

## **BIBLIOTECOMETRÍA (Librametrics). EVALUACIÓN DE COLECCIONES, UN ESTUDIO DE CASO**

*Prof. César O. Archuby (Coord.)  
UNLP/FAHCE  
coa@isis.unlp.edu.ar*

**Temas BCI:** Relación entre la edad del material bibliográfico y la circulación, Ley del 80/20 (Trueswell)

**Temas matemáticos:** Variables agrupadas, marcas de clases, frecuencias absolutas, frecuencias acumuladas, medidas de centralización, distribución estadística de las edades (lognormal), logaritmos, operaciones matriciales (producto escalar de vectores), automatización del cálculo.

### **Parte 1:**

- **Estructura de datos**
- **Reconocimiento de la matriz de datos**
- **Interpretación de los datos**
- **Estadística descriptiva**
- **Cálculos preliminares**

### **Parte 2:**

- **Automatización del cálculo**
- **Reducción de la probabilidad de error y del esfuerzo**
- **Estadística inferencial mono y bivariada**
- **Análisis de resultados**
- **Ley del 80/20**

**Consigna:**

Usted tiene el ejercicio resuelto en el Libro Resolución Parte 1.

Ver **ADVERTENCIA** al final de la página 19.

Una vez que haya estudiado y comprendido la resolución le aconsejo repetir todos los cálculos sin ayuda del texto a partir de la planilla Datos originales, en lo posible copie en un archivo nuevo sólo los datos y resuelva el mismo.

Como el objetivo de esta parte es la familiarización con Excel, con la forma de representación de los datos y con los principios del uso de la automatización, es de la mayor importancia que realice el trabajo sin ninguna ayuda, y que reporte a los docentes los puntos de mayor dificultad, tanto originados en los temas tratados como en la interpretación del texto.

## **Antecedentes, Enfoques y Alcance**

### ➤ ***Antecedentes***

La evaluación en general es una actividad de gran desarrollo y extensa aplicación en todo el mundo. Los gobiernos, fundaciones y todo tipo de institución de promoción de la ciencia, el arte, la educación, la salud, el comercio, la industria, el turismo, etc. formulan programas de promoción de actividades, invierten dinero en ellos y luego evalúan los resultados con el fin de obtener información que les permita fundamentar nuevos rumbos, tanto como justificaciones y reducciones o ampliaciones de los programas en curso.

Aunque se atribuye a Ranganathan el acuñamiento del término Librametrics en 1948 y tenemos antecedentes de estudios de características de las colecciones (Stieg en el 43 y Gosnell en el 44), la evaluación de bibliotecas en particular tomó impulso a fines de los '60 con la aparición de las computadoras (Fussler y Simon a partir de 1961). Si bien en los primeros tiempos la afectación de las computadoras a la gestión de bibliotecas fue mínima, por su altísimo costo entre otras cosas, los investigadores tuvieron un importante incentivo para intensificar el desarrollo de técnicas y métodos de evaluación de bibliotecas y de sistemas y servicios de información, en los que tuvo relevante participación Frederick W. Lancaster, cuya obra "Evaluación de la biblioteca" es de consulta obligada.

En nuestro medio el tema estuvo ausente hasta hace unos pocos años cuando fue activado, por necesidades originadas en requerimientos de la Comisión Nacional de Evaluación Universitaria. CoNEAU.

Actualmente la actividad parece definitiva aunque incipientemente instalada en el ámbito universitario por la presencia de proyectos de investigación nacionales e internacionales y actividades de autoevaluación en las bibliotecas, entre los que merecen especial atención por su vinculación con la formación de recursos humanos en el área, los trabajos finales y tesinas de licenciatura con temática "evaluación de colecciones".

### ➤ *Enfoques*

Vamos a distinguir entre dos enfoques de la evaluación de bibliotecas:

- ☀ Uno que evalúa a la biblioteca como parte monolítica de la organización a la que presta servicios
- ☀ Otro se ocupa de lo que ocurre en el interior de la biblioteca.

El primer enfoque mide ingresos y egresos, recursos y servicios, ingresos y producción, en general costos y beneficios para la organización, sin analizar el interior de la biblioteca que recibe el trato de "Caja Negra", motivo por el cual este enfoque, denominado sistémico, puede aplicarse a cualquier sistema o servicio no bibliotecario.

