Periodismo de Datos
Matemática y Estadística Para Periodistas

Por Sandra Crucianelli
www.facebook.com/sandracrucianelli
www.facebook.com/periodismodedatos
@spcrucianelli

¿Por qué Matemática y Estadística para Periodistas?

De todos los males que enfrenta el periodismo latinoamericano, el más grave en su superficialidad. Gino Lofredo (1) publicaba hace unos años en la revista Chasqui(2), que los medios parecen ir de mala gana detrás de las noticias, como si ignoraran el trascendente rol que les compete en el fortalecimiento de las democracias. Gran parte de este problema obedece al abuso de las fuentes orales en desmedro de las documentales, pero también al erróneo manejo de números dentro de la crónica periodística. El periodista no ha sido entrenado para el manejo de información numérica.

En un mundo donde casi todo está dominado por números, necesitamos herramientas que provienen de las ciencias para comprender mejor el terreno en el que se producen las noticias. La matemática y la estadística forman parte del conjunto de recursos que se deben dominar. Para quienes estudiaron periodismo con la intención de alejarse de los números, esta constituye una mala noticia. Es que las ciencias auxiliares del periodismo, y en este caso del periodismo investigativo, deben convertirse en aliadas, no en enemigas.

Nos consta a muchos, que decenas de papeles, documentos e información de toda índole suele pasar frente a los ojos de los periodistas sin que nos percatemos qué se esconde detrás de lo que tenemos entre las manos. Debemos empezar por reconocer que no estamos entrenados para manejar números y hoy día de ellos dependen nuestras economías, nuestras políticas y por ende, nuestras vidas.

Los números están presentes en los presupuestos, en las leyes, en las bases de datos, en los censos, en las estadísticas, en los sondeos de opinión, en las encuestas. Si tomáramos un diario cualquiera y comenzamos a contar cuántas crónicas incluyen algún número como dato notaríamos que son la mayoría.

El periodismo político, el que cubre las áreas de gobierno, el económico, el científico, el social, el deportivo no escapan a la necesidad de manejar con pericia los números dentro de la crónica.

En los Estados Unidos, el uso de la matemática y estadística en periodismo fue introducido como "Periodismo de Precisión" hace más de 25 años, cuando Phillip Meyer (3) publicó su primer libro, Precision Journalism (1973) seguido por un segundo volumen, The New Precision Journalism (1991).

La misma corriente siguió José Luís Dader (4) en España, bajo el mismo nombre.
En América Latina, la enseñanza de estos conceptos de manera integral, aplicada a la construcción de la noticia, es una deuda pendiente.

Definición de Número

Un número es la expresión simple de una cantidad. Puede definirse también como la totalidad de casos que comparten un atributo (dinero, personas, objetos, etc.). Obviamente, se trata de un atributo cuantitativo. Otros, como el sexo y ocupación, son atributos cualitativos, pero al ser analizados en grupos se transforman en cantidades. La pregunta que debemos formularnos entonces es, ¿qué representan los números?

La matemática está asociada a la precisión. Un atributo que debe estar presente en todo buen trabajo periodístico. Y sólo hay precisión con: 1) Datos confiables: Esto es un problema en América Latina, dada la limitante que aún se observa en materia de acceso a documentos públicos, pero lo podemos compensar con imaginación, creatividad, vocación y conocimiento, 2) Representatividad y 3) Razonamiento científico

Estas tres herramientas no son para medios ricos, sino para periodistas anticonvencionales, pero entrenados.

"El periodismo cuenta noticias, no las crea ni las construye". Esto es un mito del periodismo. En realidad podemos afirmar que "el periodismo no inventa noticias, pero puede construirlas a partir de la observación y la deducción".

Debemos pensar en los números como aliados, como herramientas ya que dicen mucho más de lo que parece.

Proceso en el Manejo de Números

Comprende los siguientes pasos:

1. Recopilación
2. Introducción en Hojas de Cálculo (Excel)
3. Análisis
4. Resumen: selección de datos, ¿qué contribuye a mi historia?)
5. Interpretación
6. Comunicación

La información respecto de cómo se obtuvieron los números, cómo se llegó a las conclusiones que se presentan, debe formar parte de la crónica. Esto hace que el artículo sea sólido y transparente. Por otra parte permite la revisión crítica y si pasa esa prueba, es válido. Cuando un periodista investiga, por ejemplo la evolución de un presupuesto público a lo largo de una década, su hipótesis inicial no debe condicionarlo. El reportero debe mirar los números e interpretar lo que éstos le cuentan con su mente abierta, aceptando como condición de borde que tal vez, intenta probar un hecho que no existe o que no puede ser comprobado tácticamente.

Hay 2 tipos de datos: 1) Continuos, tales como censos, datos gobierno: se recopilan todo el tiempo, con determinada frecuencia. El proceso no se interrumpe y 2) Categoriales, (discretos o discontinuos), que son los que agrupan información que puede recogerse aisladamente, tales como encuestas, mediciones de consultoras, etc.

Operaciones Matemáticas Básicas para Periodistas

Para comenzar, recordemos una regla principal, que todo reportero habrá estudiado en la escuela: la Regla de 3 Simple o Regla de las proporciones
Daré un ejemplo que podría aplicarse a la labor en la sala de redacción:

Se derramaron 27 metros cúbicos (m^3) de petróleo en la costa norte del país. ¿Cuántas veces se llena la piscina olímpica hondureña si su capacidad es de 6.000 litros?

