Estado constructivo de las viviendas del sector norte, Municipio de Corque.
Por: Lic. Flores

Casa	Estado constructivo
1.	R
2.	B
3.	M
4.	|B
5.	R
6.	M
7.	M
8.	B
9.	B
10.	R

Cuando el funcionario reciba el solicitado informe, puede que se extrañe un tanto. ¿Por qué? Mira, no es que esté mal lo que le entregaste, la información se ajusta a lo pedido, pero hay algo que no facilita las cosas. ¡Claro! ¡La presentación de los datos! Juzga por ti mismo la diferencia:

Estado constructivo de las casas del sector norte Municipio de Corque.
Por: Lic. Flores

Estado constructivo	Casas
Bueno:	4
Regular:	3
Malo:	3

¿No crees que ganaste con esta forma de presentar los datos? Por supuesto, sobre todo porque al resumir la información y exponerla de forma asequible, estás garantizando en gran medida que un lector abrumado por mucho trabajo dedique un minuto a tu informe, y no lo destine a la gaveta del olvido, entre otras cosas.

Esto que acabas de hacer, o sea, agrupar los datos por frecuencia de aparición, de acuerdo con la escala que más se ajusta a tus necesidades, es lo que se denomina distribución de frecuencias, que consiste en:

Una **distribución de frecuencias** es el modo en que se distribuyen las unidades de análisis entre las clases o categorías que conforman la escala de clasificación de la variable en cuestión.

Las distribuciones de frecuencias pueden clasificarse en:

Distribuciones de frecuencias
- Absoluta
- Relativa
- Acumulada
    - Absoluta
    - Relativa

Las dos primeras se utilizan cuando se tratan variables en cualquier escala, en tanto que las acumuladas se emplean cuando se estudian variables en una escala cuantitativa o dimensional.

La frecuencia absoluta es el resultado de contar los casos u observaciones (o sea, el número de observaciones) que corresponden a cada una de las clases o categorías de la escala de clasificación.

La frecuencia relativa es la importancia o peso relativos que tienen las unidades de análisis de una categoría o clase sobre el total de las unidades. Se calcula dividiendo la frecuencia absoluta de la clase en cuestión por el total de observaciones, en cuyo caso obtendrás una proporción. Si multiplicas este resultado por 100, obtendrás un porcentaje.

Las frecuencias acumuladas son las frecuencias absolutas o relativas que se acumulan hasta un intervalo de clase dado. Se calculan sumando las frecuencias (absolutas o relativas, en dependencia de la que necesites) hasta la clase deseada. La frecuencia acumulada para el último IC será el total de observaciones, si se tratare de la frecuencia absoluta acumulada; y si fuera el caso de la relativa, entonces será 1 ó 100, en dependencia de si usaste proporción o porcentaje, respectivamente.

A continuación, te presentamos mediante un ejemplo cada tipo de distribución de frecuencias.

En un estudio sobre riesgos profesionales en los trabajadores de la industria del cromo en 2010, se encontraron los siguientes datos en una muestra de 897 trabajadores:

Tiempo de trabajo en la industria (años)	**Número (frecuencia absoluta)**	**Frec. relativa**		**Frec. acumulada**		
		Prop.	%	Abs.	Prop.	%
< 1 año	41	0.04	4.0	41	0.04	4.0
1 – 3	115	0.13	13.0	156	0.17	17.0
4 – 10	304	0.34	34.0	460	0.51	51.0
> 10	437	0.49	49.0	897	1.000	100.0
Total	897	1.00	100.0	—	—	—

Si observas detenidamente, te podrás dar cuenta que:

- La suma de las frecuencias absolutas es el total de la muestra o población estudiada.
- La suma de las frecuencias relativas es igual a 1 o a 100, en dependencia de que se trate de una proporción o de un porcentaje.
- En las frecuencias acumuladas la última clase o categoría de la variable suma el total, la unidad o el ciento por ciento.

Puede que la suma de los porcentajes no sea 100 exactamente debido a las imprecisiones de la aproximación, de ahí que se admita hasta 99.9 como total; pero debes evitar cifras inferiores ya que denotan errores de cálculo.

3.6. Computo de datos

Una vez definidas las categorías, intervalos de clase o clases según el caso, debe determinarse cuantas del conjunto total de observaciones corresponden a cada una de las clases y esto requiere de un proceso de contar o computar que puede realizarse directamente en la relación donde se tienen registrados los datos, con ayuda de tarjetas en blanco, tarjetas con perforaciones marginales, palotes o por procedimientos electrónicos, según el número de datos que se hayan obtenido en cada elemento de estudiado y el número de elementos que integren el conjunto.

3.6.1. Tipos de cómputo

Existen varios métodos de cómputo y la elección de uno con preferencia a los demás, depende del número de observaciones que se estudia, de la complejidad del análisis que se intenta realizar y de los recursos económicos con que se cuenta. Así tenemos:

3.6.1.1. Método de las listas

Frecuentemente los resultados de estudio se resumen en una larga lista, en la cual se destina una línea para anotar las características o variables correspondientes a cada observación. En tales casos, el computo se concretará a buscar cuales observaciones presentan determinada característica y a contarlos mentalmente marcándolos con un signo convencional Y o X, con el fin de facilitar la verificación al final. Este método se utiliza cuando son pocas las unidades que se estudian y siempre que no se pretenda clasificarlas por más de 2 escalas a la vez. En la tabla siguiente se presente un ejemplo.

Relación de pacientes y sus características a ser computadas.

N°	Edad	Sexo	Estado civil	Procedencia	Peso (kg)	Estatura (cm)	Enfermedad
1	25	F	Casado	Potosí	63	166	Tuberculosis
2	27	M	Soltero	Oruro	78	185	Neumonía
3	26	F	Soltera	Sucre	55	155	Bronquitis
4	34	M	Casada	Tarija	72	173	Neumonía
5	22	F	Casada	Pando	52	152	Sinusitis

3.6.1.2. Método de los palotes

Consiste en registrar en una hoja de trabajo con un palote (I, III, III, etc.) por cada unidad que se cuenta, destacándose cada quinta raya cruzada las cuatro anteriores para formar un grupo compacto de cinco rayas que corresponderá a cinco observaciones. El cómputo se puede realizarse a partir de los formularios originales en los cuales se recogió la información o a partir de una lista como la anterior que la resume. Ejemplo del método de los palotes.

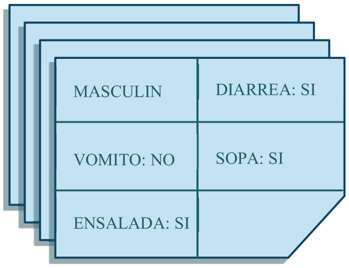
Edad	Palotes	F_i
5 – 9	IIII	4
10 – 14	~~IIII~~ ~~IIII~~ ~~IIII~~ II	17
15 – 19	~~IIII~~ ~~IIII~~ ~~IIII~~ ~~IIII~~ ~~IIII~~	25

Total	46

Este método no es aplicable en un gran número de observaciones y varias características a la vez.

3.6.1.3. Método de las tarjetas simples

Cuando el número de unidades que se estudia no es muy grande como en el caso de las investigaciones médicas habituales, las tarjetas simples son aplicables, aun tratándose de complejas computaciones. El método consiste en pasar los datos que se estudia a tarjetas de cartulina (una tarjeta para cada caso), cuyo tamaño depende del número de datos analizados, aunque un tamaño como el de las barajas, las hace más fácil manejarlas. Las tarjetas se dividen en varias casillas, de acuerdo a las variables de estudio y cada casilla se destinará a la inscripción de un dato. Por ejemplo, en un estudio sobre mortalidad en donde se hubiera investigado los siguientes datos: edad, sexo, estado civil, residencia, nacionalidad, atención médica, causa de muerte, la tarjeta podría ser de la siguiente forma:



Nº de caso ……………….	Edad ……………….	Sexo ……………….	Estado civil ……………….
Residencia ………………	Nacionalidad ……………….	Atención medica ………………	Causa de muerte ……………….

Se puede utilizar por ejemplo cuantas mujeres casadas de nacionalidad extranjera murieron por causa de la tuberculosis se procedería de la siguiente manera:

a) Separar las tarjetas por sexo.
b) Dividir el grupo de tarjetas correspondientes a las mujeres de acuerdo al estado civil.
c) El grupo de casadas se dividirá, en bolivianas y extranjeras.
d) En este último grupo se separan y contaran las muertes por tuberculosis.

3.6.1.4. Método de las tarjetas con perforaciones marginales

El material utilizado en este método consiste en: Tarjetas con perforaciones marginales, con un alicate, sacabocados, punzón, etc. Son tarjetas de tamaño variable con una serie de orificios en sus bordes, a cada uno de los cuales se le asigna la representación de una de las características o variables que se estudian. Los datos que van a inscribirse en la tarjeta pueden encontrarse en formularios especiales o pueden recogerse directamente en ella. Se usa una tarjeta para cada unidad o individuo.

Para efectuar la clasificación se asigna a cada perforación una modalidad de las características en estudio o, en su caso se procede a la codificación, que simplemente una clave mediante la cual cada dato se designa por un número.

3.6.1.5. Método de las computadoras digitales

Cuando el número de elementos y de observaciones es grande, hay necesidad de acudir a las computadoras digitales, que no son otra cosa que aparatos electrónicos que tienen la capacidad de recibir grandes cantidades de datos que permite computar con gran rapidez.

Los sistemas consisten en introducir datos y darles cierto tratamiento para obtener el producto. Entre ellos podemos citar softwares estadísticos como el Excel, SPSS, STATA, EPIDAT y otros.

Resumen

En este tema aprendiste que:

1. La elaboración de la información consiste en transformar los datos brutos en revisados, clasificados, contados y presentados para su análisis e interpretación.
2. Los pasos a seguir para la elaboración de la información: revisión, clasificación, computo, presentación y análisis de los datos.
3. La revisión de la información es llamada como control de calidad de la información las mismas tiene que ser integras, exactos y oportunos.
4. La clasificación de la información se divide en cualitativa y cuantitativa.
5. El computo de datos es un conjunto total de observaciones que requieren un proceso de contar o computar, con la ayuda del método de lista, tarjetas en blanco, tarjetas con perforaciones marginales, palotes, método de las computadoras digitales.
6. La presentación de la información se la realiza en tablas o cuadros y los gráficos estadísticos.

Ejercitación

1. La dirección de un Hospital desea estudiar la nutrición de los adolescentes del área. Una de las variables investigadas es la talla (en centímetros). A continuación, te presentamos la talla de 20 jóvenes del estudio mencionado tomados al azar.

1.	168	**6.**	170	**11.**	160	**16.**	175
2.	145	**7.**	167	**12.**	159	**17.**	161
3.	162	**8.**	163	**13.**	158	**18.**	152
4.	155	**9.**	165	**14.**	174	**19.**	147
5.	155	**10.**	165	**15.**	171	**20.**	148

Construye una escala con intervalos de igual amplitud para representar esta información, te sugerimos que utilices seis intervalos.

a) Distribuye la información a través de frecuencias absolutas, relativas y acumuladas.

2. A continuación, te presentamos la variable temperatura (en grados Celsius) clasificada en una escala de intervalos:

Temperatura
34 - 36
37 – 39
40 – 42

Determina la amplitud del segundo intervalo de clase.

a) ¿Cuál es la marca de clase del primer intervalo de clase?

b) ¿Cuál es el límite de clase inferior del tercer intervalo de clase?

Unidad didáctica N°4

REPRESENTACION ESTADISTICA

Competencia de la unidad:

Elabora e interpreta correctamente las tablas y gráficos estadísticos a partir de un conjunto de datos de la forma más adecuada.

4.1. Introducción

Cuando investigamos, obtenemos información sobre las variables del estudio. Sin embargo, por lo general es tanto el volumen de información que poseemos, que se necesita utilizar alguna técnica que permita presentarla de forma resumida.

A continuación, estudiarás las maneras de representar la información: la tabla o cuadro estadístico y los gráficos.

Te sugerimos que prestes mucha atención a este tema, pues a pesar de su sencillez, estos conocimientos se usan con bastante frecuencia, y los desatinos que usualmente se cometen hacen ineficaces el uso de estos recursos que te brinda la Estadística.

4.2. Tabla o cuadro estadístico

Es conveniente que sepas, en principio, qué es una *tabla o cuadro estadístico*:

Un cuadro estadístico es un recurso que emplea la Estadística con el fin de presentar información resumida, organizada por filas y columnas.

El cuadro estadístico tiene la finalidad de representar distribuciones de frecuencias, medidas de resúmenes y series cronológicas.

4.2.1. Partes de la tabla o cuadro estadístico

Varias son los elementos que integran una tabla estadística. Seguidamente te presentamos cada uno de ellos:

Partes de la tabla o cuadro

- Presentación (Identificación y Título)
- Cuerpo de la tabla
- Fuente
- Notas explicativas

A continuación, te presentamos cómo estos componentes de una tabla deberían ser mostrados.

1. *Identificación:* Consiste en otorgar un orden consecutivo a las tablas según orden de aparición en el texto, comenzando por el número uno, v.g. Tabla[7] 1, Tabla 2, etc.

[7] Indistintamente Cuadro, si lo deseas, pero no debieras usarlos alternativamente en un mismo trabajo: uno o el otro.

2. *Título:* Explica brevemente el contenido de la tabla o cuadro, debe ser completo y conciso. Para ser completo, el título debe responder a las preguntas qué, cómo, dónde y cuándo. Reconozcamos en un ejemplo cada una de estas preguntas:

Tabla 1. Distribución de fallecidos según grupos de edad y sexo. Municipio de Zudáñez, 2020.

