

ILUMINAR 03

ANÁLISIS DE DATOS: INTRODUCCIÓN AL TRABAJO CON HOJAS DE CÁLCULO



En esta tercera nota del módulo Iluminar de *Descifrar la Injusticia*, exploramos cómo comenzar a dar significado a los datos que hemos recopilado. Aquí, activistas y agentes del cambio encuentran herramientas para comenzar a procesar números que generarán evidencia para el cambio.

Preguntas clave

¿Cómo pueden ayudarnos las hojas de cálculo a analizar los datos que hemos recopilado?

¿Cuáles son las características básicas de las hojas de cálculo y cómo se utilizan?

¿Cuáles son algunos de los conceptos matemáticos clave para analizar datos?

ILUMINAR 03

ANÁLISIS DE DATOS: INTRODUCCIÓN AL TRABAJO CON HOJAS DE CÁLCULO

Introducción

Hay una diferencia entre datos y evidencia. Para convertirse en evidencia, los datos deben ser analizados e interpretados. Para nuestros propósitos, esto significa preguntar: ¿Qué nos dicen los datos sobre los indicadores que hemos identificado para cada una de las dimensiones del Marco OPERA? A veces, los datos se pueden leer e interpretar con bastante facilidad. Otras veces, el conjunto de datos (los datos primarios o secundarios recopilados para un análisis) es demasiado grande para leerse fácilmente en una o dos páginas. Cuando este es el caso, una hoja de cálculo puede ser una herramienta útil. Este es particularmente el caso cuando se trabaja con datos presupuestarios.

Trabajar con hojas de cálculo puede ser una nueva habilidad para muchos activistas. Esta nota te entrega habilidades básicas para usar hojas de cálculo. También describe algunos de los conceptos matemáticos clave que se pueden usar para analizar datos en una hoja de cálculo. Tiene la intención de ser una introducción, ¡Siéntete libre de omitir las partes que ya conoces!

Es importante sentirse seguro de la capacidad de trabajar con números. Nos equipa como activistas para analizar datos nosotros mismos o para colaborar con economistas, estadísticos y otros para llevar a cabo dicho análisis.

Conceptos básicos de la hoja de cálculo

Ingresar datos en una hoja de cálculo permite organizarlos, reorganizarlos y analizarlos con más detalle, para probar hipótesis sobre cómo se relacionan entre sí los diferentes

datos. **Una hoja de cálculo es esencialmente una tabla de celdas organizadas en filas y columnas.** Las columnas normalmente se representan con letras, mientras que las filas se representan con números. Cada celda se denomina por su fila y columna; "C10", por ejemplo.

Hay una variedad de programas y aplicaciones de hojas de cálculo disponibles, como Microsoft Excel, Calc de OpenOffice y Google Sheets. **Para esta nota, utilizaremos Microsoft Excel.**

IMPORTACIÓN DE DATOS SECUNDARIOS

A menos que estemos ingresando datos primarios en la hoja de cálculo, el segundo paso después de crear una hoja de cálculo es importar datos. Estos pueden ser datos primarios que hayamos registrado en otro documento, o datos secundarios de una fuente externa como el [Banco Mundial](#) o un ministerio. Si estamos descargando datos secundarios de una fuente en línea, debemos asegurarnos de que estén en el formato correcto. Tendremos la opción de elegir el formato de los datos que deseamos descargar de la fuente de datos. Uno de los formatos más comunes para descargar datos es CSV, que significa "valores separados por comas". Este es un formato que Excel puede leer fácilmente.

Para importar datos, hay que seleccionar la pestaña "Datos" en la parte superior de la hoja de cálculo. Esto te mostrará diferentes opciones para importar datos (por ejemplo, "Desde el acceso", "Desde la web", "Desde el texto"). Si seleccionas "Desde texto", se abrirá un cuadro de diálogo, desde el cual seleccionará el archivo que has descargado previamente.

Este documento está organizado de acuerdo a un innovador sistema para recolectar, analizar y presentar evidencia en tres pasos:



INTERROGAR

Mapea el problema a fondo usando el Marco OPERA para identificar indicadores y puntos de referencia.



ILUMINAR

Destaca los problemas subyacentes recolectando, analizando y visualizando datos.