Este enfoque se utilizó en el proyecto TIEB de cooperación entre la Universidad Nacional de La Plata (UNLP) y la Universidad de Granada, en el que participó el profesor titular de este seminario. Aporta información global para la toma de decisiones superiores, a un nivel de gestión que está entre la dirección de la biblioteca y la conducción de la organización a la que presta servicios o a niveles superiores como organizaciones nacionales o internacionales interesadas en el tema.

El otro enfoque, al que denominaremos especializado, centra su análisis en el interior de la biblioteca, en sus recursos, como la colección; en sus servicios, como el préstamo; en la distribución temporal de su personal, etc.

Este enfoque se utiliza en un proyecto de investigación sobre evaluación de colecciones, en la UNLP. Aporta información técnica para la toma de decisiones directivas en el nivel que está entre los responsables de los servicios y la dirección de la biblioteca.

### ➤ *Alcance*

Consideraciones de factibilidad relacionadas con la duración de las clases nos obligan a restringir el alcance de un tema que, por su amplitud, requeriría más de un seminario completo. Nos limitaremos a una parte de la evaluación cuantitativa de colecciones como es la **relación entre la edad de la colección y la circulación**, durante un año, para lo que nos apoyaremos en el estudio de una clase temática de la colección de una biblioteca universitaria real.

### ➤ *Fuente*

Desarrollaremos este tema alrededor de los datos de circulación de la clase 3 de una biblioteca universitaria real. Los datos, con pequeñas modificaciones por razones didácticas, forman parte del trabajo final del

Seminario C, Elementos de bibliometría y aplicaciones, de la Licenciatura a Distancia (LICAD) de la Universidad Nacional de Mar del Plata, presentado por la Bibliotecaria Profesional María Paula Salmoiraghi.

Para no agregar dificultades de origen bibliotecológico y por razones exclusivamente didácticas, haremos caso omiso de las diferencias entre títulos y ejemplares.

Comenzaremos con la matriz original, antes de las primeras correcciones, con el objeto de observar algunos errores y sesgos típicos en los trabajos de estudiantes de bibliotecología, sus consecuencias y sus correcciones. Ver en Excel, Libro Datos Originales.

## RECONOCIMIENTO DE LA MATRIZ DE DATOS

### ➤ **INTERPRETACIÓN**

**FUENTE.** Los datos pueden tener origen en una base de datos de préstamos o en otro medio de registro de la circulación del material. El nivel de automatización del procedimiento que determina cada valor de la matriz depende de las prestaciones del software disponible. En este caso el proceso de construcción de la matriz fue manual.

### ➤ **SIGNIFICADO DE LOS VALORES EN LAS CELDAS**

**Las celdas de la columna A de la Hoja Sólo Datos, del Libro Datos originales,** desde la fila 5 hasta la 34, que lleva como encabezado **Edad (A4)**, contienen **encabezamientos de filas** correspondientes a intervalos de clase de 5 años cada uno.

#### **Por ejemplo:**

En A5, vemos 1-5 que significa que la edad del material de toda la fila 5 tiene entre 1 y cinco años. (No admite material del año actual o edad cero, lo que puede considerarse un error)

En A15, vemos 51-55 que significa que la edad del material de toda la fila 15 tiene entre 51 y 55 años.

En A34, vemos 146-150 que significa que la edad del material de toda la fila 34 tiene entre 146 y 150 años.

**Las celdas de la fila 4,** desde la columna B hasta la Z, contienen **encabezamientos de columnas** cuyas celdas tienen valores correspondientes a la cantidad de préstamos correspondientes a las cantidades de documentos registrados en la columna.

#### **Por ejemplo:**

La celda B4 contiene el valor 0.P, encabezamiento que significa 0 Préstamos. Todas las cantidades de material bibliográfico de la columna B nunca fueron prestadas en el año

La celda M4 contiene el valor 21.P que significa 21 Préstamos. Todas las cantidades de material bibliográfico de la columna M, fueron prestadas 21 veces.

La celda Z4 contiene el valor 83.P que significa 83 Préstamos. Todas las cantidades de material bibliográfico de la columna Z, fueron prestadas 83 veces.

Las celdas de las filas 5 hasta la 34, desde la columna B hasta la Z, contienen cantidades de documentos cuya edad y cantidad de préstamos pueden verse en los encabezamientos correspondientes de fila y columna, respectivamente.

### EJEMPLOS

El valor 10 de la celda B5 significa que hubo 10 documentos con antigüedad de 1 a 5 años, que tuvieron una cantidad de préstamos igual a 0.