Un análisis del título anterior permite conocer que:

- Distribución de fallecidos es de qué trata la tabla.
- Los grupos de edad y sexo son el cómo se midió, es decir, a través de cuales variables.
- Municipio de Zudáñez es dónde se realizó el estudio.
- 2020 es cuándo se realizó el estudio.

Observa que las variables se presentan después del vocablo "según", aunque alternativamente puedes usar el término "por".

También es conveniente decirte que, en ocasiones, no es necesario dar respuestas a estas cuatro interrogantes en un título, en cuyo caso sólo deberás responder al qué y al cómo. Ello ocurre cuando estás representando información obtenida en una investigación y, en el informe final o el artículo, ya has consignado en algún apartado anterior al de Resultados, dónde y cuándo se realizó el estudio.

El otro elemento, la concisión, consiste en escribir justamente lo necesario. Elimina las preposiciones y artículos que no ayuden a la comprensión del título de tu cuadro. No obstante, no sacrifiques el mensaje redactando un título "telegráfico", no seas lacónico, pues dificultarías la comprensión de lo que quieres decir.

3. *El cuerpo de la Tabla:* Contiene los datos que constituyen el mensaje de la tabla, es el cuadro en sí mismo, formado por espacios llamados celdas, las cuales se vertebran en filas y columnas, por ejemplo:

Columna matriz (Variable)	**Encabezado de columnas**	**Total**
Encabezado o filas	Datos	
Total		

- La columna matriz: se utiliza para consignar la variable con su escala de clasificación. En caso de que el cuadro represente más de una variable, por la columna matriz representarás la que tenga más clases o categorías o la que constituye la causa, en estudios de causalidad.
- Encabezado de las columnas: describe el contenido de las columnas, se presentan las distribuciones de frecuencias, las medidas de resúmenes o la otra variable.

- Encabezado de las filas: describe el contenido de las filas.
  La última fila y la columna: se dedican a los totales.

4.2.2. Tipos de Tabla o cuadro estadístico

Los cuadros o tablas estadísticas suelen clasificarse según el número de variables que representan en:

Tablas	Unidimensionales: una variable. Bidimensionales: dos variables. Multidimensionales: tres o más variables.

Deben ser autoexplicativas, o sea, que se expliquen por sí mismas, por lo que debes evitar presentar demasiada información en ellas en aras de ganar claridad. En general, como forma de presentación se utilizan cuadros uni y bidimensionales, reservándose el uso de los multidimensionales para fines de trabajo.

4. *Fuente.* Se refiere al documento[8] de donde se extrajo la información presentada. Por lo general, las fuentes de información se clasifican en:

Fuente	Primaria: aquella de la que el investigador obtiene directamente la información utilizando diversas técnicas y métodos, v.g. la encuesta. Secundaria: aquella que existe independientemente del estudio y el investigador sólo la utiliza, v.g. el Registro de Nacimientos, las historias clínicas.

Resulta válido y oportuno aclarar que en la tabla sólo consignarás las fuentes secundarias.

Recuerda que una fuente es un documento. Frecuentemente esto se olvida, y se consignan erróneamente como fuentes algunos locales, departamentos, entre otros, como la Oficina Nacional de Estadísticas, el Archivo del policlínico, etc.

5. *Notas explicativas o aclaratorias***.** Se utilizan cuando se desea aclarar algo, por lo general del título o del cuerpo de la tabla.

4.2.3. Errores más frecuentes

No son pocos los errores que se cometen, voluntaria o involuntariamente, en la confección de los cuadros estadísticos. A continuación, te presentamos una lista de los más comunes, que podrás revisar si no deseas incurrir en los mismos.

a) Errores en la presentación.
   - Cuadros sin identificación.

[8] Cuando decimos documento no hacemos distinción de su soporte, puede estar impreso, en formato digital, etc.

- Título o encabezamiento incorrecto o inadecuado:
  -Telegráfico: título demasiado pequeño, carente de claridad.
  -Ampuloso: título demasiado extenso, que incluye vocablos que no aportan nada a la claridad del texto.

b) Errores del cuerpo.
- Errores de cálculo.
- Disposición incorrecta de los datos.
- Mostrar solamente medidas relativas (frecuentemente porcentajes) u otras medidas de resumen.
- Cuadros sobrecargados.

c) Errores en la fuente.
- No citar la fuente cuando es secundaria.
- Citar la fuente cuando es primaria.
- Consignar como fuente aquello que no es un documento (oficinas, departamentos, centros, etc.)

4.2.4. Interpretación de una tabla cuadro estadístico

Realmente parece algo tan trivial, que muchas personas lo pasan por alto muchas veces, lo que puede conllevar a interpretaciones erróneas de la tabla o en el peor de los casos, a no entenderla. Para que te evites el fiasco, echa una ojeada al orden propuesto:

a) Lee cuidadosamente el título, así sabrás de qué trata el cuadro exactamente.
b) Lee las notas explicativas. Obviamente, las mismas mejoran considerablemente la comprensión de la tabla.
c) Infórmate de las unidades de medida utilizadas.
d) Fíjate en el promedio total o porcentaje general del grupo.
e) Relaciona el promedio total con el porcentaje de cada una de las variables estudiadas.
f) Relaciona los promedios o porcentajes de las variables estudiadas.

4.3. Los gráficos estadísticos

Otra manera de presentar la información estadística es a través de los gráficos. Ellos pueden resultar muy útiles, aunque en ocasiones un uso incorrecto los convierte en instrumentos estériles. Son complemento de las tablas, por ende, deben ser más autoexplicativos que ellas.

4.3.1. Características generales

Generalmente se inscriben en los ejes de coordenadas cartesianas o ejes rectangulares, los cuales:

- Deben poseer la misma longitud, aceptándose como máximo que el eje X exceda hasta 1.5 veces al eje Y. Esto evita la introducción de falacias.

- Deben estar rotulados. Por el eje X se presenta(n) la(s) variable(s) con su escala de clasificación; en el eje Y, la distribución de frecuencias o medida de resumen utilizada.
- De ser posible, el origen de los ejes debe ser en el punto (0,0).
- Deben utilizarse números redondos.
- Debe evitarse el exceso de divisiones de los ejes.

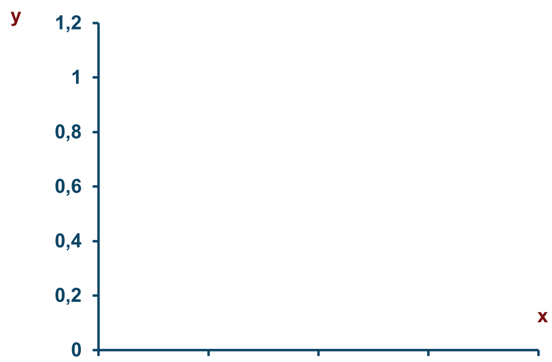


Figura: Ejes cartesianos o rectangulares

En la actualidad, con el advenimiento de las nuevas tecnologías informáticas, han proliferado los softwares que permiten la construcción de gráficos estadísticos. Al utilizarlos, debes tomar la precaución de analizar cuidadosamente el tipo de información que quieres representar, pues la mayoría de ellos ofrece varias posibilidades de representación, quedando a tu juicio escoger la más apropiada.

4.3.2. Partes del gráfico

Todo gráfico estadístico está constituido por varios elementos, los cuales te mencionamos a continuación.

Partes del gráfico

- Presentación (Identificación y Título)
- Gráfico propiamente dicho
- Leyenda
- Fuente
- Notas explicativas

1. *Identificación*: consiste en numerar los gráficos consecutivamente, por ejemplo: Gráfico 1, Gráfico 2, grafico 3, etc.
2. *Título*: Texto descriptivo del gráfico que se coloca en la parte superior, debes establecer el de la tabla que lo originó.
3. *Gráfico propiamente dicho:* verás los distintos tipos de gráficos en el epígrafe siguiente.
    - Área del gráfico: Esta es el área que se encuentra definida por el marco del gráfico y que incluye todas sus partes.
    - Series de datos: Son los puntos de datos relacionados entre sí trazados en un gráfico. Cada serie de datos tiene un color exclusivo. Un gráfico puede tener una o más series de datos a

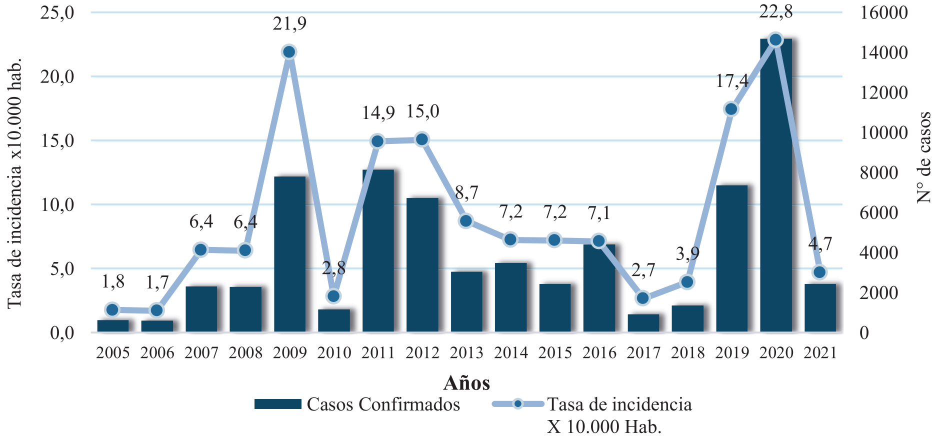
excepción de los gráficos circulares que solamente pueden tener una serie de datos.

- Ejes: Un eje es la línea que sirve como referencia de medida. El eje Y es conocido como el eje vertical y generalmente contiene datos. El eje X es conocido también como el eje horizontal y suele contener las categorías del gráfico.
- Líneas de división: Son líneas opcionales que extienden los valores de los ejes de manera que faciliten su lectura e interpretación.
- Título de eje: Texto descriptivo que se alinea automáticamente al eje correspondiente.

4. *Leyenda:* Recuadro que ayuda a identificar los colores asignados a las series de datos, su fin es identificar los elementos del gráfico (barras, sectores, etc.) con su correspondiente origen.
5. *Fuente:* la tabla que lo originó.
6. *Notas explicativas:* su uso es similar a lo descrito en las tablas.

Ejemplo:

Gráfico 1 Incidencia de dengue según número de casos positivos, Bolivia, 2005 a 2021.



Fuente: Tabla x

Nota explicativa: En grafico se puede apreciar la tasa de incidencia alta en el año 2009 con 22 casos positivos por cada 10.000 habitantes aproximadamente., en la gestión 2019 con 17 casos positivos, en el año 2020 fueron reportados 23 casos positivos por 10.000 habitantes y un decremento durante los años 2005 y 2006 con 2 casos por x 10.000 habitantes respectivamente.

Durante los últimos 17 años, el dengue en Bolivia ha mostrado un comportamiento con un patrón cíclico, que muestra que en el año 2020 se registró la mayor epidemia de dengue en el país con 14.679 casos confirmados.

4.3.3. Gráficos para representar variables en escalas cualitativa y cuantitativa discreta.

A continuación, te presentamos un grupo de gráficos que se estudiarán en este epígrafe, atendiendo al número de variables que representan.

Variables	Gráfico
1	Barras simples, Pastel
2	Barras múltiples y Barras compuestas

4.3.3.1. Gráfico de barras simples

Uso: Es un gráfico formado por barras separadas que representan a las categorías de la variable en estudio. Se utiliza cuando queremos representar una variable cualitativa o cuantitativa discreta, y la información se dispone en frecuencias absolutas o relativas, o en medidas de resumen.

La orientación del gráfico puede ser:

Vertical: las distintas categorías están situadas en el eje horizontal y las barras de frecuencias crecen verticalmente.

Horizontal: las categorías se sitúan en el eje vertical y las barras crecen horizontalmente. Suelen usarse cuando hay muchas categorías o sus nombres son demasiado largos.

Las categorías pueden ordenarse alfabéticamente facilitando su búsqueda o por sus frecuencias facilitando la comparación de los datos.

Elementos a considerar en su construcción

1. Dispón las barras separadas entre sí, para dar la idea de discontinuidad de la variable representada.
2. El ancho de las barras será opcional, pero debe ser el mismo para todas.
3. La separación entre barras debe ser igual a la mitad del ancho de ellas.
4. Si la variable es nominal, ordena las barras en orden creciente o decreciente, en dependencia de tus gustos.
5. Utiliza tantas barras como categorías tenga la variable.
6. Puedes colocar las barras en el eje vertical o en el horizontal. Comúnmente se utiliza el eje horizontal.
7. Este gráfico se origina a partir de tablas unidimensionales.

Ejemplo: Un grupo de investigadores desea conocer el comportamiento de la vulnerabilidad psicosocial en ancianos de un área de salud. Para ello aplica el cuestionario de vulnerabilidad-bienestar psicosocial del Dr. R. Pérez y obtiene los siguientes resultados:

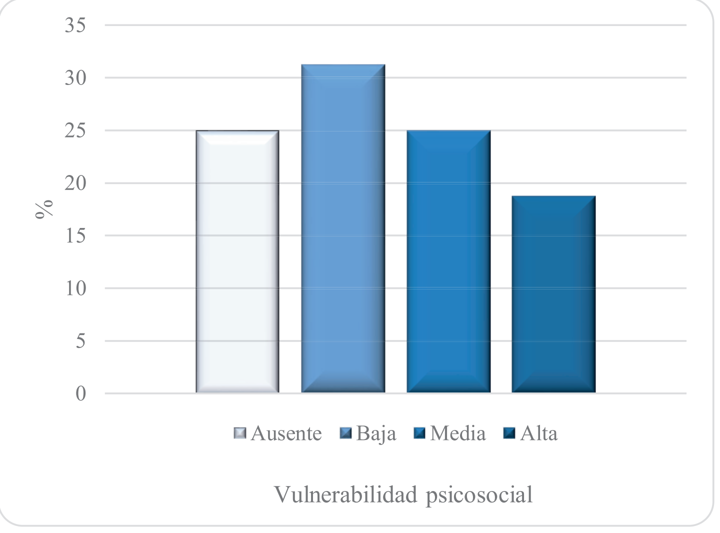
Tabla 1. Distribución de ancianos según vulnerabilidad psicosocial. Municipio de Pucarani, 2019.