INSPIRAR

Toma acción para construir poder y vigilar la responsabilidad de los tomadores de decisiones.

CONSEJO ÚTIL

Siempre que descarguemos un conjunto de datos, lo primero que hay que hacer es una copia del archivo. Cualquier cambio que hagamos debe hacerse en esta copia. Con esto, podremos volver atrás y comprobar los datos originales en cualquier momento. También es una buena práctica anotar de dónde obtuvimos esos datos y cuándo y cómo se recuperaron.

‘Si aparece el cuadro de diálogo “Asistente de importación de texto”, puedes dejar las selecciones establecidas en sus valores predeterminados:

- Paso 1 (tipo de datos): Delimitado.
- Paso 2 (delimitadores): Tabulador.
- Paso 3 (formato de datos de columna): General.

Algunos atajos de teclado útiles incluyen:

- Copiar [Ctrl + c]: copia las celdas seleccionadas en el portapapeles.
- Pegar [Ctrl + v]: pega el portapapeles.
- Cortar [Ctrl + x]: copia las celdas seleccionadas en el portapapeles y las elimina de su posición original.
- Deshacer [Ctrl + z]: deshace el último cambio realizado.
- Rehacer [Ctrl + y]: deshace un deshacer.

INGRESO DE DATOS PRIMARIOS

Si estás ingresando tus propios datos en una hoja de cálculo, es crucial hacerlo de manera consistente y estandarizada. Los datos que se ingresan de manera inconsistente son más difíciles de buscar, contar, clasificar y filtrar. Como ejemplo, aquí hay información demográfica que se ha ingresado en una hoja de cálculo:

Nombre	Sexo
Arun	M
Hyun	Masculino
Eleanor	F

El problema aquí es fácil de detectar: el sexo de los individuos se ha registrado de diferentes maneras. Introducir los datos de forma estandarizada evita estos errores. Esto significa tomar decisiones sobre cómo se pueden representar los datos de manera consistente. Los datos que tienden a registrarse de manera inconsistente incluyen:

- **Fechas y horas:** 01 de julio de 2005, 1º de julio de 2005, 1/7/2005 o 20050701 son formas de representar la misma fecha.
- **Nombres:** Por ejemplo, ¿usarás “Naciones Unidas” u “ONU”? ¿Pondrás el nombre y apellido de una persona en columnas diferentes o en la misma columna?

- **Lugares:** ¿Cuán específico se debe ser al describir datos geográficos? ¿Debemos utilizar la geografía administrativa de un país (p. ej., pueblo, ciudad, distrito), la geografía electoral (distrito, circunscripción) o la geografía operativa?

El punto clave es considerar las diferentes formas de registrar datos, hacer una elección y luego aplicarla consistentemente. Esto te ahorrará una enorme cantidad de tiempo y frustración.

El segundo problema es cómo “codificar” (o clasificar) los datos. Esto es particularmente importante si estamos registrando datos cualitativos que someterán a un análisis cuantitativo. Por ejemplo, supongamos que deseas registrar información sobre las escuelas que has visitado, clasificándolas como pequeñas, medianas o grandes. Cuando tú y tus colegas revisen la hoja de cálculo, debes estar seguro de que cada vez que veas una escuela descrita como “pequeña”, significa lo mismo. Para asegurarse de que los datos se codifiquen de forma coherente, puede diseñar un conjunto de reglas de codificación para que todos los que trabajen con los datos sepan que:

- Pequeño = entre 0 y 99 alumnos
- Medio = entre 100 y 499 alumnos
- Grande = entre 500 y 999 alumnos.

Es crucial asegurarse de que todos los que ingresen los datos sigan las mismas reglas cada vez.

BLOQUEAR FILAS Y COLUMNAS

Antes de comenzar a ordenar sus datos, puede ser útil “bloquear” la fila superior y/o la primera columna de su hoja de cálculo, especialmente si la hoja de cálculo con la que está trabajando es grande. Así, siempre podremos saber a qué se refiere cada fila y/o columna. Por ejemplo, es posible que hayamos descargado datos sobre el gasto público en educación como porcentaje del PIB de todos los países para los años 1990-2014. Esto incluirá muchas filas y muchas columnas. Para mantener visibles los años y los nombres de los países, estas filas y columnas se pueden bloquear para que no desaparezcan a medida que nos desplazamos por la hoja de cálculo.