El valor 19 de L7 significa que hubo 19 documentos con antigüedad de 11 a 15 años, que fueron prestados 10 veces cada uno, en total 190 préstamos.

El valor 4 de X10 significa que hubo 4 documentos con antigüedad de 26 a 30 años, que tuvieron una cantidad de préstamos igual a 55 cada uno, en total 220 préstamos.

El valor 2 de D31 significa que hubo 2 documentos con antigüedad de 131 a 135 años, que tuvieron una cantidad de préstamos igual a 2 cada uno, en total 4 préstamos.

El valor 0 de T15 significa que hubo 0 documentos con antigüedad de 51 a 55 años, que tuvieron una cantidad de préstamos igual a 40 cada uno, en total 0 préstamos.

### DIMENSIONES DE LA MATRIZ

La matriz de datos tiene **25 columnas**, de la B a la Z, y **30 filas**, de la 5 a la 34, lo que nos da un total de  $30 \times 25 = 750$  valores, la mayoría de los cuales son ceros. Más adelante volveremos sobre esta cantidad.

#### ➤ ***DATOS AGRUPADOS, SUCESOS, FRECUENCIAS, TERMINOLOGÍA***

Los valores de la *Variable Aleatoria* (VA) Edad, están agrupados en *intervalos de clase*, por ejemplo 1-5 años.

Cada una de las 750 celdas de la matriz contiene la *frecuencia absoluta* de un *suceso* ó evento que relaciona la edad del material bibliográfico con su circulación.

**Ejemplo:** El valor 19 de la celda L7 es la frecuencia absoluta del suceso "Documentos con edad entre 11 y 15 años que circularon 10 veces cada uno"

**No sabemos** cuántos de esos 19 documentos tienen edad 11, 12, 13, 14 ó 15 años, motivo por el cual no podremos calcular con *exactitud* las *medidas de centralización y dispersión*; sólo podremos hacerlo en forma *aproximada* tomando a la edad 13 como representativa del intervalo y haciendo los cálculos como si todos los documentos tuvieran esa edad.

Este valor central, representativo del intervalo de clase, se denomina **marca de clase** y su valor es el promedio de los extremos del intervalo:  $(11 + 15)/2$

## PREGUNTAS

### 1.- Por qué, entonces, se trabaja con *datos agrupados*?

En el **cálculo de medidas** se trabaja con datos agrupados EXCLUSIVAMENTE cuando no es posible hacerlo con datos puntuales. En este caso, por ejemplo, hemos recibido estos datos y no tenemos forma de saber cuántos documentos tienen 12 años, ni 4, ni 9, etc.

Si se puede, entonces, se debe **calcular** con datos puntuales, porque darán valores exactos.

### 2.- En qué situación o con qué objetivo puede ser conveniente el agrupamiento cuando se dispone de datos puntuales?

Sólo para facilitar la visualización de datos o resultados de los estudios. Por ejemplo, en nuestro caso, 150 edades pueden ser muchos valores para observar, sobre todo si vemos que la matriz sería de 150x25, en lugar de 30x25.

### 3.- Cómo determino el nivel de agrupamiento óptimo para mejorar la visualización?

Probando, hasta que se vea lo que queremos mostrar. Tener en cuenta que si con todos los datos puntuales se ve lo que queremos mostrar, no agrupamos.

## HIPÓTESIS Y OBSERVACIONES PRELIMINARES

### Recordemos:

Todos los valores de la matriz son cantidades de documentos que cumplen una condición de edad y otra de circulación. Ej. El valor 11 de K10 significa que hubo 11 documentos con edad entre 26 y 30 años, que circularon 9 veces cada uno

### ***Hipótesis***

Una hipótesis obvia de nuestra investigación es que la intensidad de circulación del material bibliográfico es inversamente proporcional a su edad, o sea que a menor edad mayor circulación y viceversa.

Otra hipótesis obvia es que cuanto mayor sea la cantidad de préstamos por documento, menor será la cantidad de documentos que tienen esa cantidad de préstamos, o sea que serán *pocos* los documentos con *mucha* cantidad de préstamos y viceversa.

### Observaciones:

La disposición de los datos en la matriz ya nos permite *ver* la confirmación de las hipótesis; observemos los cuadrantes marcados en la Hoja Datos Originales; al hacerlo tomemos en cuenta que cuando decimos "edades bajas", por ejemplo, nos referimos a la división hecha con los cuadrantes y no sugerimos que 70 años es una edad baja para un documento; del mismo modo la división en circulación "alta o baja" responde sólo a necesidades de estrategia didáctica.