Vulnerabilidad	Número	Porcentaje

Psicosocial		
Ausente	20	25.00
Baja	25	31.25
Media	20	25.00
Alta	15	18.75
Total	**80**	**100.00**

Fuente: Registros C.S. Pucarani

Gráfico 1. Distribución de ancianos según vulnerabilidad psicosocial. Municipio de Pucarani, 2019.



Fuente: Tabla 1

4.3.3.2. Gráfico de pastel, de sectores o circular

Un gráfico de sectores es una representación circular de las frecuencias relativas de una variable cualitativa o discreta que permite, de una manera sencilla y rápida, su comparación.

Uso: Este gráfico se utiliza cuando queremos representar una variable cualitativa o cuantitativa discreta, y la información se dispone en porcentaje. Básicamente, es un círculo dividido en sectores que representan las categorías de la variable.

Son útiles cuando las categorías son pocas. Si el gráfico tuviera muchas variables, no aportaría mucha información y sería prácticamente incomprensible.

Elementos a considerar en su construcción

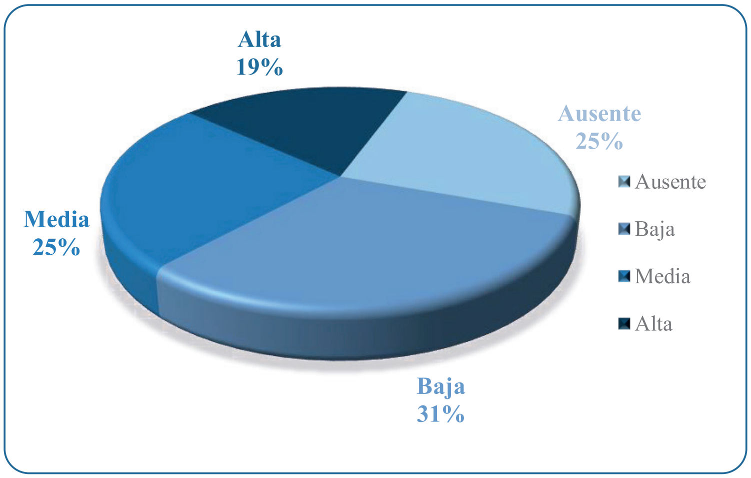
1. La totalidad de la información se representa por el número total de grados de un círculo (360º).
2. Para obtener los grados correspondientes a cada categoría, se multiplica 3. 6º por la frecuencia relativa utilizada.

Ejemplo: Utilizando la información del ejemplo anterior, y haciendo los cálculos, el gráfico quedaría de la siguiente forma:

Cálculos previos:

Vulnerabilidad Psicosocial	Porcentaje (%)	3.6 x %
Ausente	25.00	90.0
Media	31.25	112.5
Baja	25.00	90.0
Alta	18.75	67.5
Total	100.00	360.0

Gráfico 2. Distribución de ancianos según vulnerabilidad psicosocial. Municipio de Pucarani, 2019.



Fuente: Tabla 1

Nota: Con toda intención representamos los mismos datos por dos formas gráficas diferentes (barras simples y pastel), así te confirmamos la posibilidad de utilizarlos indistintamente, aunque si la variable tiene más de cinco categorías, es preferible usar las barras simples.

4.3.3.3. Gráfico de barras múltiples

Uso: Este gráfico se utiliza cuando queremos representar dos variables, las cuales pueden ser: cualitativas o cuantitativas discretas ambas, o una cualitativa y la otra cuantitativa discreta; y la información se dispone en frecuencias absolutas o relativas, o en medidas de resumen. Los datos se representan mediante barras agrupadas, como verás a continuación.

Elementos a considerar en su construcción

1. Dispondrás grupos de dos, tres o más barras, es decir, barras dobles, triples, etc.
2. El número de grupos a formar dependerá del número de categorías consignadas en la columna matriz o en la fila de encabezamiento, según tu gusto.

3. La separación entre cada grupo de barras es aproximadamente la mitad del ancho del grupo.
4. Este gráfico se origina a partir de tablas bidimensionales.

Ejemplo: El siguiente gráfico resume la información de 300 niños de un Círculo Infantil atendido por un médico de familia, atendiendo a las variables sexo y raza.

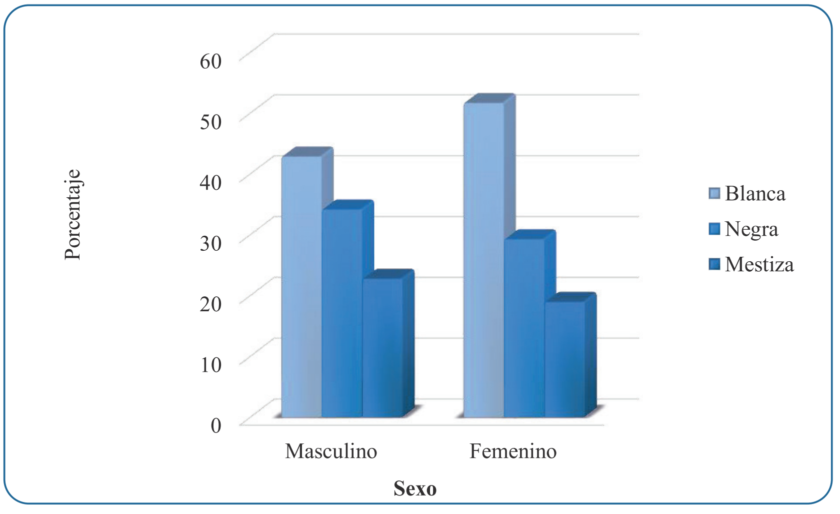
Tabla 2. Distribución de niños según raza y sexo. Centro Infantil "Jiska Lala". Municipio La Paz, 2020.

Raza	Sexo			
	Masculino		Femenino	
	No.	%	No.	%
Blanca	79	42.9	60	51.7
Negra	63	34.2	34	29.3
Mestiza	42	22.8	22	19.0
Total	184	100.0	116	100.0

Fuente: Libro de matrícula del Centro Infantil Jiska Lala

Nota: el porcentaje se calculó por columnas.

Gráfico 3. Distribución de niños según raza y sexo. Centro Infantil "Jiska lala". Municipio La Paz, 2020.



Fuente: Tabla 2

4.3.3.4. Gráfico de barras compuestas

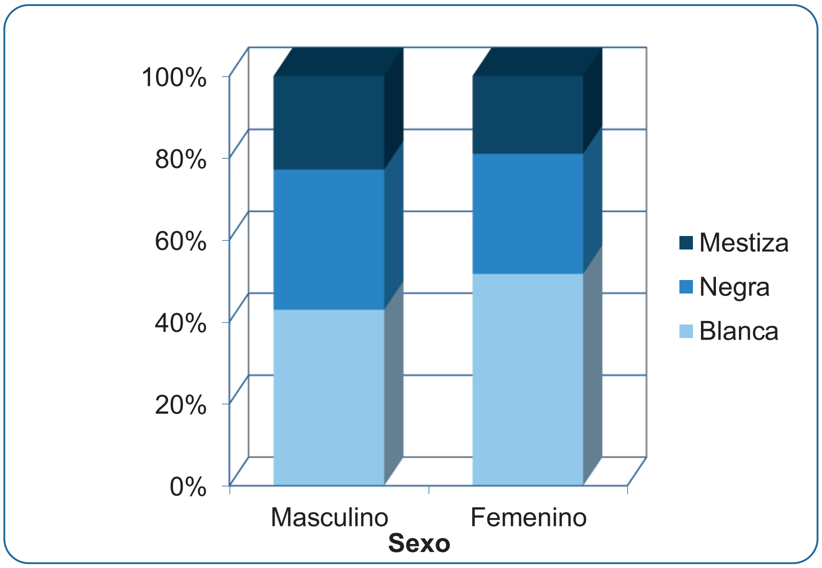
Uso: Al igual que el gráfico anterior, utiliza este cuando quieras representar dos variables: ambas cualitativas o cuantitativas discretas, o una cualitativa y la otra cuantitativa discreta; y dispongas la información en frecuencias relativas. Aquí, la información perteneciente a una variable se representa en su totalidad en una sola barra.

Elementos a considerar en su construcción

1. Cada barra representa el ciento por ciento de la información del grupo representado.
2. El ancho de las barras queda a tu gusto, pero debe ser el mismo para todas.
3. La separación entre las barras es aproximadamente la mitad del ancho.
4. Lo originan tablas bidimensionales.

Ejemplo: Utilizando la información del ejemplo anterior, el gráfico quedaría de la siguiente forma:

Gráfico 4. Distribución de niños según raza y sexo. Centro Infantil "Jiska Lala". Municipio La Paz, 2020.



Fuente: Tabla 2

4.3.4. Gráficos para representar variables en escala cuantitativa continua

A continuación, te presentamos un grupo de gráficos que se estudiarán en este epígrafe, atendiendo al número de variables que representan.

Variables	Gráfico
1	Histograma de frecuencias, ojiva*
2	Polígono de frecuencias

*: No abordaremos este gráfico por su infrecuente uso en nuestro ámbito.

Histograma.

4.3.4.1. Histograma

Uso: Este gráfico consiste en barras adyacentes, y se utiliza cuando queremos representar una variable cuantitativa continua, y la información se dispone en frecuencias absolutas o relativas, o en medidas de resumen.

Elementos a *considerar en su construcción*

1. Las barras o rectángulos se disponen unidos para dar idea de continuidad.

2. El ancho dependerá de la amplitud de los intervalos de clase en que se clasifica la variable en estudio.
3. La altura de cada IC se obtiene mediante el cociente frecuencia absoluta/amplitud.
4. Por el eje X se consigna el límite de clase inferior o real de cada intervalo.
5. Lo originan tablas unidimensionales.

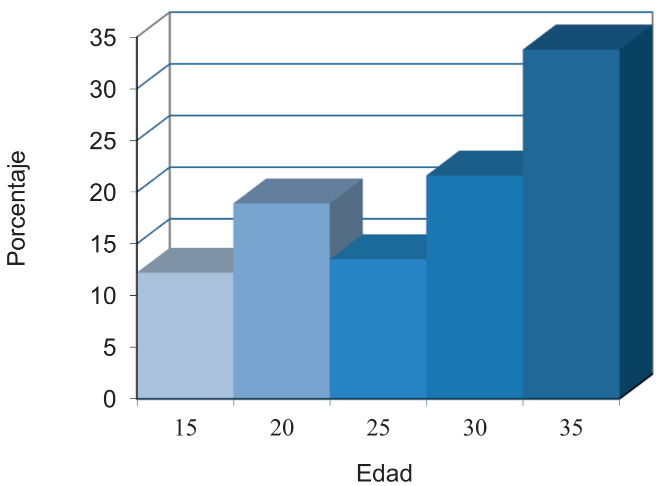
Ejemplo: A continuación, te presentamos los resultados de un estudio relacionado con las edades maternas.

Tabla 3. Distribución de recién nacidos según edad materna. Hospital General San Juan de Dios. Oruro, julio, 2020.

Edad materna	Número	Porcentaje
15 – 19	45	12.2
20 – 24	70	18.9
25 – 29	50	13.5
30 – 34	80	21.6
35 – 39	125	33.8
Total	370	100.0

Fuente: Registro de nacimientos. Hospital General San Juan de Dios.

Gráfico 5. Distribución de recién nacidos según edad materna. Hospital General San Juan de Dios. Oruro, julio, 2020.



Fuente: Tabla 3

4.3.4.2. Polígono de frecuencias

Uso: Este gráfico se utiliza cuando queremos representar hasta dos variables, de las que al menos una debe ser cuantitativa continua, y la información se dispone en frecuencias absolutas o relativas, o en medidas de resumen. Está formado por una o dos curvas que representan a cada variable estudiada.

Elementos a considerar en su construcción

a) Se pueden construir histogramas inicialmente, y luego marcar los puntos medios de cada IC (marca de clase), los cuales al unirse forman una curva.
b) Habrá tantas curvas como categorías tenga la variable discontinua.
c) Lo originan tablas uni o bidimensionales.

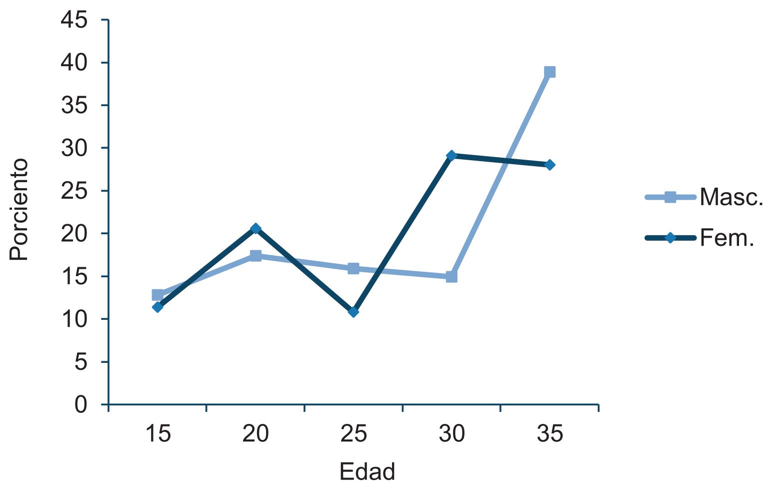
Ejemplo: A continuación, te presentamos los resultados de un estudio relacionado con las edades maternas y el sexo de los recién nacidos.