CLASIFICACIÓN DE DATOS

Lo primero que hay que hacer al mirar un nuevo conjunto de datos es familiarizarse con ellos. Esto implica clasificar los datos para que tengan más sentido. Digamos, por ejemplo, que tenemos una hoja de cálculo con datos sobre el PIB, el gasto en atención médica y la esperanza de vida para todos los países, durante varios años. Para familiarizarnos con los datos, es posible que deseemos ver los datos primero en orden por país y luego en orden por fecha. Para ordenar toda la hoja de cálculo, sigue los siguientes pasos:

1. Selecciona la pestaña “Datos” y luego el botón “Ordenar”. Esto abrirá un cuadro de diálogo adicional
2. Selecciona la columna por la que le gustaría ordenar en el menú desplegable; puede ser de forma ascendente o descendente (o de la A a la Z o de la Z a la A, si es alfabético).

- También puedes elegir varios criterios para ordenar (por ejemplo, primero por país, luego por fecha), haciendo clic en el botón "Agregar nivel".

FILTRADO DE DATOS

Es común filtrar los valores que no deseamos ver en el conjunto de datos. En el conjunto de datos mencionado anteriormente, por ejemplo, es posible que solo te interese revisar los valores más actualizados.

Para filtrar toda la hoja de cálculo, sigue los siguientes pasos:

- En la pestaña "Datos", selecciona el botón "Filtro". Ahora deberías ver triángulos junto a los nombres de las columnas en la primera fila.
- Haz clic en el triángulo junto a la columna que te interesa. Debería aparecer un cuadro de diálogo en el que puedes seleccionar y deseleccionar las opciones que deseas incluir o excluir.

Análisis de datos con hojas de cálculo

Una vez que hayas ordenado y filtrado la hoja de cálculo, es hora de comenzar el análisis. El análisis de datos ayuda a extraer respuestas a tus preguntas. Supongamos que tienes datos sobre el PIB, la población, la esperanza de vida y el gasto en atención médica como porcentaje del PIB, para todos los países durante varios años. Hay muchas preguntas sobre el derecho a la salud que podríamos responder con esta información. Por ejemplo:

- ¿Cuánto gasta cada país en total en atención médica (en USD)?
- ¿Cómo se compara un país específico con otros?
- ¿Cuánto se gasta per cápita en cada país (en USD)?
- ¿Cómo ha cambiado el gasto en cada país con el tiempo?

Es importante recordar que los datos en la hoja de cálculo no siempre entregarán toda la información que necesitamos para responder las preguntas de investigación. Por ejemplo, según los datos que tenemos, no sabemos la cantidad de dinero que se gasta en atención médica en cada país. Sin embargo, podemos calcular esta cifra usando los datos que tenemos, a saber, el PIB por país y el gasto en salud como porcentaje del PIB. Puedes calcular nuevos valores como estos usando fórmulas de hoja de cálculo.

USAR FÓRMULAS

Las fórmulas son funciones matemáticas básicas; el tipo de cosas que encontrarías en una calculadora simple. Al escribir estas fórmulas en una celda, puedes realizar cálculos. Los símbolos más comunes que usan estas fórmulas incluyen:

Las reglas matemáticas básicas sobre el orden de las funciones también se aplican al trabajar con hojas de cálculo. Por ejemplo, la fórmula $=3+5*2$ será igual a 13, no a 16. Si deseas cambiar el

Símbolo	Descripción
=	Esto es lo primero que debe ir en la celda donde ponga una fórmula, porque le dice a la hoja de cálculo que está escribiendo una fórmula.
+	Sumar
-	Restar
*	Multiplicar
/	Dividir

orden de la función, debes incluir paréntesis. Las fórmulas entre paréntesis se calcularán antes que cualquier otra fórmula. Por ejemplo, si deseas que la fórmula anterior dé como resultado 16, debe ingresarse como: $=(3+5)*2$

Sumar significa añadir valores en un rango de celdas en una hoja de cálculo. Una forma rápida de hacer esto es usar el símbolo =SUMA (en lugar de vincular cada celda individualmente con un símbolo "+"). Por ejemplo, para calcular el PIB total de un grupo de 10 países, sigue los siguientes pasos:

- Los datos sobre el PIB se encuentran en la columna B. En la primera fila vacía de la columna B, al final de la lista del PIB de los países en 2010, escribe =SUMA y luego un paréntesis abierto; =SUMA(.
- Luego selecciona la celda con el primer país del grupo. Mientras mantienes presionada la tecla MAYÚS, selecciona el último país del grupo. La celda de fórmula debe agregar automáticamente los números de celda.
- Complete la fórmula agregando un paréntesis al final; por ejemplo, =SUMA(B2:B11). Esto sumará los valores del PIB en el rango dado.

