,57
494141	1	66	56,663	70,068	23,66	1	9	15626,975	10204,268	9028,303	-11,52
437924	2	268,2	225,356	278,352	23,52	1	1	532144,075	578100,772	513944,251	-11,1
461223	1	730,28	617,759	762,786	23,48	1	4	15204850	12246993,47	10980745,22	-10,34
373822	1	327,53	284,366	350,9	23,4	1	12	552100,425	617672,656	555865,593	-10,01
492254	1	34	29,209	36,012	23,29	99	19	499618,49	399900,809	411412,526	2,88
464804	1	10	8,438	10,39	23,13	99	17	360316,725	334340,74	344628,908	3,08
162383	1	294,2	249,97	307,629	23,07	0	18	7715,75	7831,062	8075,258	3,12
350836	2	73,5	61,965	76,197	22,97	0	5	11447,1	9458,542	9763,679	3,23
493582	1	67,5	58,179	71,525	22,94	99	5	203275,55	176550,472	183258,426	3,8
368737	1	79	65,303	80,24	22,87	0	19	9243,21	7335,681	7618,002	3,85
493550	1	65,95	55,876	68,654	22,87	0	12	60543,84	58623,579	60906,194	3,89
447670	1	379,82	322,478	396,167	22,85	0	15	28559,7	25767,953	26821,779	4,09
345590	1	156	131,355	161,361	22,84	99	12	1431394,82	1319215,226	1378735,455	4,51
450183	1	46,77	40,063	49,157	22,7	99	7	473265,87	433230,143	453076,495	4,58
481847	1	124	105,219	129,033	22,63	99	18	278912,75	267216,988	279625,939	4,64
316431	1	176,705	133,624	163,782	22,57	99	9	29201,37	24279,499	25416,09	4,68
236083	1	97,35	82,209	100,692	22,48	99	0	1029845,54	1103360,074	1155341,198	4,71
494367	1	16,8	14,198	17,39	22,48	0	7	14804,98	13538,206	14179,174	4,73

Además, IterProj devuelve una norma mucho menor, por lo que se utiliza como *pulido* luego del algoritmo anterior. Esta opción se puede modificar para intercambiar mayor velocidad por precisión desde la pestaña de VBA de excel, mediante el parámetro *polish* de la función solve(). Por defecto, IterProj no se utiliza, otorgando mayor velocidad pero levemente menor precisión.

Restricciones del ejemplo 1:

Presupuesto completo

Canal 1 <= 33% total

Canal 0 <= 98% anterior

Ejemplo II

Restricciones:

Canal 1 <= 36% Total

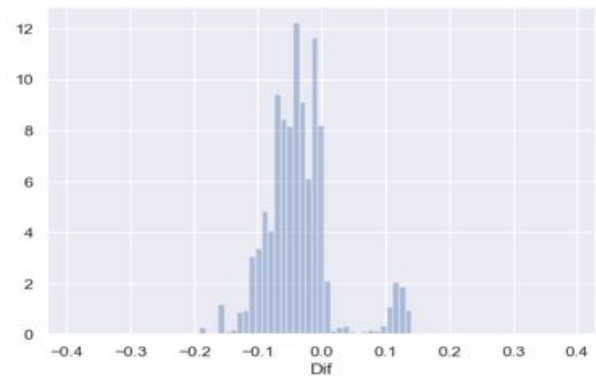
Canal 0 <= 95% Relativo

Grupo Económico 1 = 120%

Zona 120 = 105%

Presupuesto completo

	Norma q-q0	87,23	
	Max	0,15	
	Min	-0,32	
	Media	-0,036	
	Mayores a 0.70	0	
	Mayores a 0.30	5	
	Norma Aq-b	8,58	



Referencias

Operations R.: <https://notendur.hi.is/kth93/3.20.pdf>

MINLP: <https://www.mcs.anl.gov/papers/P3060-1112.pdf>

Convex Optim.: https://web.stanford.edu/~boyd/cvxbook/bv_cvxbook.pdf

Solvers: https://en.wikipedia.org/wiki/List_of_optimization_software