Tabla 4. Distribución de recién nacidos según edad materna y sexo. Hospital General San Juan de Dios. Oruro, julio, 2020.

Edad materna	**Sexo**			
	Masculino		Femenino	
	No.	%	No.	%
15 – 19	25	12.8	20	11.4
20 – 24	34	17.4	36	20.6
25 – 29	31	15.9	19	10.8
30 – 34	29	14.9	51	29.1
35 – 39	76	38.9	49	28.0
Total	195	100.0	175	100.0

Fuente: Registro de nacimientos Hosp. General San Juan de Dios.

Gráfico 6. Distribución de recién nacidos según edad materna y sexo. Hospital General San Juan de Dios. Oruro, julio, 2020.



Fuente: Tabla 4

4.3.5. Gráficos para presentar variables en el tiempo

A continuación, te mostramos el gráfico aritmético simple, utilizado en el estudio de las series cronológicas o temporales, como algunos las llaman. No obstante, es bueno que sepas que hay quienes utilizan como gráfico de trabajo el de barras simples para el tratamiento de fenómenos vistos en el tiempo.

Gráfico aritmético simple (GAS).

Uso: Este gráfico se utiliza para representar una variable a través del tiempo.

Elementos a considerar en su construcción

1. Cada categoría o clase de la variable se representa por una curva.
2. En el eje de las abscisas se consignará el año, mes, semana, etc., según la unidad en que se mida el tiempo.
3. En ocasiones, cuando los ejes no ajustan, se utiliza una escala semilogarítmica para su construcción.

Ejemplo: A continuación, te presentamos la mortalidad perinatal de Bolivia desde 2010 hasta 2018.

Tabla 5. Mortalidad perinatal según componentes. Bolivia, 2010-2018.

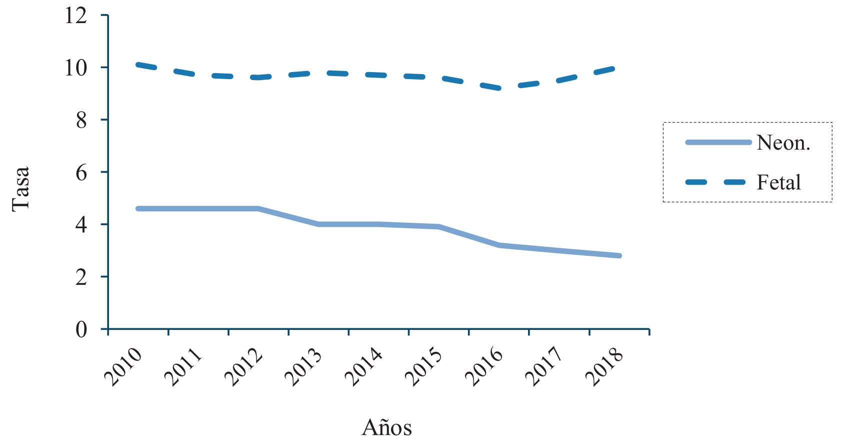
Años	Mortalidad			
	Neonatal		Fetal*	
	No.	Tasa	No.	Tasa
2010	861	4.6	1897	10.1
2011	811	4.6	1708	9.7
2012	720	4.6	1532	9.6
2013	601	4.0	1506	9.8
2014	598	4.0	1442	9.7
2015	586	3.9	1424	9.6
2016	456	3.2	1304	9.2
2017	461	3.0	1462	9.5
2018	435	2.8	1519	10.0

Fuente: Anuario Estadístico de Bolivia 2018

Tasa por 1 000 NV y defunciones fetales de 1 000 gramos y más

*: 1 000 gramos y más.

Gráfico 7. Mortalidad perinatal según componentes. Bolivia, 2010 – 2018.



Fuente: Tabla 5

Resumen

En esta unidad estudiaste que:

a) La información puede presentarse de forma resumida en tablas y gráficos estadísticos.

b) La tabla o cuadro estadístico es un recurso que se emplea con el fin de presentar información resumida, organizada por filas y columnas.

c) La estructura del cuadro estadístico incluye: Identificación, Título, Cuerpo de la tabla, Fuente y Notas explicativas.

d) Los errores cometidos con mayor frecuencia en la confección de tablas estadísticas son:

I. Errores en la presentación.
- Cuadros sin identificación.
- Título o encabezamiento incorrectos o inadecuados:
- Telegráfico: título demasiado pequeño, carente de claridad.
- Ampuloso: título demasiado extenso, que incluye vocablos que no aportan nada a la claridad del texto.

II. Errores del cuerpo.
- Errores de cálculo.
- Disposición incorrecta de los datos.
- Mostrar solamente medidas relativas (frecuentemente porcentajes) o medidas de resúmenes.
- Cuadros sobrecargados.

III. Errores en la fuente.
- No citar la fuente cuando es secundaria.
- Citar la fuente cuando es primaria.
- Consignar como fuente aquello que no es un documento (oficinas, departamentos)

a) El gráfico estadístico complementa la información previamente presentada en tablas.

b) La estructura del gráfico estadístico incluye: Identificación, Título, Gráfico propiamente dicho, Fuente, Notas explicativas y Leyenda.

c) Los gráficos más comúnmente usados en el ámbito sanitario son: de barras, pastel, histograma y polígono de frecuencias.

Unidad didáctica N°5

SÍNTESIS Y ANÁLISIS DE LA INFORMACIÓN PARA VARIABLES CUALITATIVOS

Competencia de la unidad:

Interpreta y analiza correctamente la información estadística a partir de un conjunto de datos cualitativos.

5.1. Introducción

En la investigación científica en la Atención Primaria de Salud, se utilizan con frecuencia variables cualitativas, bien por su naturaleza, o por la escala empleada. Por supuesto, una vez que la información se recogió, es necesario calcular alguna medida de resumen cuyo resultado es un indicador que deberá analizarse en un momento posterior.

En este tema te presentamos las medidas de resumen para variables cualitativas que se utilizan con mayor frecuencia en los estudios que realizas en el nivel primario de atención de salud.

5.2. Razón e índice

5.2.1. Definición:

Una razón es la relación por cociente que se establece entre las unidades de análisis que pertenecen a un grupo o categoría (a) y las unidades de análisis que pertenecen a otra categoría (b) de la misma variable. Su expresión general es: $\frac{a}{b}$.

¡Uhmm! ¿Ésa es la definición? No te desanimes, es una medida de fácil comprensión. Te la explicaremos con un ejemplo:

5.2.2. Cálculo

Supongamos que de los 400 recién nacidos (RN) de un municipio en cierto período, 300 presentaron los ojos oscuros (OO), en tanto que sólo 100 los tenían claros (OC). Aplicando la expresión general, la razón OO/OC es:

$$R = \frac{a}{b} = \frac{RN\ con\ OO}{RN\ con\ OC} = \frac{300}{100} = 3$$

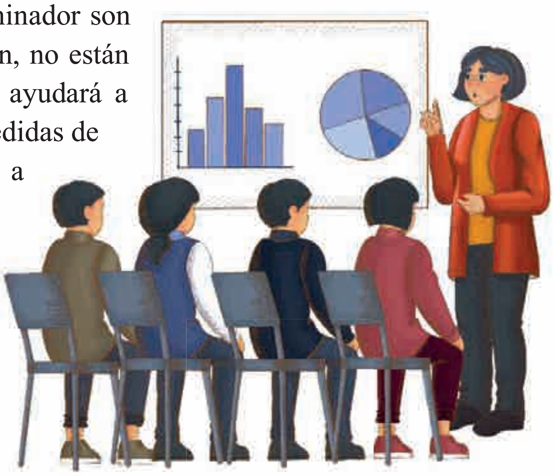
Nota: Utilizamos la letra R por razones didácticas, realmente la razón no tiene símbolo propio.

5.2.3. Interpretación

La razón ojos oscuros/ojos claros es de 3; o lo que es lo mismo, 3:1.

Pero, ¿qué significa este resultado? Expresa que hay tres recién nacidos con ojos oscuros por cada recién nacido de ojos claros en ese municipio y en ese período.

Fíjate que el numerador y el denominador son disjuntos, es decir, no se interceptan, no están contenidos uno en el otro. Ello te ayudará a establecer las diferencias con las medidas de resumen que estudiarás a continuación.



Si multiplicas el resultado obtenido por 100, entonces el nuevo número se denomina índice, de tal suerte que en el ejemplo anterior el índice sería 300. En otras palabras, en el municipio de referencia, en el período estudiado, por cada 100 bebés de ojos claros hay 300 de ojos oscuros.

5.3. Proporción y Porcentaje.

5.3.1. Definición

Una proporción es la relación por cociente que se establece entre las unidades de análisis que pertenecen a un grupo o categoría (a) de una variable y el total de las unidades de análisis estudiadas (a + b). Su expresión general es: $\frac{a}{a+b}$. Si se multiplica su resultado por 100, se obtendrá el porcentaje.

5.3.2. Cálculo

Seguiremos utilizando el ejemplo del epígrafe anterior. ¿Lo recuerdas? Por supuesto que sí. Pues bien, determinemos la proporción de niños con ojos oscuros (300) en la población de recién nacidos (400):

$$P = \frac{a}{a+b} = \frac{RN\ con\ OO}{Total\ de\ RN} = \frac{300}{400} = \frac{3}{4}$$

Alternativamente, puedes calcular el porcentaje:

$$P \times 100 = \frac{RN\ con\ OO}{Total\ de\ RN} \times 100 = \frac{300}{400} \times 100 = 75.0\%$$

Nota: Usamos la P con fines ilustrativos, pues la proporción carece de simbología.

5.3.3. Interpretación

Los resultados anteriores significan que tres de cada cuatro recién nacidos tienen los ojos oscuros; o que el 75 por ciento de los recién nacidos tiene los ojos oscuros (y, obviamente, el 25% los tiene claros).

¿No te resultan familiares estas nuevas medidas, o sea, la proporción y el porcentaje? Ya debes estarte preguntando la diferencia que existe entre éstas y la distribución de frecuencias relativas que ya estudiaste. Nada más claro: no es que sean parecidas, *son exactamente las mismas*, pero restringidas a variables cualitativas.

Observa que el porcentaje te permite analizar el aporte, el peso específico o la importancia relativa de cada categoría respecto al total.

Otro elemento que debes conocer es el siguiente: si la variable es dicotómica, puedes utilizar tanto razones como proporciones; pero si es politómica, entonces calcula sólo proporciones[9].

5.4. Tasas

Siempre que necesites medir el riesgo de que acontezca cierto fenómeno en una población determinada, dispones de un indicador valioso y único: las tasas.

Una tasa es una relación por cociente que expresa el riesgo de que ocurra cierto evento en una población y período determinados. Está compuesta por tres elementos, a saber:

$$Tasa = \frac{\text{numerador}}{\text{denominador}} \times k$$

Veamos cuáles son esos elementos:

- El numerador contiene al número de veces que ocurrió determinado fenómeno en un área geográfica y en un período determinados.
- El denominador indica el número de habitantes de la población expuesta al riesgo en la cual puede ocurrir el fenómeno.
- k es un múltiplo de 10 cuyo uso está justificado por el hecho de que habitualmente el resultado del cociente es un número fraccionario, y al multiplicarlo por una potencia de 10 se facilita enormemente la lectura y comprensión del indicador.

Esta es una medida que expresa el riesgo de ocurrencia del evento estudiado en el numerador en la población involucrada, en el tiempo y lugar establecidos.

El numerador y el denominador deben coincidir en tres aspectos:

- Naturaleza del hecho: el conjunto de elementos del numerador forma parte de la población que va en el denominador y por lo tanto se dice que son de la misma naturaleza.
- Zona geográfica.
- Tiempo en que ocurre el hecho: las tasas generalmente se calculan para períodos de un año. Como la población es variable durante el año, se considera como población representativa del período a la población estimada al 30 de junio

5.1.1. Tipos de tasas

Se pueden separar en dos grandes grupos: crudas o brutas, y específicas.

[9] De hecho, la interpretación y cálculo de razones en una politomía está proscrito.

- Crudas o brutas son aquellas que tienen en el denominador a toda la población.
- Específicas son las que en el denominador tienen una población específica, por ejemplo: los menores de 20 años.

Las tasas que más importancia revisten para nuestro desempeño en el campo de la Salud Pública son las siguientes:

Tasas de importancia y relevante en Salud Pública.	Tasas relacionadas con la natalidad Tasas relacionadas con la mortalidad Tasas relacionadas con la morbilidad

Una particularidad realmente útil de las tasas es que puedes calcularlas tanto para la totalidad de la población, como para parte de ella (por ejemplo, para el grupo de edad de cinco a nueve años, para los estudiantes, para los residentes del área rural, y así por el estilo); por otra parte, puedes calcular las tasas para todas las causas, o solamente para una de ellas (o un grupo de ellas). De este modo, tendrás calculadas tasas brutas, crudas, generales o globales si se tratara de tasas que involucren a toda la población o al total de causas; al tiempo que habrás calculado tasas específicas si incluían a una parte de la población o a una causa o grupo de ellas.