La **multiplicación** y la **división** son útiles si deseamos convertir un valor numérico en un porcentaje, o viceversa. Por ejemplo, queremos averiguar cuánto se gasta (en USD) en atención médica en total en su país. Sabemos que el PIB es de USD 100.357.000.000 y que el gasto público en salud es el 3,48 % del PIB. Con estos datos, podemos usar la siguiente fórmula:

PIB * (gasto en salud/100)
 O en una hoja de cálculo =100357000000*(3.48/100)

No necesitas escribir manualmente los números; basta con hacer clic en la celda. La fórmula se parece a: =C2*(D2/100), donde la celda C2 es el PIB de su país y la celda D2 es el gasto en salud como porcentaje del PIB de su país.

Copiar fórmulas es útil cuando deseas aplicar el mismo cálculo en filas o columnas. Para hacer esto, simplemente copia la fórmula que acabas de escribir (usando Ctrl + c) y péguela en la celda de abajo (usando Ctrl + v). Alternativamente, haz clic

en la esquina inferior derecha de la celda (el cuadrado azul) y arrastra la fórmula hasta el final de la columna.

Las **celdas vacías** pueden ser problemáticas cuando comienzas a analizar datos, porque pueden terminar produciendo valores cero. El manejo adecuado de los valores faltantes es un paso importante en la limpieza y el análisis de datos. Los grandes conjuntos de datos rara vez están completos, por lo que debes tener una estrategia para lidiar con las partes que faltan. Una forma de encontrar errores es filtrar los datos, seleccionar todos los valores "0" y eliminarlos.

Conceptos matemáticos útiles para el análisis de datos

Ordenar datos para ayudar a responder tus preguntas de investigación puede ser desalentador si estás tratando de comparar múltiples variables. Por ejemplo, si estás buscando datos relacionados con el derecho a la alimentación, es posible que desees ver: (a) disponibilidad per cápita; (b) de los principales artículos alimenticios; (c) entre municipios. Puede que te interese saber la disponibilidad *media* de estos artículos, el *mínimo* y el *máximo* disponibles, y la diferencia o la *variación* entre el mínimo y el máximo. Encontrar las respuestas a estas preguntas se conoce comúnmente como estadística descriptiva.

RANGO

La primera información que quizás desees encontrar en tus datos es el rango; en otras palabras, ¿de dónde a dónde se extienden los datos? ¿Empieza con números pequeños? ¿Números grandes? ¿Va de negativo a positivo? Todo esto es información esencial que ayudará a manejar los datos.

Mirar el rango también te ayudará a encontrar errores. Por ejemplo, supongamos que estás viendo datos relacionados con el indicador "Promedio de años de escolaridad en la población adulta". Es posible que encuentres que los datos van de 4 a 58. Claramente hay un error; la probabilidad de que cualquier adulto en cualquier país tenga 58 años de escolaridad es muy pequeña. Debes volver a los datos y comprobarlo.

Entonces, ¿cómo se encuentra el rango? Simplemente revisa los datos y encuentra los valores mínimo y máximo: el más bajo y el más alto, respectivamente. Por ejemplo, supongamos que tienes los siguientes datos sobre el promedio de años de escolaridad en la población adulta de tu propio país, así como de varios países vecinos:

8.5, 5.8, 6.5, 7.6, 10.2, 8.4, 7.3, 7.2, 9.2, 9.3

- Pregunta: ¿Cuál es el rango de su conjunto de datos?
- Respuesta: El número más bajo (mínimo) es 5.8 y el número más alto (máximo) es 10.2. Por lo tanto, el rango es de 5.8 a 10.2.