1 litro = 0,001 metro cúbico = 0,001 m^3

Aplicando la Regla de las Proporciones (dos veces):

Si 0,001 m^3......................... es 1 litro
En 27 m^3 ....X = (27x1)/0,001 = 27.000 litros

Si 6.000 litros............1 piscina
27.000 litros .... X = (27.000 x 1) / 6.000 = 4,5 piscinas

Conclusión periodística noticiable: "Con el petróleo derramado se podrían llenar 4,5 veces la piscina olímpica". Esto ayuda al editor a dimensionar un problema. Puede ser que las autoridades marítimas subestimen el problema; entonces es necesario un análisis independiente para decidir, por caso, si la noticia va en primera plana o merece un espacio menos relevante.

Las comparaciones siempre son útiles a la hora de redactar una crónica. Ayudan al lector a comprender mejor la magnitud de cierto problema.
Por ejemplo: "Para la construcción y montaje de la represa de Itaipú, fueron empleadas más de un millón de hojas, entre planos y listas de materiales. La superposición de esta documentación formaría una pila con una altura equivalente a un edificio de 50 pisos.

El volumen total del hormigón armado sería suficiente para construir un estadio como el Maracaná de Brasil en la capital de cada país del mundo. El hierro y acero empleados serían suficientes para construir 380 Torres Eiffel. Con su actual capacidad de descarga, el vertedero tiene capacidad para evacuar el equivalente 40 veces el caudal medio de las Cataratas del Iguazú"

Hubiera sido imposible escribir el texto anterior sin que el redactor conociera la Regla de las Proporciones. Nótese que los números no aparecen. Como bien señala Ron Nixon (4) "los números no son la historia. Las palabras cuentan la historia".

Los números y su propio análisis pueden darle al tema una cuota extra de profundidad y autoridad, pero son solamente el respaldo.

Apoyarse en la cronología de los hechos también ayuda a redactar mejores crónicas o guiones. La audiencia entiende mejor si ciertos procesos se describen en orden cronológico. La mayoría del trabajo del reportero que usa números consiste en estudiar variables (aquello que varía a lo largo del tiempo: presupuesto, inversión, producción, etc.) Es posible escribir acerca de los resultados del análisis de estas variables sin involucrar números en la historia.

Un consejo que suele darse es no excederse en la cantidad de números que forman parte de una crónica. Cuando apele a ellos, es mejor usarlos escasamente. Podríamos establecer como regla no emplear más de 3 números en un párrafo. La información numérica adicional puede ir en gráficos. Para lograr este objetivo, sustituir palabras por números es lo mejor. En vez de poner 50 %, prefiera decir "la mitad". Si está hablando de incrementos, prefiera decir "aumentó el doble" en vez de reportar un incremento porcentual del 100 %. De esa forma, estará usando los resultados de su análisis sin llenar la crónica de números que pueden marear, distraer o aburrir al público.

No hay que perder la perspectiva de que en general, los números que manejamos involucran de una u otra manera, a personas. Los censos hablan de personas. Las contribuciones a las campañas políticas no son más que personas dando dinero a otras personas. Los datos del gobierno como impuestos, producción, desempleo, ayuda social, también son acerca de personas. Por eso, es necesario no sentir temor frente al manejo de los números y recordar que están basados en personas reales, viviendo en un mundo real. Apelar a los ejemplos es otro recurso de oro. Es difícil que la gente entienda cuanto es una parte en un billón. O cuán pesadas son 23 toneladas.

Las conversiones de medidas se apoyan en tablas. Nunca supe cuán pesado podría ser un metro cúbico de arena contenido en una bolsa gigante, hasta que en plena obra de ampliación de mi vivienda observé que tuvo que ser transportada con la ayuda de máquina. Cuatro albañiles no podían moverla. Eran 2,5 toneladas de arena.

Para conversiones eficaces:
http://jumk.de/calc/capacidad.shtml (en la barra superior, seleccione categoría)
http://www.megaconverter.com/Mega2/
http://www.xe.net/ucc/
http://www.calculator.com/calcs/conv.html
http://www.calculator.com/pantaserv/simplefeet.calc
http://www.onlineconversion.com/
http://www.convert-me.com/en/

El Caso del Derrame de Petróleo

Ya se explicó la cuestión de la equivalencia con el ejemplo del derrame y las piscinas. En el ejemplo citado, no hubo conflicto con las unidades, pero generalmente, para el caso de derrames de petróleo, se suele reportar el dato en unidad de peso o masa. Supongamos que la autoridad marítima reporta un accidente en el mar e informa que un buque derramó 520 toneladas de petróleo. El reportero desea saber cuántas piscinas llena si el volumen de la que toma como referencia es de 5.000 litros.
Como podrá intuirse, es imposible pasar de una unidad de peso a una de volumen si no se dispone de un tercer dato: la densidad. En efecto, para convertir masa en volumen hay que dividirla entre su densidad. La fórmula resulta:

V = m/ d m= masa d=densidad

densidad del petróleo: 0,8 Kg/lt

V = 520.000 Kgs/0,8 Kg/lt = 650.000 litros

1 piscina
650.000 lt-------- x= 130 piscinas

Medidas de Superficie y Volumen

De las muchas fórmulas que existen para dimensionar superficies o volumen hay al menos tres que pueden ser de utilidad al reportero:

- Superficie Cuadrado (caso Plaza) = Lado al cuadrado = L^2

- Superficie Rectángulo (caso calle) = Lado Mayor x Lado menor = L x l

- Volumen de un Cuadrilátero (caso piscina) = Largo x Alto x Ancho

Caso Densidad de Concurrencia

1 paso de hombre equivale a = 50 centímetros ó 0,5 metros ó ½ metro

Cálculo empírico de cuantificación de personas:
En 1 metro cuadrado (1 m^2) se acomodan 4 personas comprimidas, 2 a distancia normal de ½ metro y 1 dispersa.