El **cuadrante superior izquierdo** corresponde a *edades bajas y circulación baja*. Tiene *pocos ceros* lo que indica *mucha presencia de documentos prestados pocas veces*. La mayoría de los ceros están en su parte inferior derecha, indicando que el suceso "mayor edad – mayor circulación", dentro de este cuadrante, se produjo cero veces.

El **cuadrante inferior derecho** corresponde a *edades altas y circulación alta*. Todas celdas tienen valor cero lo que indica que no hubo *alta circulación* de material *antiguo*. La frecuencia del suceso "edad alta–circulación alta", siempre fue cero.

El **cuadrante superior derecho** corresponde a *edades bajas y circulación alta* y presenta *importante cantidad de ceros*, con mayor concentración en su parte inferior derecha (edad alta – circulación alta = frecuencia cero, en este cuadrante)

El **cuadrante inferior izquierdo** corresponde a *edades altas y circulación baja*. Tiene muchos ceros. Vemos que sólo hay valores distintos de cero en las 3 primeras columnas que corresponden a las más bajas cantidades de préstamos: "*alta edad – baja circulación*"

### Observaciones y precisión

En los párrafos anteriores utilizamos el lenguaje natural con terminología imprecisa, abierta a interpretaciones diversas como es habitual en ambientes no científicos, para expresar las observaciones referidas a la relación entre la edad del material bibliográfico y su circulación.

Cuando decimos terminología imprecisa nos referimos a expresiones como "edades altas", "edades bajas", "alta circulación", "baja circulación", "mucho presencia", "mayoría de ceros", en las que queda sujeto a interpretación hasta donde llegan las edades bajas y las altas, tanto como la circulación alta y baja, las mayorías y la mucha presencia.

### Precisemos los términos y los límites de los cuadrantes

#### ● Dividiremos las 30 edades en dos grupos de 15 intervalos de clase cada uno:

El primero de ellos incluirá las edades entre 1 y 75 años, celdas A5 a A19 y el segundo incluirá las edades entre 76 y 150 años, celdas A20 a A34.

El primer grupo corresponde a "edades bajas" y el segundo a "edades altas", ahora con definiciones precisas que no dejan lugar a interpretaciones diversas.

#### ● Dividiremos las 25 cantidades de préstamos, en dos grupos, uno de 13 cantidades y el otro de 12.

El primero de ellos incluirá cantidades de préstamos entre 0 y 22, celdas B4 a N4, y el segundo incluirá cantidades de préstamos entre 25 y 83, celdas O4 a Z4.

El primer grupo corresponde a "baja circulación" y el segundo a "alta circulación", también con definiciones precisas no sujetas a interpretación.

Ahora tenemos límites precisos que nos permiten definir con precisión los cuatro cuadrantes.

**Cuadrante superior izquierdo:** Celdas B5 a N19 (fondo canela)

*(Edades bajas, circulación baja, pocos ceros, mucha presencia de documentos prestados)*

A partir de esta definición podremos decir “Cuadrante superior izquierdo” sin caer en imprecisiones ya que la frase es equivalente a “Celdas B5 a N19”

**Cuadrante superior derecho:** Celdas O5 a Z19 (Fondo verde)

*(Edades bajas, circulación alta, importante cantidad de ceros)*

**Cuadrante inferior izquierdo:** Celdas B20 a N34 (Fondo amarillo)

*(Edades altas, circulación baja, muchos ceros)*

**Cuadrante inferior derecho:** Celdas O20 a Z34 (Fondo violeta)

*(Edades altas, circulación alta, todos ceros)*

Precisados los límites de los cuadrantes, vamos a precisar las observaciones en cursiva.

## CAMBIO DE EXPRESIONES DE LENGUAJE NATURAL POR EXPRESIONES QUE INCLUYEN LENGUAJE MATEMÁTICO:

- En lugar de **circulación baja**, diremos:  
**cantidades de préstamos de 0 a 22**
- En lugar de **pocos o muchos ceros**, diremos:  
**el X% de los valores del cuadrante son ceros**
- En lugar de **mucha o poca presencia de documentos prestados**, diremos:  
**el Y% de los documentos de la colección fueron prestados alguna vez.**

Como vemos, para las dos últimas expresiones deberemos *contar* y *calcular*, ya que para calcular el X% es necesario conocer la cantidad total de valores del cuadrante y la cantidad de ceros del mismo. Para contar la cantidad de ceros usaremos la función contar.si de la siguiente manera:

### Cuadrante superior izquierdo

En la celda **A36** ingresamos un cero.