En una hoja de cálculo, puedes hacer esto ordenando los datos de menor a mayor, o con las fórmulas **=MIN** y **=MAX**, usando paréntesis para seleccionar las celdas que desees incluir en el cálculo; por ejemplo, **=MIN(D12:D84)**.

CONTAR

La siguiente información importante que quizás desees determinar es para cuántas cosas tiene datos. ¿Cuántos países? ¿Cuántos hogares? Y así.

¿Cómo obtener esta información? Simplemente cuéntala. En el conjunto de datos anterior, por ejemplo, hay 10 observaciones.

Si el conjunto de datos es demasiado grande para contarlo manualmente, puedes usar las fórmulas **=CONTAR** cuando tengas números en sus celdas o **=CONTARA** cuando no haya números en sus celdas. Nuevamente, usa paréntesis para seleccionar las celdas que desea incluir en el cálculo; por ejemplo **=CONTARA(A5:A2089)**.

Esto puede parecer simple. Sin embargo, cuando se trata de análisis e interpretación de datos, es muy importante. Por ejemplo, si estás comparando datos sobre hospitales en tu país, ¿los datos de 10 hospitales son suficientes para hacer esa comparación?

PROMEDIOS

La siguiente información relevante es el **valor central** y cómo se distribuyen los datos en relación a ese valor. ¿El valor central da una buena indicación de todo el conjunto de datos, con una distribución equitativa de los puntos de datos por encima y por debajo de él? ¿O la distribución está "sesgada", en el sentido de que hay un pico en un extremo del rango de datos con una cola larga hacia el otro? La distribución te dice qué tipo de descriptores adicionales son prácticos de usar. Hay varias maneras diferentes de responder a estas preguntas.

La **media** (o "promedio") es la forma más común de ver el "valor central". Resulta familiar en los informes; por ejemplo, el promedio de desempleo ha aumentado en el país X o las tasas promedio de alfabetización han disminuido en el país Y.

Entonces, ¿cómo se calcula la media? La media es la suma de todos los valores en el conjunto de datos dividida por la cantidad de valores que hay. Por ejemplo, supongamos que quieres determinar el ingreso familiar promedio. Tienes los siguientes datos sobre el ingreso familiar anual promedio, medido en dólares, para varios hogares: 1120, 241, 876, 201, 112, 345, 567, 156, 154, 1345

- Pregunta: ¿Cuál es el promedio del conjunto de datos?
- Respuesta: Suma los ingresos que tienes para todos los hogares (1120 + 241 + 876 + 201 + 112 + 345 + 567 + 156 + 154 + 1345 = 5117). Luego divide ese número por el número de hogares (5117/10). La respuesta es 511.7.

En una hoja de cálculo, puedes calcular esto usando la fórmula **=PROMEDIO**.

La media brinda una buena estimación de lo que es "normal" cuando el resto de sus datos se distribuyen uniformemente por encima y por debajo. Sin embargo, si el resto de sus datos está "sesgado" hacia algún lado, puede ser más apropiada una medición diferente del promedio. Por ejemplo, si hay un pequeño número de personas con ingresos extremadamente altos en un

grupo de población, usar la media hará que el ingreso per cápita promedio parezca más alto de lo que realmente es.

La **mediana** es el valor numérico que separa la mitad superior de los valores del conjunto de datos de la mitad inferior. Es útil cuando el resto de los datos no están distribuidos uniformemente a ambos lados de la media. En el ejemplo anterior del ingreso familiar promedio, el valor del ingreso promedio es 511.7. Sin embargo, este es un número bastante grande en relación con la mayoría de los valores en el conjunto de datos. Esto ocurre porque hay unos pocos hogares con ingresos muy altos, lo que sesga los datos. En este caso, la mediana puede proporcionar una mejor estimación de lo que es “normal” que la media.

Entonces, ¿cómo se calcula la mediana? En primer lugar, hay que ordenar los datos (en forma ascendente o descendente, no importa) y el valor en el medio del conjunto de datos es la mediana. Si hay un número par de valores en el conjunto de datos, tome el promedio de los dos valores centrales.

- Pregunta: ¿Cuál es el ingreso familiar mediano?
- Respuesta: Primero, ordene los datos: 112, 154, 156, 201, 241, 345, 567, 876, 1120, 1345. Hay 10 valores y los dos valores del medio son 241 y 345. Halle el promedio entre estos dos números ($241 + 345 / 2 = 293$). La mediana es 293.