Así las cosas, estarás en plena facultad de hallar tasas brutas de mortalidad, de natalidad, o bien específicas por edad, por sexo, por edad y sexo a la vez, entre muchas otras. Teniendo a tu disposición los datos adecuados, podrás hallar una tasa tan específica como desees.

Existe en punto cardinal en el manejo de las tasas: la población expuesta al riesgo en cuestión. Como ya sabes, este es el denominador de la ecuación, y de su correcta determinación depende la fidelidad del cálculo. Nunca serán suficientes las medidas que tomes para asegurarte que estás empleando el dato acertado. No creas que es muy difícil saber que estás errado o en lo cierto, el problema radica en que muchas veces se pasa por alto este “detalle” de forma involuntaria.

Probablemente te habrás preguntado: «Bueno, ¿y qué tanto problema con el denominador?» ¡Ah! Es que ahí radica el quid de la cosa. Recuerda que calculas una tasa para medir el riesgo de ocurrencia de un evento o fenómeno en una población, pero no en cualquier población, sino en la población expuesta a ese riesgo. Esto quiere decir que sólo podrás calcular la tasa de mortalidad por cáncer de útero en las mujeres de cierta ciudad, puesto que sería imposible calcularla en los hombres; del mismo modo que no puedes calcular la tasa de morbilidad por cáncer de pulmón de los habitantes de Pueblo Mocho en 1999, utilizando para ello a los habitantes que tenía el pueblo en 1979, o a los habitantes de Palma Mocha en 1999. ¿Satisfecha tu inquietud?

También haz de saber que las poblaciones están sometidas a constantes cambios en lo que a su número atañe, determinados por los nacimientos y defunciones y por los movimientos migratorios (emigración e inmigración), que provocan que no sea la misma a lo largo de todo el año. De ahí que, por convenio, se tome la población

existente ***a mediados del período***[10] ***o población media*** para el cálculo de las tasas.

Por otra parte, debes tener especial cuidado al calcular tasas para poblaciones pequeñas, como la que usualmente manejan los Consultorios, pues suelen volverse inestables, ya que cualquier evento "mueve "mucho la tasa, y a veces no guarda relación el resultado obtenido con la magnitud del evento acontecido.

Bueno, ya estamos en condiciones de particularizar en las tasas más relevantes en la práctica diaria.

5.1.2. Tasas relacionadas con la natalidad

El estudio de la natalidad está relacionado con el número de nacimientos ocurridos en una población y tiempo determinados, así como la distribución que siguen de acuerdo con ciertas características. Como ves, todo gira en torno a la medición de la misma, y una de las formas de conseguirlo es utilizando las tasas.

Ahora nos tropezamos con una contrariedad: la población expuesta al riesgo es muy difícil de definir, ya que tener un hijo no involucra a toda la parte femenina de la población, y va más allá, pues otros factores de índole psicosocial actúan en tal decisión. Por estas razones, verás que se han buscado soluciones alternativas a esta situación.

5.1.2.1. Tasa bruta de natalidad

Comencemos por la tasa bruta de natalidad. La misma expresa cómo se comportan los nacimientos en un área y tiempo determinados. Su cálculo es sencillo:

$$TBN = \frac{Total\ de\ nacidos\ \ vivos\ en\ lugar\ y\ tiempo\ x}{población\ total\ en\ lugar\ y\ tiempo\ x} \times 1000$$

Por ejemplo, la tasa cruda de natalidad del departamento de Oruro en el año 2021 fue:

- Total, de nacidos vivos en el departamento de Oruro durante 2021: 9664
- Total, de habitantes en el departamento de Oruro durante 2021: 548537[11]

$$TBN = \frac{9664}{548537} \times 1000 = 17.6$$

Bien, ya tienes el número calculado. Pero ¿es suficiente con eso? Claro que no, necesitas saber qué significa, a fin de manejarlo apropiadamente. En primer lugar, debes informar el resultado de la siguiente forma: «La tasa bruta de natalidad en el departamento de Oruro en el año 2021 fue de 18 nacido

[10] se toma la población del 30 de junio.

[11] A menos que se indique lo contrario, todos los datos que aparecen proceden de estimaciones poblacionales del Sistema Nacional de Información en Salud y Vigilancia Epidemiológica, Bolivia,2021.

vivos por cada 1000 habitantes», ello significa que durante 2021 en Oruro nacieron como promedio 18 niños por cada 1000 habitantes.

Esta tasa tiene el inconveniente de no tomar en cuenta a las personas realmente expuestas al evento, pero por su sencillez y facilidad de comprensión es la medida más generalmente utilizada.

De ahora en adelante nos limitaremos a enseñarte cómo calcular e interpretar el indicador. Continuemos entonces.

5.1.2.2. Tasa general de fecundidad

Este indicador mide la natalidad, pero tomando en cuenta solamente a la población femenina en edad reproductiva o fértil (15 a 49 años). El hecho de que se restringe el denominador no inyecta especificidad a la tasa, pues continúa siendo una mezcla de diversos grupos de edades con situaciones diversas; a menos de que se mueve a la par de la tasa cruda de natalidad. Se calcula de la siguiente forma:

$$\text{TGF}=\frac{\text{Total de nacido vivos en lugar y tiempo x}}{\text{Mujeres en edad fértil en el lugar y tiempo x}}\times 1000$$

Así, la tasa de fecundidad del departamento de Chuquisaca en el año 2021 fue:

$$TGF = \frac{13744}{164245} \times 1000 = 83.6$$

Interpretación: En el departamento de Chuquisaca, durante 2021, nacieron como promedio 84 niños por cada 1000 mujeres en edad fértil (15 a 49 años).

5.1.2.3. Tasa de fecundidad específica por edad

Esta es una tasa específica, que usualmente se calcula para grupos quinquenales comprendidos entre 15 y 49 años.

$$TEF_{edad} = \frac{\text{NV de mujeres de edad i en lugar y tiempo X}}{\text{mujeres de edad i en lugar y tiempo X}} \times 1000$$

$$TEF_{15-49} = \frac{8734}{164245} \times 1000 = 53.2$$

Interpretación: Durante 2021 en Chuquisaca nacieron como promedio 53 niños por cada 1000 mujeres de 15 a 49 años de edad.

5.1.2.4. Tasas brutas de mortalidad

Esta tasa expresa el riesgo que tienen todos los habitantes de cierta población, en un momento dado, de morir por cualquier causa.

$$TBM = \frac{\text{fallecidos en lugar y tiempo X}}{\text{población en lugar y tiempo X}} X\ 1000$$

En el 2021, en el departamento de Santa Cruz, esta tasa fue:

$$TBM = \frac{19382}{3363377} = 5.76$$

Esto significa que, en 2021, en el departamento de Santa Cruz fallecieron como promedio 6 personas por cada 1 000 habitantes.

5.1.2.5. Tasa de mortalidad por edad

Ahora te presentamos una tasa de mortalidad específica, que solo mide el riesgo de morir que tienen las personas del grupo de edad analizado. Su cálculo se logra restringiendo el denominador a las personas de la edad deseada, e incluyendo en el numerador a los fallecidos en esa edad.

$$TME_{edad} = \frac{\text{fallecidos de edad i en lugar y tiempo X}}{\text{población de edad i en lugar y tiempo X}} X1000$$

Por ejemplo, en 1998, en Cuba, la tasa de mortalidad en personas de 60 años y más fue:

$$TME_{60\ y\ +} = \frac{52\ 558}{1\ 440\ 127} = 50$$

Así, durante 1998, en Cuba fallecieron como promedio 50 individuos de 60 y más años por cada 1 000 personas de ese grupo de edad.

5.1.2.6. Tasa de mortalidad por sexo

El cálculo de esta tasa es muy similar a la anterior, con la diferencia de que te restringes a un sexo en particular. Expresa el riesgo de morir de las personas de ese sexo en esa población, en el período especificado. Para calcularla, sustituye el numerador por el total de defunciones del sexo analizado, y el denominador por el total de habitantes de ese sexo en el lugar y momento deseados.

$$TME_{sexo} = \frac{\text{fallecidos de sexo i en lugar y tiempo X}}{\text{población sexo i en lugar y tiempo X}}\ X\ 1000$$

En nuestro país, durante 1998 la tasa de mortalidad del sexo femenino fue:

$$TME_{fem} = \frac{34\ 692}{5\ 550\ 426} X1000 = 6$$

Interpretación: en Cuba, en 1998 fallecieron como promedio 6 mujeres por cada 1000 féminas.

5.1.2.7. Tasa de mortalidad por causa

Análogamente, puedes conocer el riesgo a que están sometidos los habitantes de cierto lugar, en un momento definido, de morir por una causa de muerte dada. Ahora el numerador está formado por las defunciones debidas a la causa en cuestión, mientras que el denominador incluye al total de población.

$$TME_{causa} = \frac{\text{fallecidos por causa i en lugar y tiempo X}}{\text{población en lugar y tiempo X}}\ X\ 100000$$

En el departamento de Cochabamba, durante 2021 la tasa de mortalidad por enfermedades de la Covid-9 fue:

$$TME_{enf.covid-19} = \frac{3403}{2086930} X\,100000 = 163$$

Interpretación: en el departamento de Cochabamba, en el año 2021 fallecieron como promedio 163 personas por enfermedad del Covid-9 por cada 100 000 habitantes.

5.1.2.8. Tasa de mortalidad infantil

Arribamos a un punto de suma importancia al analizar la situación de salud de una comunidad. Este indicador es una especie de señal hacia la cual se dirigen los ojos de todo aquel que, tome interés en el estudio de las características de una población.

Es un indicador que toma como población expuesta al riesgo a los nacidos vivos en período estudiado, y se calcula de la siguiente forma:

$$TMI = \frac{\text{fallecidos menores de 1 año en lugar y tiempo X}}{\text{NV en lugar y tiempo X}} X\,1000$$

Una de las tareas fundamentales para el Estado, es la reducción de la mortalidad infantil, en este sentido a través del Ministerio de Salud se implementó el programa público Bono “Juana Azurduy” creado en abril de 2009 para el uso de los servicios de salud, en beneficio de mujeres gestantes, niños y niñas menores de dos años, así mismo se implementó la Ley de Prestaciones de Servicios de Salud Integral de 30 de diciembre de 2013 (Ley 475), y ley modificatoria 1152 de 20 de febrero de 2019 “Hacia el Sistema Único de Salud Universal y Gratuito” La reducción de la mortalidad aún se constituye en tarea pendiente, por lo que el trabajo debe centrarse principalmente en la atención y disminución de factores de riesgo que conllevan al incremento de las tasas de mortalidad. En Bolivia en el periodo de quinquenio 2011-2016 previo a la EDSA 2016 se registró una tasa de mortalidad infantil de 24 defunciones por cada mil nacidos vivos. Ello significa que en el periodo de 2011 a 2016, en Bolivia Fallecieron como promedio 24 niños por cada 1000 nacidos vivos

Este indicador tiene la singularidad de que puede descomponerse en varios indicadores, que miden con más especificidad el comportamiento de la mortalidad en el menor de un año. Estos componentes son los siguientes:

5.1.2.9. Tasa de mortalidad neonatal

La tasa de mortalidad neonatal y se define como el cociente entre el número de nacidos vivos que mueren antes de cumplir 28 días de edad durante un año dado y el número de nacidos vivos durante el mismo año, en determinado país, territorio o área geográfica, expresado por 1.000 nacidos vivos mide la probabilidad de morir durante aproximadamente en el primer mes de vida.

$$TMN = \frac{\text{fallecidos menores de 28 dias en lugar y tiempo X}}{\text{NV en lugar y tiempo X}} X\,1000$$

En Bolivia en el quinquenio de 2011-2016 se registró según EDSA una tasa de mortalidad neonatal de 15 por mil nacidos vivos. Lo que quiere decir que, en el quinquenio de 2011 a 2016, en Bolivia fallecieron como promedio de 15 niños menores de 28 días por cada 1000 nacidos vivos.

5.1.2.10. Tasa de mortalidad neonatal precoz

Al calcular esta tasa conocerás el riesgo de morir de los bebés con menos de siete días de nacidos. Su cálculo estriba en sustituir el numerador de la TMI por las defunciones ocurridas en recién nacidos de menos de siete días en el período y lugar estudiados.

$$TMNP = \frac{\text{fallecidos} < \text{ de 7 dias en lugar y tiempo X}}{\text{Numero de NV en lugar y tiempo X}} X\,1000$$

En 2020 según los reportes de SEDES La Paz tuvimos una TMNP de 7.5 por 1000 NV. De este modo, puedes decir que, en el departamento de La Paz durante el año 2020, fallecieron como promedio 8 niños de menos de 7 días por cada 1000 nacidos vivos.

5.1.2.11. Tasa de mortalidad neonatal tardía

Conforme calculaste el riesgo de muerte de los bebitos menores de siete días, puedes conocer también el de siete en adelante y menores de 28 días, cerrando así el diapasón en la etapa neonatal de la vida. Sólo tienes que sustituir el numerador de la tasa anterior por las defunciones de niños de 7 - 27 días en la población de tu interés, durante el período que necesites.

$$TMNT = \frac{\text{fallecidos} \geq \text{ 7 a } < \text{28 dias en lugar y tiempo X}}{\text{Numero de NV en lugar y tiempo X}} X\,1000$$

Para el departamento de La Paz la TMNT en 2020 fue de 3.2 por cada 1000 nacidos vivos, lo que quiere decir que, en 2020, en el departamento de La Paz fallecieron como promedio 3 niño de 7 a 27 días por cada 1000 nacidos vivos.