Esto es útil a la hora de calcular la cantidad de personas que concurrieron a una manifestación pública. Por ejemplo: ¿Cuántas personas integraron una manifestación popular muy concurrida realizada en una calle?

La columna de manifestantes ocupaba 150 metros de largo y 8 metros de ancho.

Solución:

Superficie = Lado Mayor x Lado menor = LM x Lm
S = 150 metros x 8 metros
S = 1.200 m^2 (metros cuadrados)

Si en 1 m^2 ------------------- hay 4 personas (muy concurrido)
En 1.200 m^2 --------------- X = ¿Cuántas habrá?

X = (1.200 x 4)/ 1 = 4.800 personas

Había aproximadamente 4.800 personas

En este caso se utilizó la regla de 3 simple o Regla de las Proporciones. El Razonamiento es sencillo: Si en un metro cuadrado es posible que (muy comprimidas) se ubiquen 4 personas, entonces en 1.200 metros cuadrados, ¿cuántas se ubicarán?
Para eso hay que hacer lo siguiente: primero se multiplican los dos extremos cruzados conocidos (1.200 x 4) y finalmente ese producto se divide por el número del extremo cruzado donde está la incógnita.

Ejercicio Adicional:

Imaginen que como reporteros van a cubrir una manifestación popular que se realiza por las calles de la ciudad. La extensión de la multitud es de aproximadamente 200 metros y el ancho de la calle es de 8 metros. La concurrencia era muy numerosa y casi no había distancia entre los presentes.
a) ¿Cuántas personas había?

La superficie sería de 1600 metros cuadrados. La cantidad de personas depende de como estaban dispuestas. Muy apretados, caben cuatro por metro, dos si están dispuestos en forma normal y 1 si pueden caminar con las manos extendidas. En el primer caso, muy apretados habrian 6400 En el segundo caso, dispuestos en forma normal (2 por metro cuadrado) 3200 En el tercer caso serían 1600

b) ¿Qué técnica de observación usarían para calcular a ojo, es decir aproximadamente, la extensión total de la multitud y de esa forma calcular las dimensiones del rectángulo que ocupa la manifestación?

La técnica de observación aconsejada para determinar qué figura geométrica ocupa una multitud es observar desde lo alto. Ir a un balcón o a una terraza y determinar a ojo y papel lo más preciso que se pueda dónde comienza la multitud, dónde termina y qué espacio ocupa...porque puede ser que la distribución no sea uniforme; que dentro de un rectángulo la tercera parte esté muy concurrida, la otra tercera parte normal y la restante dispersa...entonces tienen que calcular tres veces conforme la regla de las proporciones y sumarlas. Imposible cubriendo un evento desde el piso saber bien esto; por eso se aconseja una observación aguda desde la altura...

Números que Engañan

Imaginen que están cubriendo una manifestación popular en una plaza cuadrada. Pero tiene un monumento de base circular en el centro. Obviamente, para calcular la superficie de la plaza y poder averiguar cuántas personas concurrieron, hay que restar la superficie del círculo de la superficie del cuadrado.
La fórmula de la superficie de un círculo es

S = ¶ x r^2

Se lee como el producto de la constante "Pi" que es 3,1416 multiplicado el cuadrado del radio de la circunferencia. El radio es el punto que va desde el centro al borde.

Del mismo modo, si quisieran calcular el perímetro de un círculo, por ejemplo la base de un deposito cilíndrico de combustible que estalló y originó un incendio, la fórmula será P = ¶ x D (el producto de Pi por el diámetro, siendo el diámetro el doble del radio).

Conociendo este último dato, y para que vean cómo los números a veces pueden engañar nuestras percepciones de la realidad, les comento un caso que John Allen Paulos (6) cita en uno de sus libros y siempre menciona en sus clases a periodistas: Imaginen una cuerda que da la vuelta alrededor de la tierra a ras del suelo sobre el ecuador. ¿Cuánta cuerda adicional habrá que añadir para que la cuerda alargada de la vuelta alrededor de la tierra un metro por encima del suelo, también siguiendo al ecuador, si saben que el diámetro de la tierra es de 13.000 km? Uno podría pensar que se necesita mucha cuerda.

Veamos cómo se resuelve:

Resolución:
P = ¶ x D

Diámetro tierra: 13.000 km
Un metro por encima de la tierra: hay que agregar 1 m de un lado y 1 m del otro: 2 m = 0,002 kilometros
Diámetro 1 metro por encima: 13.000 km + 0,002 = 13000,002 kilometros
Perímetro Tierra: 3,1416 x 13. 000 = 40840,8
Perímetro 1 metro por encima = 3,1416 x 13000,002 = 40840,80628
40840,806 – 40840,8 = 0,006 kilometros = 6,28 metros

Criterios de Redondeo

Muchas veces introducimos datos numéricos en las crónicas sin caer en la cuenta de lo que intentamos transmitir.

Por ejemplo, ¿es correcto decir o escribir que el promedio de empleados públicos por oficina del gobierno es de 20,3? ¿Acaso una persona puede fraccionarse en una tercera parte, como lo indicaría el decimal?

Eliminación de 1 decimal: Se utiliza el criterio del 5

- 20,17 = 20,2 (7 es mayor que 5, por lo tanto, para eliminar el 7, se añade una unidad al número 1 después de la coma)
- 0,173 = 0,17 (3 es menor que 5, por lo tanto, las unidades anteriores al número que se quiere eliminar permanecen sin modificación)

Eliminación de 2 decimales: Se utiliza el criterio del 50

- 0,1798 = 0,18 (98 mayor que 50)
- 3,4919 = 3,49 (19 menor que 50)

Cuantificación de personas u objetos no fraccionables (casos judiciales, etc.)