En la celda **A37** ingresamos  $=\text{CONTAR.SI}(B5:N19;A36)$ , ó  $=\text{CONTAR.SI}(B5:N19|A36)$ . Esta expresión invoca a la función contar.si y le pide que cuente, en el rango B5:N19, la cantidad de veces que aparece el valor de la celda **A36**. Estamos pidiendo que cuente la cantidad de ceros en el cuadrante superior izquierdo.

El resultado es el valor de la celda **A37**: 69 ceros.

La cantidad de celdas del cuadrante es  $13 \times 15 = 195$ , celda A38.

El % de ceros es  $= 69 / 195 * 100 = 35,4\%$

En la celda A39 ingresamos:  $=\text{A37}/\text{A38} * 100$

***SI REPETIMOS EL CONTEO Y EL CÁLCULO PARA LOS OTROS TRES CUADRANTES OBTENEMOS:***

En la celda **B37**:  $=\text{CONTAR.SI}(O5:Z19|A36) = 134$  **Superior derecho**

La cantidad de celdas del cuadrante es  $12 \times 15 = 180$ , celda B38

El % de ceros es  $= 134 / 180 * 100 = 74,4\%$  ( $B39=B37/B38 * 100$ )

En la celda **C37**:  $=\text{CONTAR.SI}(B20:N34|A36) = 166$  **Inferior izquierdo**

La cantidad de celdas del cuadrante es  $13 \times 15 = 195$ , celda **C38**

El % de ceros es  $= 166 / 195 * 100 = 85,1\%$  ( $C39=C37/C38 * 100$ )

En la celda **D37**:  $=\text{CONTAR.SI}(O20:Z34|A36) = 180$  **Inferior derecho**

La cantidad de celdas del cuadrante es  $12 \times 15 = 180$ , celda **D38**

El % de ceros es  $= 180 / 180 * 100 = 100\%$  ( $D39=D37/D38 * 100$ )

Si esta división en cuadrantes fuera un modelo representativo de la relación entre la edad y la circulación, concluiríamos que se confirma la hipótesis obvia: ***así como crecen la edad y la circulación, decrece la cantidad de documentos con tal edad y circulación.***

## OTROS

En la celda E37 vemos el valor 549 que es la cantidad de ceros de la matriz obtenida por la suma de las cantidades de ceros de los cuadrantes. En la celda F37 vemos el mismo valor pero calculado sobre la matriz completa.

En la celda E39 vemos que el porcentaje de ceros de la matriz es 73. Comparando con los porcentajes de los cuadrantes vemos que el superior izquierdo está significativamente por debajo de la cifra general (menos de la mitad), mientras que los otros tres la superan.

## COMENTARIOS

*Independientemente de la utilidad que pueda tener este indicador y de sus limitaciones (sólo distingue entre "0" y "no 0", sin tomar en cuenta si el "no 0" es 1 ó 50), hemos mostrado una forma de extraer información de los datos, una forma de "leer" los datos, con ayuda de la computadora, además de familiarizarnos con la estructura de los datos, el significado de sus valores y las operaciones de contar.*

*Este tipo de actividades, que suele recibir el nombre de "Análisis exploratorio de datos", además de aportar alguna información sobre el problema bajo análisis nos ayuda a orientar el estudio y a elegir técnicas y herramientas.*

*Podrá observarse que la estructura de datos en forma de matriz facilita el trabajo.*

## REFUTACIÓN AL ESTUDIO PREVIO

Para que la cantidad de ceros fuera un indicador aceptable, deberíamos completar la matriz con varias columnas que están ausentes, como las columnas correspondientes a 11, 12, 13,....20 préstamos, y tantas otras que no fueron incluidas porque todas sus celdas tendrían ceros y casi ninguna información.

*Qué utilidad tiene para la evaluación de la colección saber que ningún documento fue prestado 11 ó 12 veces, cuando varios fueron prestados 10 ó 21 veces, por ejemplo? Parece que poca o ninguna, evidentemente es una cuestión de azar que ningún documento haya sido prestado entre 11 y 20 veces y que varios hayan tenido 21 préstamos.*

## COMIENZO DE LA EVALUACIÓN

El eje principal sobre el que giran las actividades de un proceso de evaluación de una colección es el conocimiento de características propias de la misma y de su relación con el medio en el que actúa.