5.1.2.12. 1. Tasa de mortalidad post neonatal

$$TMP = \frac{\text{fallecidos} \geq \text{28 dias y} \leq \text{a1 año en lugar y tiempo X}}{\text{NV en lugar y tiempo X}} X\,1000$$

En el caso de la mortalidad post neonatal en nuestro país, en el período 2011-2016 llega a 10 por mil nacidos vivos, observándose una reducción del 56%, dado que este indicador para el período 2003-2008 era de 23 por mil nacidos vivos. De este modo, puedes decir que, en Bolivia, durante el periodo 2011 - 2016, fallecieron como promedio 10 niños mayores de 27 días a 1 año de edad por cada 1000 nacidos vivos.

Resumen

En este tema estudiaste que:

1. Las medidas de resumen para datos cualitativos más frecuentemente utilizadas son las razones, las proporciones y las tasas.
2. Cada uno de esos indicadores tiene diferente interpretación. Así, los más refinados son las tasas, pues expresan el riesgo de ocurrencia del evento consignado en su numerador.
3. Debes tener cuidado al calcular las tasas para poblaciones pequeñas, por ejemplo, en el Consultorio Médico de la Familia, porque suelen ser inestables.
4. Las tasas pueden dividirse en generales y específicas.
5. En el ámbito sanitario, las tasas más usadas son las de natalidad, mortalidad y morbilidad.

Unidad didáctica N°6

SÍNTESIS Y ANÁLISIS DE LA INFORMACIÓN PARA VARIABLES CUANTITATIVOS

Competencia de la unidad:

Interpreta y analiza correctamente la información estadística a partir de un conjunto de datos cuantitativos.

6.1. Introducción

En la unida anterior viste cómo resumir las variables cualitativas, ya lo fuesen por naturaleza o porque manejaste los datos en escala cualitativa. La información cuantitativa también es dable de ser resumida, con las consabidas ventajas que de ello se generan, pues corrientemente es necesario representar un conjunto de datos por un número que en la medida de lo posible logre describir a dicho conjunto. Para obtenerlo, podrás disponer de tres grandes grupos de medidas de resumen: de tendencia central, de posición, y de dispersión o variación, las cuales verás en los apartados siguientes.

Antes de empezar

En esta unidad nos valdremos de algunas notaciones matemáticas para representar las fórmulas de cálculo de los distintos indicadores.

Con el ánimo de refrescarte la memoria, te mencionamos los elementos propios del lenguaje matemático que empleamos.

Para representar las distintas variables utilizamos las letras del alfabeto. Por ejemplo, si medimos la edad y la talla de cinco individuos, pudiésemos referirnos a la primera como la variable X, y a la segunda como la variable Y, de ahí que cada vez que hagamos alusión a X, sabemos que en realidad estamos hablando de la edad de las personas estudiadas (si fuera Y, entonces hablamos de la talla). Te aclaramos que para escoger las letras no hay una regla específica, eso queda a tu decisión.

Supongamos que los resultados de un estudio fueron los siguientes:

Individuos	Edad (años)	Talla (cm)
	X	Y
María	24	130.0
Juan	26	120.0
Rosa	27	140.0
Jaime	25	150.0

Ruth	23	110.0

Por otra parte, pudieras referirte a la edad por las letras ed, o algo por el estilo. Lo que debe quedar claro, es a qué te refieres con la simbología utilizada, pues X puede significar edad en un estudio, pero sexo en otro, por citar un ejemplo.

Ahora bien, llegamos a otro punto que necesita ser definido. Para referirte a las edades de las cinco personas, y suponiendo que las representaste por la letra X, pudieras escribir entre otras tantas formas:

X(María) = 24   X(Rosa) = 27   X(Ruth) = 23

X(Juan) = 26   X(Jaime) = 25

Como habrás visto, es un procedimiento bastante tedioso el utilizar los paréntesis u otra forma similar de identificación de los datos. En su lugar, puedes usar los subíndices, resultando algo por el estilo: X1 = 24, X2 = 26, X3 = 27, X4 = 25, X5 = 23. De la misma forma, las tallas serían: Y1 = 130, Y2 = 120, Y3 = 140, Y4 = 150, Y5 = 110.

También resultan muy útiles los subíndices generales (por lo general son letras), los cuales hacen alusión a un grupo de valores de la variable. Por ejemplo, para decir que hay cinco valores de la edad, puedes escribir: Xi (i = 1, 2, 3, 4, 5); o lo que es lo mismo: Xi (i = 1, …,5). Observa que lo anterior no dice que la variable toma los valores 1, 2, 3, 4 ó 5, sino que hay cinco mediciones de la misma, ¿claro?

Supón ahora que sumas todos los valores de la variable peso, entonces escribirías lo siguiente:

$$X1 + X2 + X3 + X4 + X5 = 650$$

En el caso que nos ocupa, resulta fácil y rápida esta forma de escritura, pero si fuesen, digamos, ¡200! valores…

En esta situación, se utiliza la letra griega sigma mayúscula Σ, que representa el símbolo de ***sumatoria***, el cual antecede a la variable en cuestión y se acompaña de dos anotaciones: una encima y otra debajo, como lo siguiente: $\sum_{i=1}^{n} X_i$. Esto se lee como “la suma de las Xs desde i hasta n”, o sea, las Xs cuyo subíndice van desde los valores especificados en i hasta n.

Retomando el ejemplo anterior, hubieses escrito $\sum_{i=1}^{5} X_i$. Si quisieras sumar solamente los tres valores del centro, entonces escribirías: $\sum_{i=2}^{4} X_i$.

De la misma forma, $\sum_{i=1}^{5} X_i Y_i$ significa $X_1 Y_1 + X_2 Y_2 + X_3 Y_3 + X_4 Y_4 + X_5 Y_5 = (24\cdot 130) + (26\cdot 120) + (27\cdot 140) + (25\cdot 150) + (23\cdot 110) = 16\ 300$.

Al mismo tiempo, te recordamos que ***elevar al cuadrado*** un número es multiplicarlo por sí mismo, y se representa por el supra índice 2, o sea, $13^2 = 169$ (porque $13\cdot 13 = 169$). Entonces, $\sum_{i=1}^{5} X_i^2$es la representación matemática de lo siguiente —utilizando los datos del ejemplo—: $X_1^2 + X_2^2 + X_3^2 + X_4^2 + X_5^2 = 24^2 + 26^2 + 27^2 + 25^2 + 23^2 = 576 + 676 + 729 + 625 + 529 = 3125$.

La operación inversa de elevar al cuadrado no es dividirlo por sí mismo, sino *extraerle su raíz cuadrada*, que se representa por el símbolo de radical $\sqrt{\ }$, quedando bajo la barra horizontal a lo que se le extrae la raíz cuadrada. De ahí que $\sqrt{4}$ = ±2, porque 2·2 = 4, pero –2 · –2 = 4 también, por lo que debe especificarse en buena lid el símbolo ±. Para algunos datos del ejemplo anterior: $\sqrt{24}$ = ±4.89; $\sqrt{26}$ = ±5.09; $\sqrt{27}$ = ±5.19.

Por último, seguramente recordarás que el *valor absoluto o modular* de un número es él mismo sin el signo asociado, esto es, se toma la magnitud positiva del número. Se representa por dos barras verticales que enmarcan al número deseado, v.g. $|3|$ = 3, y $|-3|$ = 3.

6.2. Medidas de tendencia central

Seguramente, lo primero que estarás preguntándote es: «¿Por qué de tendencia central?». Bueno, estriba en que ellas están constituidas por un número al que se acercan o "tienden" la mayoría de las observaciones. Con otras palabras, alrededor de él se agrupan las observaciones de la serie de datos, puesto que es en sí el centro de la serie, aunque ello no significa que este número tiene que estar representado en la serie (de hecho, muchas veces no ocurre así). Veamos en detalle cada una de estas medidas.

6.2.1. La media aritmética

Puedes encontrar la medida que te vamos a presentar con diversos nombres, entre ellos los más utilizados son promedio, promedio aritmético e incluso simplemente media. Al respecto debemos aclararte que existen otras medias que no son aritméticas, pero cuando decimos media a secas, nos referimos a la aritmética.

La media aritmética es una cifra que obtienes al sumar todos los valores observados y dividirlos por el número de valores. ¿No te resulta familiar? Claro, estás muy acostumbrado al cálculo del promedio. Se denota por los símbolos μ (letra griega mu) ó $\overline{X}$ (se lee equis media), la distinción entre estos dos símbolos se hará importante en cursos sucesivos, cuando abordemos temas de Estadística Inferencial. Por lo pronto, utilizaremos el último de ellos. La media conserva las unidades de medida de la variable en su estado original, o sea, que la media de un grupo de edades en años se expresará asimismo en años.

El cálculo de la media dependerá de cómo aparezcan los datos, de tal suerte que, para datos simples12, la fórmula es:

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

Veamos un ejemplo. Supón que deseas conocer la estatura media de cinco adolescentes de tu consultorio. Estas son las observaciones (datos) de la medición de cada uno de ellos (en cm):

170.0 150.0 130.0 160.0 140.0

[12] O sea, datos no agrupados ni distribuciones de frecuencias.

$$\overline{\mathrm{X}} = \frac{170 + 150 + 130 + 160 + 140}{5} = 150 \text{ cm}$$

Este resultado indica que, en promedio, los adolescentes miden 150 centímetros. Sencillamente, no hemos hecho otra cosa que decir: «más o menos, los muchachos miden 150 centímetros».

Si los datos aparecen en una distribución de frecuencias, entonces calcula la media de la siguiente manera:

$$\overline{X} = \frac{\sum_{i=1}^{n} f_i \cdot X_i}{n}$$

donde:

- fi son las frecuencias absolutas;
- Xi son las categorías de la variable;
- n: total de observaciones.

Por ejemplo, estos son los pesos (en kilogramos) de 10 pacientes representados por sus frecuencias absolutas:

Peso (X_i)	Número (f_i)	f_i x X_i
60	3	180
61	1	61
62	2	124
63	3	189
64	1	64
Total	10	618

$$\overline{X} = \frac{168}{10} = 61.8\ Kg$$

Con este resultado puedes decir que los 10 pacientes pesan 61.8 kilogramos (o mejor, 62) como promedio.

Ahora bien, puede que los datos estén agrupados en una escala de intervalos, entonces el cálculo se realiza como verás. A continuación, te mostramos los resultados de la frecuencia cardíaca (FC), en latidos cardíacos por minuto, de 100 pacientes ingresados en el Servicio de Medicina Interna de cierto hospital:

FC	Pacientes (fi)
21 – 40	1
41 – 60	17
61 – 80	38
81 – 100	40
101 – 120	4
Total	100

La media aritmética de una serie de datos agrupados está dada por:

$$\overline{X} = \frac{\sum_{i=1}^{n} MC_i \times f_i}{n}$$

donde:

- MC_i son las marcas de clase de los intervalos,
- f_i son las frecuencias absolutas de los intervalos.
- n: total de observaciones.

Procedamos, a calcular:

1. Ante todo, necesitarás conocer las marcas de clase (MC_i) de cada intervalo, para lo cual suma el límite inferior (LI) al superior (LS) del intervalo en cuestión, y luego divídelo por dos.
2. Luego, multiplicarás la marca de clase por la frecuencia absoluta (f_i) del intervalo.
3. Sumas todos los números resultantes del paso anterior y, finalmente, divides la suma por el total de observaciones (n).

Los cálculos pertinentes por intervalos son:

Intervalo	Marcas de clase	MC x fi	
1er intervalo:	[21 + 40]/2 = 30.5	30.5 x 1 =	30.5
2º intervalo:	[41 + 60]/2 = 50.5	50.5 x 17 =	858.5
3er intervalo:	[61 + 80]/2 = 70.5	70.5 x 38 =	2679.0
4º intervalo:	[81 + 100]/2 = 90.5	90.5 x 40 =	3620.0
5º intervalo:	[101 + 120]/2 = 110.5	110.5 x 4 =	442.0
		Total =	7630.0

Por último, tenemos: $\overline{X} = \frac{7630}{100} = 76.3$, por lo que puedes decir que, en promedio, los pacientes estudiados tenían una frecuencia cardiaca de 76 latidos por minuto.

Hay ocasiones en que los datos numéricos se hacen difíciles de manejar, bien porque sean muchos, o porque constan de varios dígitos. En esta situación, te aconsejamos agruparlos primero y luego calcular la media. Por otra parte, es bueno que recuerdes que el mayor monto de la información estadística aparece agrupado, por lo que te verás obligado a utilizar la fórmula estudiada.

Cometeríamos un grave error si no habláramos de las propiedades de la media. Ciertamente, entre las más notables tenemos que:

a) Es fácilmente entendible por la mayoría de las personas (o, al menos, es fácil de explicar su significado);
b) Siempre existe, y puede calcularse para cualquier dato numérico;
c) Es única, o sea, un grupo de datos sólo tiene una media;
d) Toma en cuenta a todos los valores de la serie de forma individual, esto es, recorre la serie completa.