- 17,6 personas = 18
- 15,3 personas = 15
- 15,5 personas = Entre 15 y 16
- 19,8 bebés = 20 bebés
- 58,2 ancianos = 58 ancianos

Notación Científica

Muchas estadísticas, censos y en especial, informes medioambientales o económicos contienen números extremadamente grandes como para que los incluyamos en una crónica o reportaje como notación numérica pura. Por consiguiente, hay que buscar otra manera de expresarlos.

La forma más regular de expresar números grandes es usando el exponencial, es decir elevar una base, el número 10 por ejemplo, a otro número. Significa que hay que multiplicar ese número base tantas veces como el número que indica el exponente.
De este modo, si se reporta el valor 10^3 (diez elevado al cubo) de cierta variable que se está midiendo, no se usará esa notación sino simplemente "mil" o 1.000, ya que 10 x 10 x 10 = 1.000

Veamos algunos ejemplos de notaciones que podrían utilizarse para expresar los diferentes números en un texto, sea con palabras y/o números más sencillos de entender para la audiencia.

 10^2 = cien = 100

 4 x 10^3 = cuatro mil = 4.000

 ¿Cuántos días son 994060 segundos? = once días y medio = 11,5 días
1 día = 24 horas = 1440 minutos = 86440 segundos.
Luego 994060 segundos con once días y medio.

 ¿Cuántos años son 103 millones de segundos? = 3,2 años
1 día = 86440 segundos; luego 103 millones de segundos son 1191,57 días; luego dividido entre 365 días (1 año) el resultado es 3,2 (o 3.2 según la notación de científica norteamericana)

 Si el prefijo kilo es el que se usa reemplazar al exponente 10^3 (diez al cubo =10 x 10 x 10) ¿cuál será el prefijo que usamos para 10 elevado a la secta potencia = Mega.

 Si el disco duro de mi PC tiene una capacidad de 80 Giga, ¿qué significa y qué exponente le correspondería? = Cada giga son 1024 megas. Los 80 Giga originales serían entonces 83.886.080 Kb (kilo bytes) o 85.899.345.920 Bytes. En notación científica aproximada, los 80 Giga serían 86 x 109. Pero como la idea es simplificar, y tomando en cuenta que un documento de Word de una página escrita ronda los 25 Kb y una foto de baja resolución insume 200 Kb, se podría expresar esos 80 giga como un disco con capacidad para almacenar más de tres millones de documentos de Word de una página, o 420 mil fotos de baja resolución.

Equivalencias

Philip Meyer (5) , el padre del periodismo de precisión escribió la memorable frase: “Si tu madre te dice que es tu madre, verifícalo”. Conceptualmente hablando es necesario recordar que:

- 1/ 2 (un medio) es una fracción: “La mitad de los policías estaba fuera de la zona de peligro” (Sería el 50 %)

- 25 % es un porcentaje: “El 25 % de los niños no va a la escuela” (o lo que es lo mismo también, se puede decir, la cuarta parte)

- 1 de cada 3 es una proporción: “1 de cada 3 mujeres es alérgica” (o la tercera parte, o aproximadamente el 33 %)

- 5 % es lo mismo que 1 de cada 20

- 10 %, es lo mismo que la décima parte o uno cada diez

- 20 % es lo mismo que la quinta parte o uno de cada cinco

- 25 % es lo mismo que la cuarta parte o lo que es igual a uno de cada cuatro

- 30 % es lo mismo que 3 de cada 10

- 33 % es lo mismo que la tercera parte o 1 de cada 3

- 40 % es lo mismo que 2 de cada 5

- 50 % es siempre la mitad

- 60 % es lo mismo que la tres quintas (3/5) parte o 3 de cada 5

- 66 % es lo mismo que las dos terceras partes (2/3) o 2 de cada 3

- 70 % es lo mismo que 7 cada 10

- 75 % es lo mismo que la tres cuarta parte (3/4) o 3 de cada 4

- 80 % es lo mismo que cuatro de cada cinco

- 90 % es lo mismo que 9 de cada 10

Partes por Millón

Partes por millón (abreviado como ppm) unidad empleada usualmente para valorar la presencia de elementos en pequeñas cantidades (traza) en una mezcla. Generalmente suele referirse a porcentajes en peso en el caso de sólidos y en volumen en el caso de gases. Se abrevia como ppm. También se puede definir como "la cantidad de materia contenida en una parte sobre un total de un millón de partes."

Ejemplo: Supongamos que tenemos un cubo homogéneo de un metro de arista, cuyo volumen es un metro cúbico (m^3). Si lo dividimos en 'cubitos' de un centímetro de lado, obtendríamos un millón de 'cubitos' de un centímetro cúbico,(cm^3 o cc). Si tomamos uno de esos 'cubitos', del millón total de 'cubitos', tendríamos una parte por millón.

El caso de los Años Luz

Los números abstractos algunas veces constituyen una buena manera de construir una imagen mental de nuestra propia experiencia.

Las personas pesamos tantas libras (o kilogramos), los árboles miden determinados pies (o metros) de altura y las ciudades se encuentran a ciertas millas (o kilómetros) de distancia.

Un año luz es la distancia que viaja un haz de luz en un año. A una velocidad promedio de 300.000 km/seg (186.000 millas/segundos), un año luz equivale a 9.461.000.000.000 kilómetros.

¿Cuánto es un billón de dólares?

Para todos los países, excepto los angloparlantes, un billón representa un millón de millones, es decir:

1.000.000 x 1.000.000 = 1.000.000.000.000

En los Estados Unidos e Inglaterra, un billón son mil millones:

1.000 x 1.000.000 = 1.000.000.000

A esta medida (la de los mil millones), en Francia se la comenzó a conocer como millardo

Por lo que:

- Un Millardo: nueve ceros

- Un Billón en USA e Inglaterra: nueve ceros

- Un Billón en América Latina y otros países: 12 ceros

En el Diccionario Hispánico de dudas se aclara que Millardo es la adaptación gráfica de la voz francesa milliard, „mil millones‟

"Los ingresos brutos se situaron en 1,1 millardos (1146 millones) de dólares".