En una hoja de cálculo, puedes calcular esto usando la fórmula **=MEDIANA**.

La **moda** es el valor que aparece con mayor frecuencia en un conjunto de datos. A veces, ni la media ni la mediana nos dicen realmente lo que queremos saber. Por ejemplo, para conocer el número promedio de niños por hogar matriculados en la escuela, podemos tener el siguiente conjunto de datos:

0, 1, 1, 1, 1, 2, 2, 2, 3, 5

El número medio de niños matriculados en la escuela por hogar es de 1,8 y la mediana es de 1,5. Pero lo que realmente se quiere averiguar es cuántos niños están matriculados en la escuela en la mayoría de los hogares. Vemos que “1” niño es la respuesta más frecuente. Esta es la moda.

En una hoja de cálculo, esto se puede calcular con la fórmula **=MODA**.

En el caso de que más de un valor sea el más frecuente, el conjunto de datos puede ser bimodal (dos valores medios) o multimodal (más de dos valores medios).

VARIACIÓN

Otro elemento relevante es el tamaño de la variación en el conjunto de datos. Esto es crucial cuando se analizan datos agregados y desagregados. Por ejemplo, digamos que estamos investigando la realización del derecho al trabajo y queremos saber la tasa de desempleo promedio. Sin embargo, también podemos querer probar qué tan representativo es el promedio de diferentes municipios dentro del país. Hay dos medidas comunes para hacer esto.

La **desviación estándar** es una medida de cuánto, en promedio, los valores que muestran los datos están fuera de la media. Los siguientes tres pasos muestran cómo calcularla:

1. Suma el cuadrado de las diferencias entre los valores y la media.
2. Divide esa suma por el número de valores menos uno.
3. Saca la raíz cuadrada.

En una hoja de cálculo, puedes hacer esto con la fórmula **=STDEV**.

Supongamos que tenemos los siguientes datos sobre tasas de desempleo como porcentaje de la población activa total para cuatro municipios diferentes:

1, 2, 3, 4

- ¿Cuál es la desviación estándar del conjunto de datos?
- La media es 2,5, por lo que siguiendo los pasos 1 a 3:

Valor	Diferencia con la media	Diferencia al cuadrado
1	-1.5	2.25
2	-0.5	0.25
3	0.5	0.25
4	1.5	2.25

1. Suma el cuadrado de las diferencias entre los valores y la media = 5.
2. Divide esa suma por el número de valores menos uno = $5 / (4 - 1) = 5 / 3$.
3. Saca la raíz cuadrada = 1.291.

Si los datos se distribuyen normalmente (todos los puntos de datos se distribuyen uniformemente a ambos lados de la media), entonces el 68.27 % de los puntos de datos estarán dentro de una desviación estándar de la media y el 95.45 % de los puntos de datos estarán dentro de dos desviaciones estándar de la media. Cuanto mayor sea el valor de la desviación estándar, mayor será la variación en el conjunto de datos.

Usando el ejemplo del desempleo, una desviación estándar de 1.291 significa que el 68,27 % de los puntos de datos estarán dentro de esta distancia de la media (suponiendo una distribución normal). Esto sugiere que puede haber una variación geográfica razonable en el desempleo, por lo que informar sólo la media puede ser engañoso.

La **desviación absoluta mediana** es similar a la desviación estándar, pero se usa con la mediana en lugar de la media. Los siguientes dos pasos muestran cómo calcularla:

1. Calcular la mediana y las diferencias absolutas entre cada valor y la mediana.
2. Calcular la mediana de las diferencias.

Para tomar el mismo ejemplo de desempleo anterior, pero con datos de un municipio adicional, nuestros datos son 1, 2, 3, 4, 5.

1. La mediana es 3 y las diferencias absolutas son 2, 1, 0, 1, 2.
2. Ordenar las diferencias 0, 1, 1, 2, 2 y la desviación absoluta mediana es 1.