El conocimiento sobre cualquier tema se obtiene mediante actividades diversas agrupadas bajo el nombre de investigación, actividad que comienza con la formulación de preguntas, continúa con las correspondientes respuestas y finaliza con nuevas preguntas surgidas durante el estudio.

Comenzaremos con la formulación de preguntas.

### PREGUNTAS Y CÁLCULOS INICIALES

#### *Es muy grande la proporción de la colección que no salió del estante?*

Para responder esta pregunta necesitamos dos valores: La cantidad de documentos que no salieron del estante y la cantidad de documentos de la colección.

#### **Paso 1**

Cálculo de la cantidad de documentos que no salieron del estante.

La cantidad de documentos que nunca se prestaron (B4) es igual a la suma de la columna B, de acuerdo a su encabezamiento. Como vemos en B35 =SUMA(B5:B34), hay **606** documentos con cero préstamos.

#### **Paso 2**

Cálculo de la cantidad de documentos de la colección

Este cálculo puede hacerse con más de un procedimiento:

1. Podemos calcular la cantidad de documentos de la colección sumando todos los valores de las celdas de la matriz, ya que hemos visto que cada valor de la matriz corresponde a una cantidad de documentos de la colección. Por lo que hemos visto, en la matriz están todos los documentos y ninguno está repetido, por lo que si sumamos los valores de todas las celdas de la matriz tendremos la cantidad de documentos de la colección en estudio.
2. Podemos calcular la cantidad de documentos que fueron prestados tantas veces como indica el encabezamiento de columna que registra la fila 4, desde B4 (1.p) hasta Z4( 83.p), y luego sumar los 24 valores, que dará el mismo total que el procedimiento anterior.

Por ejemplo, para calcular la cantidad de documentos que fueron prestados una sola vez, ingresamos =SUMA(C5:C34), en la celda C35, donde vemos que son 137 documentos.

De la misma forma, la suma de las demás columnas dará la cantidad de documentos correspondiente a la cantidad de préstamos que muestra el encabezamiento de cada columna (Fila 4).

### **Veamos dos formas de hacer el trabajo:**

#### **■ Proceso manual asistido por computadora**

La forma “manual asistida por computadora” consistirá en repetir el cálculo de la columna B para cada columna, desde la C hasta la Z, por lo que en cada celda habrá que ingresar una expresión como = suma (C5:C34), en total 25 expresiones, con el consiguiente trabajo y posibilidad de cometer algún error en cada expresión.

Si bien las dos desventajas del cálculo manual asistido por computadora (+ trabajo y + errores) justificarían por sí mismas el intento de automatizar el proceso, es importante poner de manifiesto que la principal ventaja de la automatización es otra.

En el trabajo manual asistido por computadora la *cantidad de trabajo realizado por la computadora es directamente proporcional al trabajo realizado por nosotros*, por ejemplo, en el caso anterior la computadora sumó 25 columnas (claro que sumó 30 valores por cada columna en forma automática) pero nosotros debimos escribir 25 fórmulas de cálculo.

Aunque es obvio, no debemos perder de vista el hecho de que la automatización subrayada en el párrafo anterior fue posible porque, además de la estructura, el tipo de datos(números) permite sumar. Si para mejorar la comprensión de la matriz, en lugar de registrar en B5 el valor 10 hubiéramos registrado 10D (10 Documentos), se vería con mayor claridad que 10 documentos fueron prestados 0 veces, pero no podríamos sumar la columna. Esto es lo que ocurre con la fila 4.

Si en lugar de sumar 25 columnas tuviéramos que hacerlo con 100, nuestro trabajo sería muy repetitivo, pesado, tedioso, en fin, contrario al “Principio del menor esfuerzo” que, según G.K. Zipf, guía la conducta humana y, además, la probabilidad de no cometer algún error sería casi nula. Para cantidades mayores el calificativo pasa de “muy difícil” a “prácticamente imposible”.

***En el proceso automático no hay relación proporcional entre el trabajo realizado por la computadora y el realizado por nosotros.***

#### **Ejemplo:**

La realización automática de la suma de los 30 valores de la columna B, nos exigió el ingreso de la fórmula “= suma (B5:B34)”. Si en lugar de 30 valores tuviéramos que sumar 3000, el trabajo de la computadora sería 100

veces mayor, pero el nuestro sería el mismo: ingresar la fórmula “= suma (B5:B3004)”

### ● **Proceso automático**

Consistirá en el “arrastre” de la celda B35 hasta la Z35.