Esta última resulta ser sumamente importante, pues la media calculada representa a todos los valores de la serie, siendo precisamente lo que se quería lograr. Ahora bien, no siempre esto resulta beneficioso, como verás en este ejemplo: imagínate que se deseaba saber la edad promedio de las personas reunidas en un salón de cierto Círculo Infantil, para lo cual se escogió al azar uno de los que poseía dicho centro. En el momento de la medición, se encontraban presentes en el salón escogido siete bebés y la educadora que los cuidaba, siendo sus edades las siguientes (m: meses, a: años):

18m 10m 12m 16m 20m 12m 14m 34a

Edad media: 510 meses/8 = 63.75 meses = 5.3 años

¿Viste eso? Ahora tenemos que, en promedio, las personas allí reunidas tenían 64 meses de edad (¡Bueno, 5 años es algo más fácil de entender!) ¿Crees que sea cierto ese dato? Claro que no, está bastante lejos de la realidad, mas no está mal hecho el cálculo. Matemáticamente es impecable, pero la lógica dice que algo falló.

El motivo por el que apareció un resultado tan dispar es la presencia de un dato discordante en el conjunto: la edad de la educadora. Cuando en una serie de datos encuentras algún dato que se aparta de los demás de forma llamativa, entonces puedes nombrarlo dato(s) aberrante(s). Si calculásemos la media con las edades de los pequeños solamente, entonces hubiese sido de 15 meses.

En resumen, si los datos son relativamente homogéneos, la media aritmética es una buena medida de resumen; pero si existen valores muy alejados de la mayoría (datos aberrantes), entonces se distorsiona mucho y deja de reflejar la realidad existente.

6.2.2. La mediana

Aquí tienes otra de las medidas de tendencia central. Al igual que la media, puedes utilizarla para describir el "centro" de un grupo de datos. No tiene un símbolo específico que la denote; nosotros usaremos med o mediana en lo adelante.

La mediana es la observación que divide a una serie ordenada de datos en dos partes iguales, o sea, es la observación que ocupa la posición central de una serie ordenada.

De lo antedicho se deduce que lo primero que tienes que hacer para calcular la mediana es ordenar la serie, ya sea en orden creciente o decreciente. Luego, buscarás cuál de los valores es la mediana, lo cual dependerá del número total de observaciones o datos que tengas.

De tener los datos simples, si tienes un número n impar de observaciones, la del centro es la medida buscada, como lo es 32 en esta serie: 41, 40, 36, 32, 26, 21, 20. Fíjate que a ambos lados de la mediana hay la misma cantidad de números.

En este caso, por simple observación llegaste al resultado, pero puedes valerte de calcular $\frac{n+1}{2}$ para saber la posición de la mediana, comenzando a contar por cualquiera de los dos extremos de la serie. En el ejemplo anterior el resultado es

(7+1)/2 = 4, y el cuarto puesto lo ocupa el 32, no importa por cuál extremo comienzas a contar.

La contrapartida ocurre cuando el total de datos es un número par; entonces la mediana es la media aritmética de los valores del centro de la serie, como sucede en el ejemplo: 20, 24, 33, 39, 45, 51, 75, 80. Los valores del centro son 39 y 45, su media es 42, y es este el valor de la mediana de esa serie.

No debe causarte extrañeza tal proceder, pues si aplicásemos la fórmula de la posición, entonces la mediana ocuparía el lugar (8+1)/2 = 4.5, esto es, la mitad entre los números 4 y 5 de la serie. Siendo el 39 y el 45 los lugares 4º y 5º respectivamente, entonces 42 es el centro entre ellos. ¿De acuerdo?

Se te puede presentar la situación de que tengas una serie con varios valores iguales, como 50, 54, 56, 56, 56, 56, 60, 62. Aquí la mediana es 56, claro está. Recuerda que ella es el valor central del grupo, y sería un atentado abierto a la lógica cuestionarse cuál de los 56s es la mediana.

También puedes calcular la mediana para datos agrupados. Supón que tienes las edades de 100 individuos de 20 a 54 años de tu área de salud:

Intervalo	Edad (años)	Frecuencia Absoluta	Frecuencia Absoluta Acumulada
1	20 – 24	25	25
2	25 – 29	12	37
3	30 – 34	14	51
4	35 – 39	9	60
5	40 – 44	10	70
6	45 – 49	12	82
7	50 – 54	18	100
	Total	100	-

Nota: la primera columna cumple funciones didácticas solamente, y la última contiene datos de trabajo.

Para el cálculo de la mediana en series agrupadas, se utiliza la siguiente expresión:

$$med = LRI + \left[\frac{\frac{n}{2} - FAA}{f}\right] \cdot c$$

donde:

- LRI es el límite real inferior del intervalo que contiene a la mediana.
- n es el total de observaciones de la serie de datos.
- FAA es la frecuencia absoluta acumulada del IC que antecede al que contiene a la mediana.
- f es la frecuencia absoluta del IC que contiene a la mediana.
- c es la amplitud del IC que contiene a la mediana.

Ilustremos, a través del ejemplo anterior, los pasos que seguirás para hallar la mediana de la serie ordenada:

1. Calcula la mitad de las observaciones: n/2 = 100/2 = 50.
2. Determina el IC que contiene a la mediana: busca el IC cuya frecuencia absoluta acumulada (FAA) sea la primera que excede el número que acabas de obtener. En el caso que nos ocupa, será el 3er intervalo, ya que las FAA de los intervalos 1 y 2 son menores que 50.
3. Halla el límite real inferior (LRI) del intervalo de clase que contiene a la mediana, calculando la semisuma entre el límite superior del intervalo que le antecede (LSa) —que es el 2º intervalo— y el límite inferior del propio intervalo (LIp).

$$LRI = \frac{LSa + LIp}{2} = \frac{30 + 29}{2} = 29.5$$

4. Determina la FAA del IC que antecede al que contiene a la mediana. En este caso es 37.
5. Resta el resultado anterior a la mitad de las observaciones: 50 – 37 = 13.
6. Ahora, calcula la amplitud del intervalo de clase que contiene a la mediana:

c = límite real superior (LRS) – límite real inferior (LRI)

LRS = [LS propio intervalo + LI intervalo siguiente]/2 = (34+35)/2 = 34.5

c = 34.5 – 29.5 = 5

7. Por último, f = 14.

Ahora, ya estás en condiciones de saber quién es la mediana del ejemplo:

$$med = LRI + \left[\frac{\frac{n}{2} - FAA}{f}\right] \cdot c = 29.5 + \frac{13}{14} \cdot 5 \approx 34$$

Así las cosas, ya puedes decir que la mediana de la serie en cuestión es de 34 años de edad, o que la edad mediana de la serie es 34 años.

Esta medida de resumen posee las propiedades siguientes:

a) Su cálculo es sencillo;
b) Siempre existe, y puedes calcularla a cualquier conjunto de datos numéricos;
c) Es única;
d) Se puede calcular en series con límites abiertos, excepto cuando la propia mediana caiga en un límite abierto, pero esto es sumamente improbable; y
e) No se afecta fácilmente por valores extremos.

La cuarta y quinta propiedades hacen que se prefiera esta medida sobre la media en situaciones en que la escala sea abierta o que existan valores aberrantes. Ahora bien, en la mayoría de los casos lógicamente, salvo los citados se prefiere conocer la media como medida de tendencia central.

Para ilustrar lo planteado en la quinta propiedad, volvamos al ejemplo de las edades de los niños del Círculo Infantil y su educadora. Si calculamos la mediana de esos datos, ésta sería:

Datos ordenados: 10, 12, 12, 14, 16, 18, 20, 34

Mediana: $(14 + 16) / 2 = 15$ meses, resultado que sí refleja con certeza el centro de los datos.

Quizá una desventaja imputable a la determinación de la mediana radica en el ordenamiento previo de las observaciones, faena que pudiese devenir tediosa y hasta impracticable de ser un número considerable de datos, pero recuerda las potencialidades que te brindan los softwares existentes en el mercado actual, que facilitan enormemente el trabajo[13] al calcular la mayoría de estos indicadores.

6.2.3. La moda

Ahora conocerás una medida realmente sencilla, tanto de determinar como de interpretar. Es muy intuitiva, y consiste en el valor, clase o categoría que aparece con más frecuencia en una serie de datos; o sea, es el que más se repite. Por ejemplo, si de seis pacientes, tres tienen 20 años, y los otros tienen 18, 21, y 25 respectivamente, entonces dirías que 20 años es la moda, o edad modal.

La mayor ventaja de la moda radica en que no requiere cálculo alguno, para beneplácito de algunos que no cuentan a las Matemáticas entre su círculo de amistades. Sin embargo, puede que no exista, e incluso puede no ser única. Por ejemplo, la serie 2, 5, 6, 6, 6, 7, 8, 8, 8, es una serie bimodal, pues cuenta con el seis y el ocho como modas.

6.3. Medidas de tendencia no central (Posición)

Como recordarás, las medidas estudiadas con anterioridad se nombran de tendencia central por representar el centro del conjunto de observaciones. Verás ahora un grupo de medidas que establecen una posición en una serie ordenada de datos, son una referencia a partir de la cual podrás decir: por encima de este valor están las ¾ partes de las observaciones, o algo por el estilo. De estas medidas, llamadas cuantiles por algunos, veremos a los cuartiles, deciles y percentiles.

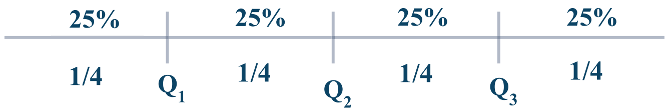
Medidas de tendencia central: Es encontrar en una serie de datos o en una distribución de frecuencias valores específicos, además proporcionan información resumida de la variable objeto de estudio.

[13] No tienes que acudir obligatoriamente a un paquete estadístico especializado: por ejemplo, Microsoft® Excel realiza cálculos como el que nos ocupa.

Nos facilitan información sobre la serie de datos que estamos analizando.

6.3.1. Los cuartiles

Los cuartiles, representados por la letra Q, son valores que dividen una serie ordenada de datos en cuatro partes iguales (cuartos), de tal suerte que por debajo del primer cuartil (Q1) se encuentra ¼ parte (el 25%) de los datos, y por ende, el 75% (las ¾ partes) está por encima de ese cuartil. La mitad de las observaciones cae por debajo del segundo cuartil y, lógicamente, la otra mitad está por encima de él; y las ¾ partes de los datos están por debajo del tercer cuartil, como muestra la figura.



Indudablemente, uno de estos nuevos amigos tiene un aire familiar con cierta conocida. ¿Ya te disté cuenta? Claro, es Q2, quien coincide con la mediana.

Para hallar los cuartiles de una serie, seguiremos un procedimiento similar al utilizado para conseguir la mediana.

Para una serie de datos simples o datos no agrupados, el cuartil en cuestión Q_i será el valor que ocupe la posición $\frac{n+1}{4}q_i$, siendo qi el cuartil deseado y n el total de observaciones.

Para encontrar la posición:

$$Q_1 = \frac{(n+1)}{4} \qquad Q_2 = \frac{2(n+1)}{4} \qquad Q_3 = \frac{3(n+1)}{4}$$

Si los datos están agrupados, entonces el cuartil será el resultado de calcular:

$$Q_i = LRI + \left(\frac{\frac{nq_i}{4} - FAA}{f}\right)c$$

donde:

q_i: es el cuartil deseado (i = 1, 2, 3);
LRI: límite real inferior del intervalo que contiene al cuartil;
n: total de observaciones;
FAA: frecuencia acumulada absoluta de la clase que antecede a la del cuartil;
f: frecuencia absoluta de la clase que contiene al cuartil;
c: amplitud del IC que contiene al cuartil.

Para calcular el primer cuartil, volvamos a auxiliarnos del ejemplo de las frecuencias cardiacas en ciertos pacientes ingresados.

Frecuencia Cardiaca	Pacientes	Frecuencia Absoluta

		Acumulada
21 – 40	1	1
41 – 60	17	18
61 – 80	38	56
81 – 100	40	96
101 – 120	4	100
Total	100	-

1. Busca en cual clase está el cuartil: $\frac{nq_i}{4} = \frac{100 \cdot 1}{4} = 25$; con este valor y siguiendo el procedimiento que seguramente recuerdas de la mediana, rápidamente sabrás que el cuartil está en el tercer intervalo (porque su frecuencia acumulada absoluta es la primera en sobrepasar 25).
2. Los LR de ese IC son: LRI = (60 + 61) / 2 = 60.5; LRS = (80 + 81) / 2 = 80.5. Luego: c = 80.5 – 60.5 = 20.
3. FAA = 18, f = 38.
4. Determina el cuartil:

$$Q_1 = 60.5 + \left(\frac{25 - 18}{38}\right) 20 = 64.2$$

Bien, ya tienes el valor del cuartil, solo tienes que interpretarlo adecuadamente, lo cual lograrás si leíste el comienzo del acápite. La cuarta parte de los pacientes, o el 25% de ellos, tiene 64 o menos latidos cardiacos por minuto; o puedes decir que el 75% (¾ partes de los pacientes) tiene más de 64 latidos cardiacos por minuto.

6.3.2. Los deciles

Veamos ahora los deciles. Muy parecidos a sus parientes los cuartiles, ellos son nueve (D1, D2…, D9) que dividen a una serie ordenada de datos en diez partes iguales (décimos). Por ejemplo, por encima de D6 hay cuatro décimos, quedando seis décimos debajo de él, y así por el estilo. Curiosamente, el quinto decil coincide con el segundo cuartil, y, por consiguiente, con la mediana.