NORMALIZACIÓN

Una vez que tengamos una idea de los datos con los que estamos tratando, podemos comenzar a hacer comparaciones. Por ejemplo, en una investigación sobre la medida en que se cumple el derecho a la salud, es posible que deseemos comparar el gasto público en salud con el de otro país que difiere en varios aspectos. Comparar el valor total del gasto público en atención médica en este caso no nos dirá algo si, por ejemplo, un país es muy grande y el país con el que lo comparamos es muy pequeño. Lo más probable es que el país grande gaste mucho más en salud en total que el país pequeño. ¿Significa esto que el país más grande está cumpliendo con su obligación de garantizar el derecho a la salud mejor que el país más pequeño? No necesariamente. Para responder a esa pregunta, los dos países deben compararse en igualdad de condiciones. Esto generalmente se hace usando un indicador que nos dice qué tan grande es un país; a menudo el tamaño de su población. Para comparar el gasto del gobierno en salud, se puede dividir el gasto total por la población. Esto se llama **normalización**.

Aplicar estos conceptos al analizar datos presupuestarios

Analizar el presupuesto de un gobierno a través del lente de los derechos humanos implica algunos cálculos específicos. Esto se debe a que las cifras de un presupuesto siempre son relativas.

CONSEJO ÚTIL

Si trabajamos con un conjunto de datos muy grande, puede llegar un punto en el que la hoja de cálculo nos quede pequeña. Si esto sucede, hay que usar otro software de base de datos como SPSS, Stata o R. A diferencia de una hoja de cálculo, que está diseñada para poder “leerse” en la pantalla, la forma en que se almacenan los datos en una base de datos a menudo está completamente oculta para el usuario. Esto permite formas abstractas y complejas de almacenar grandes cantidades de datos y brinda al usuario más flexibilidad en cómo usarlos. Dicho esto, las bases de datos son una forma mucho más técnica de almacenar y analizar datos, y si las usas, es posible que debas trabajar con estadísticos, programadores y diseñadores.

Analizar si las cifras presupuestarias son altas o bajas, por ejemplo, implica preguntar alto o bajo *en relación a qué*. Tomando un ejemplo hipotético, digamos que el presupuesto de vivienda social de un país pasa de 100 millones de dólares en 2015 a 200 millones de dólares en 2020. Duplicar el presupuesto de vivienda social puede parecer bastante significativo. Sin embargo, ¿qué pasaría si el presupuesto general del gobierno se triplicara en ese mismo período? Como parte porcentual, el presupuesto de vivienda social en realidad se reduce. ¿Qué pasa si, debido a la inflación, el costo de construir viviendas sociales aumenta un 150% en el mismo período? En este escenario, el poder adquisitivo del gobierno disminuye, por lo que el presupuesto no se estira tanto. Debido a esta relatividad, a menudo es necesario convertir las cifras presupuestarias.

CONVERTIR NÚMEROS EN PORCENTAJES O RATIOS

El uso de porcentajes es una forma de hacer que los datos sean comparables. Por ejemplo, es posible que queramos comparar los ingresos fiscales de dos países. Una forma de hacerlo es normalizar los ingresos fiscales como porcentaje del PIB de ambos países, lo que puede mostrar que el primer país recauda una cantidad relativamente alta de ingresos, mientras que el segundo no.

CÁLCULO DE ASIGNACIONES PER CÁPITA

Para descubrir patrones discriminatorios de asignación de recursos, el presupuesto también debe evaluarse en términos de cómo distribuye los beneficios entre los hogares y las personas. Para hacer esto, es posible que debamos calcular el gasto per cápita; por ejemplo, por región o municipio. Puedes hacerlo simplemente dividiendo la asignación total a la región o municipio por la población total de esa región o por el grupo de población total que nos interesa (como niños en edad escolar o mujeres en edad reproductiva).

AJUSTE POR INFLACIÓN

Es importante comparar las asignaciones presupuestarias a lo largo del tiempo y evaluar si ha habido un aumento o una disminución en los montos asignados a diferentes sectores. Esto es importante porque nos dice algo acerca de si un gobierno está tomando medidas para hacer realidad “progresivamente” estos derechos.