Si en lugar de 25 columnas fueran 250, la computadora haría 10 veces más trabajo, mientras que nuestra tarea se reduciría a prolongar el “arrastre”. No olvidemos que en cada una de las 250 columnas la computadora haría 30 sumas, o sea que hará 7500 sumas mientras nuestro trabajo se reducirá a realizar el “arrastre”.

### **La calidad de los datos**

Esto último, que fue posible por contar con el tipo y la estructura de datos adecuados, debe alertarnos sobre la importancia de contar con datos correctos para cualquier tipo de tratamiento automático. Si bien el dato correcto también es importante en el tratamiento manual, éste nos permite advertir los errores de los datos en el momento del uso, mientras que en el tratamiento automático es imposible.

#### **Paso 3**

Cálculo de la cantidad total de documentos

En la fila 35, desde la columna B hasta la Z, tenemos la cantidad de documentos que fueron prestados desde 0 veces hasta 83 veces por lo que ningún documento de la colección en estudio ha quedado fuera de la lista. Sólo faltaría sumar los valores desde B35 hasta Z35 para obtener el total de documentos.

En AA35 vemos el valor 1294 correspondiente a la fórmula =suma (B35:Z35)

### **OBSERVACIÓN**

Si sólo nos interesara la cantidad total de documentos, bastaría con ingresar =suma(C5:AA34), claro que no tendríamos la discriminación por cantidad de préstamos de la fila 35.

#### **Paso 4**

Cálculo del % correspondiente a los documentos que nunca fueron prestados

% de nunca prestados =

Cantidad de “nunca prestados” / Cantidad total de documentos x 100

=606 / 1294 x 100 = 46,8% (Ver B36)

### RECUERDAN LA PREGUNTA INICIAL?

■ *Es muy grande la proporción de la colección que no salió del estante?*

**Finalmente tenemos la respuesta:**

*El 47% de los documentos no salió del estante durante el año del estudio*

**Finalmente tenemos la respuesta a la pregunta**

### ANEXO

#### 1. OBSERVACIONES FINALES SOBRE LA AUTOMATIZACIÓN DEL CÁLCULO

La automatización de los procesos de cálculo depende de dos condiciones referidas a los datos:

1. La **estructura** de los datos
2. El **tipo** de los datos

La automatización mediante fórmulas y "arrastrés", como acabamos de hacer, sólo es posible si los datos tienen una **estructura** que lo permite, como las filas o las columnas.

Los cálculos sólo son posibles si los datos son de **tipo** numérico, con o sin automatización.

**Veamos los siguientes problemas:**

#### Problema 1

**Calcular la cantidad de préstamos correspondiente a material bibliográfico con edad entre 1 y 5 años que haya sido prestado 9 veces**

El valor 6 de K5 (9.P) nos indica claramente que hubo 6 documentos, cuya edad estaba entre 1 y 5 años, que fueron prestados 9 veces cada uno, por lo que el total de préstamos es igual a:  $6 \times 9 = 54$ .

Como se ve, el cálculo manual resulta sumamente fácil, pero vislumbramos problemas para cuando tengamos que calcular el total de préstamos correspondientes a esa edad, como veremos en el Problema 2..

El cálculo manual asistido por computadora mejora algo pero poco ya que, en la planilla Excel correspondería multiplicar la celda K5 por la celda K4, pero el valor de la celda K4 no es numérico, por lo tanto no podemos multiplicarlo ni hacer operaciones aritméticas. Por lo mismo no podremos definir fórmulas aritméticas ni tampoco arrastrarlas.

Sólo nos queda por hacer lo que hizo la creadora de la matriz: hacer =K5 x 9 y almacenar el resultado (54) en alguna celda. Si quisiéramos conocer la cantidad de préstamos correspondiente a todos los documentos cuya edad estaba entre 1 y 5 años, deberíamos repetir la operación para cada cantidad de préstamos, registrar los resultados en celdas "borrador" y luego sumar uno por uno.