Es voz de reciente incorporación al español, cuyo uso es recomendable para desterrar el empleo de la palabra billón con este sentido, calco rechazable del inglés americano y que puede dar lugar a peligrosas confusiones (v. billón).

Billón. Voz procedente del francés billion, „un millón de millones (1012)‟ . Es inaceptable su empleo en español con el sentido de „mil millones‟ , que es el que tiene la palabra billion en el inglés americano. Para este último sentido,
debe emplearse la voz millardo (v. millardo), procedente también del francés, o la equivalencia española mil millones.

Diccionario de la Real Academia:
Millardo (Del fr. milliard). 1. m. Econ. Mil millones. Un millardo de pesos.

La Real Academia acogió esta palabra en diciembre de 1995. La palabra millardo, no viene a resolver el problema del billón norteamericano, sino más bien a complicar las cosas. La duda no radica en utilizar mil millones o millardo, sino en saber cuándo billón significa 'mil millones', según el uso norteamericano, y cuándo 'un millón de millones', según el uso mayoritario en Europa.

Consejo:

- Evite el Millardo y el Billón.

- Millardo: Es mejor hablar de 1.000 millones.

MEDIDAS DE TENDENCIA CENTRAL

En una distribución de datos, éstos tienen a acumularse hacia el centro

Las medidas de tendencia central más comunes son:

a) **Media Arimética** (Valor Medio o Promedio)
b) **Mediana** (Valor Central)
c) **Moda** (Valor Más frecuente)

Promedio o Media Arimética

Es la sumatoria (Σ) de todos los casos, dividida por el número de casos.

P= $\Sigma\, n / N$

Σ = sumatoria

n: cada dato

N: cantidad datos totales

Se distorsiona si tiene valores extremos o no representativos del resto.
Veremos un ejemplo en el que el Promedio no debe ser utilizado por falta de representatividad. Supongamos que nos informan que en una oficina de la administración pública hay 7 categorías principales de salarios cuyos sueldos en dólares son: 8.000 - 5.000- 1.200 – 700 – 500 – 400 - 300.

Si se suman esos valores y se divide entre 7 (número de datos totales) se obtiene el promedio mensual de salarios para los empleados de esa oficina

P= 2.300 $

El promedio, en este caso no representa a ninguna de las categorías.

Cuando esto ocurre hay que calcular lo que se conoce como **Media Ponderada,** es decir, dividir por intervalos las distintas clases y considerar la cantidad de empleados en cada categoría.

Veamos un ejemplo sobre otra escala salarial más detallada: (Los números entre paréntesis corresponden a la cantidad de empleados)

(1) 8000 $
(9) 3200 $
(18) 2599$
(36) 2200$
(32) 1800$
(21) 1500$
(15) 1200$
(12) 1080$
(57) 850$
(86) 750$
(92) 560$
(105) 420$
(210) 360$

Clase 1: 8000 (1 funcionario) = No se promedia, ya que distorsionaría cualquier otro valor.

Clase 2: (incluye a quienes ganan valores entre 2.200 y 3.200)
P= 2.457 $ (63 funcionarios) = Lo que se hizo aquí fue (9 x 3.200) + (18 x 2.599) + (36 x 2.200) y a este resultado global se lo dividió entre 63 que es la suma de 9 + 18 + 36. El método consiste en ajustar cada categoría salarial al número de empleados cuya diferencia salarial no sea demasiado significativa considerando los extremos de la tabla.

Clase 3: (incluye a quienes ganan valores entre 1.800 y1.080)
P= 1.501 $ (80 funcionarios)

Clase 4: (incluye a quienes ganan valores entre 850 y 360)
P= 517 $ (550 funcionarios)

Mediana (ó Valor Central)

Es el valor que representa el punto central de una serie de datos. La mitad de los datos recogidos está por encima y la otra mitad por debajo. También se dice que el 50 % de los valores quedan por debajo de ese dato (la mediana o valor central) y el otro 50 % por encima.

Al igual que la Moda, que veremos a continuación, no tiene demasiada aplicación en las crónicas periodísticas ya que su uso no es popular, pero es importante que el reportero comprenda su significado porque en muchos reportes estadísticos pueden aparecer como variables y resulta **importante que no confundan este valor con el promedio.**

Para calcular la mediana, dada una serie de datos numéricos, lo que se hace es ordenar los datos de menor a mayor.

Ejemplo 1
7-10-10-**12-**13-15-17
Mediana: 12

Ejemplo 2
7-10-**10-12**-13-15
Se toma el P de los centrales internos. En este caso= 11
La mediana es 11.

A veces resulta difícil entender para qué sirve la mediana, y además, es de difícil aplicación en una crónica periodística. No obstante, es una medida de tendencia central y todo periodista debe conocer su significado y el procedimiento para calcularla, en especial cuando se tiene un conjunto de datos pares.

No recomiendo usar a la mediana (o valor central) en especial porque los lectores pueden confundirlo con el promedio. Y a veces el número coincide, pero generalmente no.

Se utiliza mucho en estadística básica, cuando tenemos muchos datos, que podemos ordenar de manera creciente (lo que nos dará una línea ascendente en el plano), y queremos decir que la mitad de los casos relevados cayeron por debajo de ese valor y la otra mitad está por encima de ese valor.