Sin embargo, esto no es tan simple como mirar las cifras presupuestarias de diferentes años. Las cifras presupuestarias se consignan en términos “nominales”, lo que significa que

CONSEJO ÚTIL

International Budget Partnership ha desarrollado una calculadora de inflación. Para utilizarla basta con introducir los importes nominales de cada año y el Índice de Precios al Consumidor (IPC) de cada año y calcular. Ver: https://www.bls.gov/data/inflation_calculator.htm

no tienen en cuenta la inflación. Por esta razón, es necesario convertir las asignaciones de montos “nominales” a “reales”. Esto hace que las cifras presupuestarias de diferentes años sean “equivalentes” a los valores actuales de uno de los años, y permite realizar comparaciones válidas a lo largo del tiempo. En otras palabras, valor real = valor nominal ajustado por inflación.

FÓRMULA PARA AJUSTAR POR INFLACIÓN

$$\text{Valor Real} = \frac{\text{Valor nominal del año objetivo} \times \text{índice de precios al consumidor (IPC) del año base}}{\text{IPC del año base}}$$

Por ejemplo, el dinero de 2010 en valores de 2000 se calcularía como:

$$\text{Valor Real} = \frac{\text{Valor de 2010} \times \text{IPC del 2000}}{\text{IPC del 2010}}$$

Interpretación y cómo evitar conceptos erróneos comunes

Si bien se requiere simplificación para comprender lo que significan los datos, como cuando presenta evidencia en un formato gráfico, es crucial mantenerse lo más cerca posible de la “historia completa”.

CORRELACIÓN NO ES CAUSALIDAD

En general, es extremadamente difícil establecer causalidad entre dos observaciones correlacionadas. Hay varias razones por las que las conclusiones de sentido común acerca de la causa y el efecto pueden estar equivocadas. Por ejemplo, es posible que deseemos informar sobre la relación entre la educación y la salud. Usando un gráfico de dispersión, se tiene el promedio de años de escolaridad en un eje y la esperanza de vida promedio en el otro. El diagrama de dispersión revela que las dos variables están altamente correlacionadas positivamente: un mayor promedio de años de escolaridad está asociado con una mayor esperanza de vida promedio. Pero, ¿puede decirse que un mayor promedio de años de escolaridad provoca una mayor esperanza de vida promedio? No. Puede haber varias razones para esta asociación:

- Un mayor promedio de años de escolaridad puede causar una mayor esperanza de vida promedio.
- Una mayor esperanza de vida promedio puede causar un mayor promedio de años de escolaridad.
- Mayores años de escolaridad promedio y mayor esperanza de vida promedio son consecuencias de una causa común, pero no se causan entre sí. La causa común puede ser un ingreso promedio más alto, por ejemplo, o una distribución más equitativa del ingreso.
- No existe relación entre el promedio de años de escolaridad y la esperanza de vida. La correlación es casual.

CAMBIO EN PORCENTAJE Y CAMBIO DE PUNTO PORCENTUAL

El cambio en porcentaje y el cambio de punto porcentual a veces pueden confundirse en la interpretación de los datos. Por ejemplo, si un valor cambia del 5% al 10%, ¿cuál es el cambio porcentual? Un error común sería responder el 5%. Esto es incorrecto. El cambio porcentual en cuestión es en realidad 100%. Es, sin embargo, un cambio en cinco puntos porcentuales. La elección de informar el cambio porcentual o el cambio de punto porcentual dependerá de su pregunta. A menudo es útil reportar ambos.

REFLEXIONES FINALES

Desarrollar la confianza necesaria para trabajar con números nos ayuda a equiparnos mejor como activistas, ya sea para analizar datos nosotros mismos o para colaborar con economistas, estadísticos y otros en la realización de un análisis. Hacerlo puede llevar tiempo. Implica experimentar y explorar, lo que significa jugar con los datos que tenemos. Por ejemplo, podemos comparar variables, desagregarlas, mirarlas a lo largo del tiempo, etc. Esto nos ayuda a ver qué es lo más relevante para nuestra investigación.

Dicho esto, es muy importante mantener un registro de cualquier cambio realizado en los datos. Este es especialmente el caso si diferentes personas están trabajando en un conjunto de datos. Esto hace posible realizar un seguimiento de los errores que se cometen, lo que facilita volver atrás y corregirlos.

También es importante recordar que procesar números no es el único paso para interpretar los datos. A menudo, todavía hay más en lo cual profundizar para responder por qué una situación es como es. Hacerlo es crucial para decidir qué conclusiones sacar de su investigación. Este tema se aborda en [la quinta nota de este módulo](#).