### Problema 2

**Calcular la cantidad de préstamos correspondiente a material bibliográfico con edad entre 1 y 5 años que haya sido prestado entre 0 y 83 veces**

Aquí Excel ayuda algo más ya que nos permite la escritura de la suma total, por ejemplo, en una celda libre, que podría ser AB5, ingresamos:

=B5 x 0 + C5 x 1 + D5 x 2 + E5 x 3 + F5 x 4.....+ L5 x 10 + M5 x 1,  
y no seguimos sumando porque las cantidades de documentos restantes son todas iguales a cero. Esta última verificación se hace "a ojo" y, aunque es muy fácil, el tedio del trabajo repetitivo fatalmente resultará en errores, como veremos en Problema 3 más abajo.

### Problema 3

**Calcular la cantidad de préstamos correspondiente a material bibliográfico con edad entre 36 y 40 años, que haya sido prestado entre 0 y 83 veces**

En **AB12** ingresaríamos:

=B12\*0 + C12\*1 + D12\*2 + E12\*3 + F12\*4 + G12\*5 + H12\*6 + I12\*7 + J12\*8, y no seguimos porque *siguen todos ceros*. El resultado es 37.

Si seguimos mirando la fila 12, veremos que el valor de la celda T12 no es cero sino 2 y si observamos el valor de T4, vemos que corresponde a 40 préstamos por lo que el 2 de T12 significa que hubo 2 documentos con 40 préstamos cada uno, o sea 80 préstamos sólo aquí, y acabamos de dar un total de 37 para toda la fila!!

Este error, debido a que ya cansados de tanta repetición no “vimos” el 2, es real, fue cometido por la autora de la matriz, junto a unos increíblemente pocos más, como cambiar inadvertidamente una letra identificadora de columna por otra, o cambiar un signo \* por un +, o viceversa, u omitir un sumando, todo producto de un trabajo manual, tedioso y repetitivo para el que fueron hechas las computadoras, no las personas.

#### **Problema 4**

**Calcular las cantidades de préstamos para toda la matriz: todas las edades y todas las cantidades de documentos**

Aplicaremos el mismo procedimiento que en los dos problemas anteriores, claro que habrá que repetirlo 30 veces, una por fila y, si bien los 549 ceros (E37) y los 30 valores de la columna O.P nos ahorrarán muchas multiplicaciones y sumas, no evitaremos las restantes, con la consiguiente carga de trabajo desagradable y alta probabilidad de cometer errores.

#### **Observación importante**

Evitando todas las multiplicaciones por cero “ganamos” mucho tiempo si trabajamos en forma “manual, asistidos por computadora”. Claro que casi todas las fórmulas serán distintas porque los ceros estarán en ubicaciones distintas lo que impedirá la operación base de la automatización: el “arrastre”, que aplica siempre las mismas fórmulas.

## **2. ADAPTACIÓN DE LOS DATOS PARA FACILITAR LA AUTOMATIZACIÓN DEL CÁLCULO**

(Comenzamos el desarrollo de Resolución Parte 1Final)

### **Paso 1**

En la Hoja Resolución Parte 1, quitar las letras y los puntos de la fila 4.

Seleccionar las celdas de B4 hasta Z4 y darles formato general.  
En la celda B4, en lugar de 0.P, quedará 0.

En la celda K4, en lugar de 8.p, quedará 8, y así desde B4 hasta Z4.

**Paso 2** Insertar una columna entre la A y la B y darle formato general.

Hacer **B5 = (1+5)/2**: aparecerá un 3 que es la Marca de clase del intervalo de clase 1-5

Los extremos inferior y superior de los intervalos de clase están separados por 5 unidades de los correspondientes al intervalo siguiente, por ejemplo:

1 está a 5 unidades de 6 y 6 está a 5 unidades de 11 y así hasta 141 que está a 5 unidades de 146

5 está a 5 unidades de 10 y 10 está a 5 unidades de 15 y así hasta 145 que está a 5 unidades de 150.

De la misma forma, los centros de los intervalos o marcas de clase, estarán separados por 5 unidades, por ejemplo,  
3 está a 5 unidades de 8 y 8 a 5 unidades de 13, y así hasta el final.

Por este motivo definiremos como fórmula general de las marcas de clase, para este caso a la siguiente:

Desde la segunda hasta la última, cada marca de clase es igual a la anterior más 5. Así, hacemos  $B6 = B5 + 5$ , y "arrastramos" hasta el final. Ver columna B.

**ADVERTENCIA.** La inserción de la columna B para las marcas de clase produce el desplazamiento de todas las columnas en Resolución Parte1 Final. Ver esa diferencia de Resolución Parte 1 Final con Resolución Parte1 y Datos Originales.

Con esto damos por finalizada la Parte 1.