Por ejemplo en estudios clínicos, cuando un laboratorio quiere probar la eficacia de un nuevo medicamento y entrega un dossier de prensa. También cuando se analizan tablas de frecuencia. Por ejemplo, frecuencias de temperatura climática registradas a lo largo del año.

Hace unos años supe de un gobierno centroamericano que utilizaba este dato para crear confusión en los reportes de prensa que entregaba su ministerio de Economía. Esto ocurría cuando el promedio de cierta variable no daba un buen resultado a sus intereses y entonces incluían el valor de la mediana (que en apariencia los favorecía más) por lo que todos los periodistas creían que se trataba de un promedio.

Aunque ese no era el caso, debo aclarar que en muchos reportes estadísticos (como los médicos o económicos), se aconseja usar a la mediana y no a la media arimética o promedio, porque el conjunto de datos tiene valores muy extremos, que no son promediables, por lo que la media arimética se vería fuertemente distorsionada
La idea es que no confundan el concepto y así eviten una posible manipulación.

<u>Moda</u>
Es el valor más frecuente
No confundir con “mayoría”. A veces coincide pero otras no.

Ejemplo:
16-20-**16**-17-**16**-23-12
Moda: 16

El programa Excel calcula automáticamente estos valores, por eso es importante que el reportero maneje este programa informático, al cual nos referiremos durante las sesiones en los foros.

Medidas de Proporción

Un número solo no dice nada.

Para que adquiera significado es necesario efectuar una comparación. Por eso las medidas de proporción establecen básicamente comparaciones de X cantidad con respecto a Y cantidad

a) Razón
Es el cociente de una cantidad dividida por otra. Se la define como la principal operación de transformación o "normalización" estadística, esto es, definir una norma para expresar datos primarios. Para calcular una razón se divide la cantidad que se quiere "normalizar" (estudiar), por la cantidad "normalizadora" (referente).

Veamos un ejemplo sobre un dato censal: ¿Qué sería "razón de feminidad?

Si en un estudio, el total de la población son 300 personas, de las cuales 200 son mujeres y 100 son hombres, dicha razón es 200/100= 2

Significa: 2 mujeres por cada hombre

Hoy día Se está usando el término "razón" como ardid para enmascarar estadísticas en lugar de otras medidas de proporción, como el porcentaje, de mayores niveles de comprensión.

b) Índice

Es un tipo de medida que usa más de dos indicadores u observaciones para resumirlos en un resultado, relacionados con un mismo fenómeno.

Expresa la variación de un conjunto de valores. El más conocido es el Índice de Precios al Consumidor. Otro ejemplo, usado en muchos países es el IME (Índice Multivariado de Educación), resume seis factores del proceso educativo en un solo número.

Los índices se calculan por métodos estadísticos del análisis factorial EL IPC (Índice de Precios al Consumidor), reúne a 20 categorías de datos o más, conforme el país en el que se lo calcule.

c) Proporción

Es la frecuencia de casos en una categoría, dividida por el número de casos de todas las categorías. Se trata de la razón entre una parte y la totalidad.

Las proporciones varían siempre entre 0 y 1.
La suma de las proporciones siempre da 1

Ejemplo: Población total: 1560 personas (850 Mujeres y 710 Hombres). Proporción Mujeres (850/1560=0,54) y Proporción H: (710/1560=0,46).

Si multiplicamos por 100 el valor de una proporción lo que se obtiene es el porcentaje.

d) Tasa

De vital importancia para el reportero, **expresa la frecuencia de casos con relación a un número fijo que se toma como referencia.** Se refiere a números referentes (per cápita, cada 10.000, cada 100.000).
Se usa en crónicas periodísticas para comparar situaciones en distintas ciudades o países, porque permite usar el mismo número tomado como base, que es fijo e independiente de la población total.

Ejemplo:

- **Tasa de escolarización: número de escolares por cada 100.000 niños en edad escolar.**
- **Tasa de ocupación hospitalaria: Número de camas ocupadas en hospitales por cada 10.000 camas hospitalarias.**
- **Tasa de delito: Número de delitos denunciados por cada 10.000 habitantes** (pude calcularse sobre otra base, como 1.000, 100.000, etc.) La base depende del tamaño del escenario en estudio y/o de la convención que se haya decidido en el lugar de uso.

Las tasas se construyen de diferente manera conforme el país o ciudad en que se estudien.

Hay definiciones universales para algunos casos, pero en otros, como se explicó antes, la base es diferente; por eso es importante que el reportero, conozca el significado real de las tasas que se calculan en su país, como las que aparecen en el censo

Ejemplo:

Título periodístico: "Alarmante Ola de Robos en la Ciudad" (Bahía Blanca)
Pero resultó que eso no era cierto, ya que cada ciudad tiene una cantidad de habitantes diferentes.

Ciudad	**Delitos**	**Habitantes**	**Tasa x c/1.000 Hab.**
Bahía Blanca	536	284.313	
C. de Patagones	22	27.759	
Dorrego	42	16.469	
Pringles	12	23.765	
Rosales	39	60.879	
G. Chavez	29	11.967	
Monte Hermoso	14	5.603	
Puan	12	16.952	
Saavedra	16	19.751	
Tornquist	15	11.686	
Tres Arroyos	121	57.110	
Villarino	23	26.438	

Calculando la tasa de delitos por cada 1.000 habitantes resulta que Dorrego tiene un valor mayor al resto: 3 delitos por cada 1.000
Entonces: ¿Es el título el adecuado?

e) Porcentaje

Expresa una cantidad como un número de partes por cada 100 unidades.
Recuerde que, como ya dijimos, toda proporción puede ser transformada en %, pero no todo % puede ser transformado en proporción.

A diferencia de las proporciones, un porcentaje puede ser mayor de 100. (No confunda porcentaje puro con puntos porcentuales. Veremos esa diferencia en el foro de la semana)

MEDIDAS DE CAMBIO

Lo que no cambia no es noticia. Lo que cambia sí.
Toda variación implica un cambio y los cambios suelen contener noticias de relevancia.

Se calculan a partir del estudio de variables

Variable: Aquello que se modifica (o varía) conforme pasa el tiempo
Ejemplo: Partidas presupuestarias, accidentes, robos, clima, niños desnutridos, etc.
La medida de cambio más utilizada en periodismo es la VARIACION PORCENTUAL

Variación Porcentual:

El cálculo de variaciones porcentuales es la operación de máxima importancia en el análisis de tablas numéricas. Resulta vital que el reportero entienda cómo se calculan e interpretan

Por ejemplo: Supongamos que el cuadro muestra la evolución de la Deuda Externa, conforme aparece a continuación.

AÑO	DEUDA (en millones de dólares)	VARIACION NETA (en millones de dólares)
1991	58	-
1992	74	16
1993	192	118
1994	320	128
1995	415	95
1996	512	97
1997	640	128
1998	720	80
1999	960	240 (mayor valor neto)
2000	1080	120
2001	1280	200

La VARIACION NETA es la cantidad en millones de dólares, que se agrega cada año, a la deuda del año anterior. Es un número absoluto que se calcula mediante una simple resta. Para el caso del ejemplo, es la resta que se hace con el valor de cada año en estudio, respecto del valor del año anterior. Para el año 1991 no hay variación neta calculada, porque se desconoce el monto de la deuda del año anterior.

El uso de la variación neta, o números abstractos para expresar cambios es ALTAMENTE INADECUADO en periodismo, ya que no permiten las comparaciones.

Un valor puede decir mucho o poco, depende de qué valor tiene para esa misma variable en otra circunstancia.

Por ejemplo, si un candidato a presidente obtiene en una encuesta el 42 % de intención de votos, ese número puede ser mucho o poco; depende del resultado obtenido en la medición anterior.

Si una encuesta previa arrojó el resultado de 61 % de intención de votos el valor menor tiene un significado: la aprobación popular bajó y es un dato malo para el candidato y su partido. Ahora, si una encuesta previa había medido 17 % de intención de voto para ese mismo candidato y en la encuesta posterior midió 42 % entonces el significado es otro.

Otro ejemplo es el de la cantidad de delitos administrativos (por citar una variable). Si se cometen 200 delitos de este tipo en un año, en un país con 120 millones de habitantes, eso tiene un significado diferente a si se cometen 200 delitos en el mismo año, pero en un país con 5 millones de habitantes. Es claro que la tasa de delitos contra la administración pública por cada 10.000 habitantes será mayor en el segundo país que en el primero.

Retomando el ejemplo de la tabla, el observador podría pensar que el año en que mayor aumento de la deuda externa hubo fue 1999, ya que ese año el incremento en millones de dólares fue de 240, pero no es la forma correcta de analizar la evolución de una variable, ya que para cada cálculo, no se toma un número fijo como referente. Todas las bases de referencia en cada resta que se hace para cada año, son distintas.
Por eso, reitero, es importantísimo usar medidas de cambio para expresar variables que están o pueden cambiar a lo largo del tiempo. **La más usada es la variación porcentual.**

<u>¿Cómo SE CALCULA LA VARIACION PORCENTUAL?</u>

Volvamos al caso del ejemplo anterior; si en vez de calcular la variación neta, hubiéramos calculado la variación porcentual para cada año respecto del año inmediato anterior, se habría caído en la cuenta que **el año de mayor endeudamiento fue 1993. Ahí había una noticia escondida en la tabla.**

Sólo había que hacer algunos cálculos para descubrirla, ya que no saltaba a simple vista viendo los números puros.

Veamos cómo calcular la variación porcentual de la deuda de 1992 respecto de 1991.
74 millones – 58 millones = 16 millones (deuda agregada en 1992 respecto de 1991)

Si 58 millones ……………........... 100%
16 millones (la diferencia) .... X= ¿??

X= (16 x 100) / 58 = 28%

Hagamos todos los cálculos y volvamos a la tabla anterior, esta vez calculando todas las variaciones porcentuales

AÑO	DEUDA (en millones de dólares)	VARIACION NETA (en millones de dólares)	VARIACION PORCENTUAL
1991	58	-	
1992	74	16	28 %
1993	192	118	159% (mayor endeudamiento)⚜
1994	320	128	66%⚜
1995	415	95	30%⚜
1996	512	97	23%⚜
1997	640	128	25%⚜
1998	720	80	12,5%⚜
1999	960	240 (mayor valor neto)	33%
2000	1080	120	12,5%⚜
2001	1280	200	18,5%⚜

Si el resultado da negativo (porque el Valor Final es menor que el Valor Inicial, la variación porcentual es negativa (signo -).
En un caso así no hay incremento, sino decrecimiento (caso caída de la bolsa = bajó cuatro puntos, significa que la variación porcentual fue negativa).

Puntos Porcentuales

"El candidato A medía la semana pasada 10 % de intención de voto. Hoy mide 12 %."

El reportero escribió que la intención de voto del candidato aumentó 2 % (dos por ciento)

¿Es correcto?

NO = La variación **neta** fue de 2 puntos porcentuales.

Pero la variación **porcentual** fue mayor (¿de cuánto?)

<u>**Tantas veces más**</u>

- ✶1991: se adeudaban 58 millones
- ✶2001: se adeudan 1280 millones

El reportero escribe: "Ahora se debe 22 veces más dinero que hace diez años". Falso.

División: 1280/58: 22 (Pero el 22 contiene la base, o lo que se adeudaba en el primer año tomado en estudio)
Correcto: "21 veces más que hace 10 años"

Otro ejemplo:
El asesino tiene 20 años
La víctima tiene 60 años

El reportero escribió: "La víctima es tres veces más vieja que el asesino". Falso. Si fuera tres veces más vieja tendría 80. En este caso es dos veces.

<u>Interpolación de datos Externos</u>
Mientras una variable va creciendo o decreciendo, hay otras variables que también sufren modificaciones permanentes e impactan directamente en la que estamos estudiando.

Por ejemplo, si en una ciudad X la cantidad de accidentes de tránsito es de 120 por día y en otra ciudad Y es de 190 por día, estos números no se pueden considerar aislados ya que la cantidad de habitantes de una ciudad difiere de la otra

Veamos un ejemplo aplicado a sueldos de empleados públicos

En 1980: ganaban en promedio 11.133 $ anuales
En 2008: ganan en promedio 19.000 $ anuales

Ahora ganan más: Falso

Hay que considerar el Índice de Precios al Consumidor: IPC

Datos IPC
1980 = 38.8
2008 = 90,9

<u>Haciendo cuentas:</u>
38,8 (IPC 1980) ---------- 11.133$ (el sueldo en 1980, con 38,8 de IPC)
90,9 (IPC 2008) -------- x = 26.082,21$ **(SUELDO IDEAL)**

SUELDO IDEAL = EL QUE DEBERÍAN GANAR PARA MANTENER EL MISMO PODER ADQUISITIVO

La conclusión es que si ahora ganan 19.000, entonces tienen menor poder adquisitivo)

<u>¿Cómo calcular la Caída del Salario?</u>
Caída del Salario: Sueldo Ideal – Sueldo Real
Caída del Salario: 26.082,21 $ - 19.000 $ = 7.082,21 $

<u>¿Cómo se calcula la Caída del Salario Porcentual?</u>
26,082,21 $ (salario ideal) ------------- 100%
7.082,21 $ (caída del salario) --------- x = 27 %
El **salario cayó un 27 %**

Recuperación Porcentual
Con este concepto se comete uno de los errores más comunes = Caso Bolsa Dow Jones (DJ)

Día 1: El Índice DJ cerró con 1759,89 puntos
Día 2: El Índice DJ cerró con 1569,26 puntos

La diferencia entre el día 2 y el día 1 es de = 190,63 puntos
Conclusión: Perdió el 10,83 %.
¿Qué porcentaje tenía que tener el Día 3 para recuperarse? ¿10,83?
NO. La base es otra. En puntos es 190.63, pero el % es diferente.

Haciendo cuentas:
1569,26........100%
190,63..........**x= 12,14%**

MEDIDAS DE DISPERSIÓN

Las medidas de dispersión nos informan cuánto una variable se alejó de lo esperado. Lo esperado está directamente relacionado con el promedio.

¿Por qué un avión que cae es noticia? Porque lo esperado es que eso no ocurra ya que el promedio de aviones que se caen es muy bajo respecto de la cantidad de aviones que vuelan.

Las medidas de dispersión más frecuentemente usadas con la **Varianza,** la **Covarianza** y la **Desviación Estándar.**

Varianza:
Es un indicador de cambio. Mide la dispersión calculada, respecto de la media arimética (promedio)de una serie de datos. Nos indica cuánto una medición se aleja de lo esperado

De este modo, la varianza se convierte en noticia.

La varianza siempre será mayor que cero. Mientras más se aproxima a cero, más concentrados están los valores de una serie de datos, alrededor de la media. Por el contrario, mientras mayor sea la varianza, más dispersos están.

No es intención de esta clase introducirlos en el cálculo de la varianza, pero sí que comprendan su significado ya que frecuentemente, en censos o reportes estadísticos pueden encontrar este tipo de terminología.

En estadística, el **análisis de varianza** (**ANOVA)** sirve para comparar si los valores de un conjunto de datos numéricos son significativamente distintos a los valores de otro o más conjuntos de datos.

En conclusión, cuando observen una varianza alta con relación al promedio de una serie de valores, deténganse en ese dato y busquen más información consultando fuentes académicas, ya que probablemente detrás de la varianza puede estar escondida una noticia.

Ejemplo de variable con varianza baja: Enfermeras trabajando en hospitales públicos menores de 40 años. (El promedio de la edad de las enfermeras es de 35; luego, las de 40 tienen una varianza baja; las de 60, tendrán una varianza más alta)

Covarianza

Cuando dos fenómenos varían al mismo tiempo, se dice que "covarían"
Por ejemplo, todos los años varía el número de nacimiento de niños prematuros, pero al mismo tiempo también varía la condición de fumadoras de las madres que los dan a luz.

Hay dos posibilidades:
1) Un fenómeno depende del otro
2) Uno explica al otro

Podemos realizar investigaciones para tratar de encontrar respuestas a los diferentes interrogantes que nos plantean las estadísticas.

Desviación Estandar (DS)

Muchas veces la Varianza no aparece en un reporte estadístico, pero sí la DS. Es una forma más común de mencionar a la Varianza ya que la DS está definida como la raíz cuadrada de la varianza.

¿Cómo saber si varían mucho o si la desviación es grande?

Siempre compararlos con la media

Ejemplo: número de crímenes cometidos en distintas ciudades. (Variación entre una y otra).

Recomendaciones

Mire los números con detenimiento

✶Vea qué variables está analizando
✶Analice qué posibles cruces puede hacer en las tablas
✶Revise sus procedimientos y cálculos
✶Para visualizar mejor los datos de una tabla e interpretar mejor su contenido, grafique sus conclusiones
✶No pierda perspectiva: Mire hacia atrás y hacia delante.