La posición del decil está dada por la expresión $\frac{n+1}{10} d_i$ en una serie de datos simples, siendo n el total de observaciones, y d_i el decil deseado (i = 1, 2, 3, 4, 5, 6, 7, 8, 9).

$$D_1 = \frac{(n+1)}{10} \qquad D_5 = \frac{5(n+1)}{10} \qquad D_9 = \frac{9(n+1)}{10}$$

Para datos agrupados, usemos nuevamente el ejemplo de la frecuencia cardiaca. Los pasos para calcular el decil son en esencia iguales que si fueses a calcular el cuartil, haciendo los cambios siguientes:

$$D_i = LRI + \left(\frac{\frac{nd_i}{10} - FAA}{f}\right) c$$

donde:

D_i: es el decil deseado (i = 1, 2…, 9);
LRI: límite real inferior de la clase o intervalo que contiene al decil;
n: total de observaciones;

FAA: frecuencia acumulada absoluta de la clase que antecede a la del decil;
f: frecuencia absoluta de la clase que contiene al decil;
c: amplitud del IC que contiene al decil.

Calculemos el primer decil:

$\frac{nd_1}{10} = \frac{100 \cdot 1}{10} = 10$, por tanto, el decil está en el segundo intervalo.
LRI = 40.5; c = 20; FAA = 17.

$$D_1 = 40.5 + \left(\frac{10-1}{17}\right)20 = 51.01$$

Y ya puedes decir que la décima parte de los pacientes (el 10% de ellos) tenía 51 o menos latidos cardiacos por minuto; o lo que es lo mismo, el 90% tenía una frecuencia cardíaca de más de 51 latidos por minuto.

¿Eso era todo? Mira, todo resultado, por muy bien calculado que esté, por muy laborioso o sencillo que haya sido llegar hasta él, la interpretación debe estar avalada por un conocimiento previo del problema en cuestión. Realmente, no es improbable que alguien te replicara, no sin cierta sorna: «Médico, ¿pudiese Ud. explicarnos cómo nos las arreglamos para manejar frecuencias cardíacas impares?» A lo que tendrías que hacer de tripas corazón y rectificar la cifra, escogiendo para ello el número par precedente o siguiente.

6.3.3. Los percentiles

Los percentiles son observaciones que dividen a una serie ordenada de datos en cien partes iguales, motivo por el que hay 99 de ellos. Sí, ya sabemos que el término te resultó conocido apenas lo viste. Casi de la familia. Pues sí, se trata de los mismos percentiles que utilizas, por ejemplo, para conocer la evaluación nutricional de uno de tus bebitos. Bueno, ahora verás cómo se construye un percentil, procedimiento que conoces en parte porque es muy análogo al utilizado para determinar las medidas de posición estudiadas.

Para una serie de datos simples, la posición del percentil estará dada por $\frac{n+1}{100}p_i$, siendo n el total de observaciones, y pi el percentil deseado (i = 1, 2…, 99).

$$P_1 = \frac{(n+1)}{100} \qquad P_{50} = \frac{50(n+1)}{100} \qquad P_{99} = \frac{99(n+1)}{100}$$

Para datos agrupados, el percentil deseado será el resultado de computar:

$$P_i = LRI + \left(\frac{\frac{np_i}{100} - FAA}{f}\right)c$$

donde:

P_i: es el percentil deseado (i = 1, 2,…, 99);
LRI: límite real inferior de la clase o intervalo que contiene al percentil;
n: total de observaciones;

FAA: frecuencia acumulada absoluta de la clase que antecede a la del percentil;

f: frecuencia absoluta de la clase que contiene al percentil;

c: amplitud del IC que contiene al percentil.

Haciendo uso de nuestro fiel amigo, el ejemplo de las frecuencias cardiacas, calculemos el percentil 95 de la serie:

$\frac{np_{95}}{100} = \frac{100 \cdot 95}{100} = 95$, por lo que el percentil está en el 4º intervalo.

LRI = 80.5; FAA = 40; c = 20.

$$P_{95} = 80.5 + \left(\frac{95-56}{40}\right)20 = 100$$

Con este dato, puedes afirmar que el 95% de los pacientes examinados tenía una frecuencia cardíaca de 100 o menos latidos por minuto, o lo que es lo mismo, el 5% tenía frecuencias superiores a los 100 latidos por minuto.

6.4. Medidas de dispersión o variación:

Hasta el momento has visto algunas medidas que proporcionan información de una serie de datos numéricos, con la característica que un solo número es el encargado de esto. Quizá podría parecerte que con esas medidas sería suficiente para resumir y describir los conjuntos de los cuales proceden, sin embargo, múltiples circunstancias exigen la descripción de otros rasgos de los datos existentes. Verás por qué planteamos esto.

Volviendo al ejemplo de la medición de la frecuencia cardiaca (FC, epígrafe 4.1.1), imagina que el médico a cargo de la sala (que no fue quien hizo el estudio) realice el siguiente análisis: «tomando en cuenta la FC media, no tengo motivos para preocuparme por la salud de los pacientes, pues, en general, ostentan cifras dentro de límites normales; por ende, tenemos que encaminar nuestros esfuerzos hacia otros problemas».

¿Qué opinas acerca de esto? Sí, estamos de acuerdo contigo: en principio, no está nada mal, ya que interpretó correctamente el indicador. Pero, hay algo que debes recordar: la realidad es que, hasta donde él sabe, puede que la mayoría de los pacientes sea la que tiene frecuencias cardiacas en los límites considerados normales; pero puede que sea solamente la mitad, mientras la otra mitad permanezca en franca bradicardia (o taquicardia). Claro, este es uno de los casos "extremos", más el hecho de que sea algo poco probable no significa que sea imposible. De hecho, nota que, en el ejemplo, el intervalo de clase que ostenta mayor frecuencia absoluta (40) es el de 81 a 100 latidos por minuto, o sea, no es precisamente el que contiene a la media.

Veamos otra situación. Fíjate en estas dos series cuyas medias y medianas coinciden (53). Sin embargo, no podríamos decir que son semejantes, si tomamos en cuenta sus datos:

Serie A: 50, 51, 52, 53, 54, 55, 56

Serie B: 23, 33, 43, 53, 63, 73, 83

Por supuesto que no podemos. En la serie A los datos están más juntos, están más cercanos (distan una unidad entre sí), lo que no ocurre en la serie B, cuyos datos están mucho más alejados entre sí. Si analizas las distancias entre las observaciones y su media, en la serie A la primera y última observaciones están a tres unidades del centro de la serie, mientras que en la otra esas observaciones distan 30 unidades de su media. ¿Claro?

Ante una situación parecida, se necesitan otros parámetros acerca de la serie, como aquellos que miden cuán alejadas o agrupadas están las observaciones unas de otras o de la media.

El grado de agrupación o alejamiento de los datos de una serie es lo que recibe el nombre de variabilidad, variación, esparcimiento o dispersión de los datos, la cual puede ser absoluta o relativa. Con el fin de estudiarla, esto es, de conocer hasta qué punto las observaciones están agrupadas o esparcidas, la Estadística nos facilita las medidas de variabilidad, variación, esparcimiento o dispersión, las cuales verás en este epígrafe.

6.4.1. Medidas de variabilidad absoluta

6.4.1.1. La amplitud

Esta es la más sencilla de las medidas de dispersión. Consiste en la diferencia entre el mayor valor de la serie y el menor, o sea, restar ambos valores. Por ejemplo, la amplitud de los datos de la serie A vista en el acápite anterior es de 56 - 50 = 6, y la de la serie B es de 60. Ya conoces esta medida

Como ves, es muy fácil de determinar. Da una descripción rápida de la variabilidad de un grupo de observaciones. No es muy descriptiva de la misma porque sólo toma en cuenta los valores extremos de la serie, mira el siguiente ejemplo:

Serie A: 150, 160, 170, 180, 190, 200

Serie B: 150, 190, 197, 198, 199, 200

Resulta obvio que ambas tienen la misma amplitud (50), pero salta a la vista que ambas no tienen la misma dispersión. En la primera, los valores se sitúan de forma bastante dispersa entre los extremos; en la segunda la mayoría está cercana al valor mayor de la serie.

Si queremos describir de manera más eficiente la dispersión de un cúmulo de datos, tendremos que echar mano a otra medida. Veamos qué sorpresa nos depara la Sra. Estadística.

6.4.1.2. La desviación media

Al hablar de dispersión, lo hacemos la mayoría de las veces tomando en cuenta a la media aritmética del conjunto de observaciones. Así, cuando decimos que la dispersión de una serie es pequeña, es porque los datos están agrupados en la cercanía de su media, siendo grande si los datos están alejados de ella. Esto sienta las bases para utilizar como referencia a las distancias para referirnos a la dispersión, o sea, que sería procedente definirla en función de las distancias que existen entre los números y su media. Ahora bien, estas distancias pueden ser vistas como la desviación entre los elementos en cuestión; con otras palabras, si hay mucha distancia, decimos que se desvió mucho el número de la media, y así por el estilo.

Si promediásemos las desviaciones entre cada número y la media, o sea, sumarlas y dividirlas por la cantidad de números de la serie, obtendríamos una medida de la variación promedio del conjunto de datos dada por:

$$\frac{\sum_{i=1}^{n}(X_i - \overline{X})}{n}$$

Por desgracia, realizar este cálculo te sería tan provechoso como no hacer ninguno, pues el resultado final siempre es cero[14], debido a razones matemáticas establecidas. Por ejemplo, considera la siguiente serie: 2, 3, 4, 5, 6. Su media es 4, y si calculásemos lo planteado:

$$\frac{(2-4)+(3-4)+(4-4)+(5-4)+(6-4)}{5} = \frac{(-2)+(-1)+0+1+2}{5} = \frac{0}{5} = 0$$

La solución a este inconveniente es hallar la diferencia modular de las desviaciones, de esa manera sólo tomarás en cuenta la magnitud de dichas desviaciones, esto es, hallar el módulo de las diferencias. De esta forma, estarás calculando la desviación media (DM) o desviación promedio, cuya fórmula para datos simples es:

$$DM = \frac{\sum_{i=1}^{n}|X_i - \overline{X}|}{n}$$

donde:

X_i son las observaciones de la serie (i = 1,…, n);

$\overline{X}$: media aritmética de la serie;

n: total de observaciones.

Ilustremos lo antedicho calculando la desviación media de las siguientes mediciones del peso (en libras) que corresponden a cinco estudiantes de tu área de salud.

150.5, 180.8, 145.3, 127.9, 130.5

Ante todo, calcula la media:

$$\overline{X} = \frac{150.5 + 180.8 + 145.3 + 127.9 + 130.5}{5} = 147\ libras$$

Luego halla las desviaciones de cada observación con respecto a su media:

Xi	|Xi - |
150.5	3.5
180.8	33.8
145.3	1.7
127.9	19.1

[14] Estriba en que la media de las desviaciones es justamente la media misma. Como escapa a los propósitos de este curso entrar en formalismos matemáticos, si estás interesado puedes acudir a la literatura especializada para indagar al respecto.

130.5	16.5
	74.6

Finalmente, DM = 74.6 □ 5 = 14.92 libras.

Calculada la medida deseada, puedes decir que, en promedio, el peso de los estudiantes se desvía casi 15 libras del promedio general de 147 libras.

Si los datos de que dispones están agrupados, entonces calcularás la desviación media de la siguiente forma:

$$DM = \frac{\sum_{i=1}^{n} \left| MC_i - \overline{X} \right| \cdot f_i}{n}$$

donde

MC_i: marca de clase de cada IC;
f_i: frecuencia absoluta de cada intervalo de clase;
$\overline{X}$: media aritmética de la serie;
n: total de observaciones.

Utilizando nuevamente el ejemplo del acápite 4.1.1, calculemos la desviación promedio de esos datos.

($\overline{X}$ = 76.3 latidos/minuto)

FC	Pacientes (f_i)	MC_i	$\lvert MC_i - \overline{X} \rvert fi$
21 – 40	1	30.5	(30.5 – 76.3) = 45.8
41 – 60	17	50.5	(50.5 – 76.3) = 438.6
61 – 80	38	70.5	(70.5 – 76.3) = 220.4
81 – 100	40	90.5	(90.5 – 76.3) = 568.0
101 – 120	4	110.5	(110.5 – 76.3) = 136.8
Total	100		1409.6

Entonces, DM = 1409.6/100 = 14.09 latidos por minuto, lo que significa que las frecuencias cardiacas de los pacientes de alejan, en promedio, 14 latidos por minuto de la media de 76.

Particularicemos en algo: si se tratase de una serie con valores aberrantes, debiste calcular la mediana en vez de la media; entonces debes usar la mediana para calcular la desviación media, sustituyendo la media en la fórmula por la mediana.

La desviación media es una medida que se utiliza poco en la práctica, sobre todo si son muchos datos o si éstos tienen muchos lugares decimales, pero principalmente debido a razones de índole matemática que no abordaremos en este curso. De todas formas, optamos por dártela a conocer con el fin de que lograras entender cabalmente las medidas que verás a continuación.

6.4.1.3. La varianza y la desviación estándar

Como recuerdas, para calcular la desviación media te viste obligado a utilizar las diferencias modulares para obtener un resultado válido, pues de lo contrario no hubieses obtenido nada. Pues bien, existe otra medida que se vale de elevar al cuadrado las desviaciones de los datos con respecto a su media. Dicha medida recibe el nombre de varianza o variancia. Se denota por los símbolos S^2 ó σ^2 (letra griega sigma minúscula al cuadrado)

Varianza de la población $(\boldsymbol{\sigma^2})$ La varianza se define como la media aritmética de los cuadrados de las diferencias de los datos con su media aritmética.

Varianza de la Muestra $(\boldsymbol{S^2})$ La fórmula de la varianza de la muestra es diferente a la de varianza de la población.

Su cálculo (para datos simples) se verifica según la fórmula:

	Varianza	Desviación estándar
Población	$\sigma^2 = \frac{\sum_{i=1}^{N}(X_i - \mu)^2}{N}$	$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}}$
Muestra	$S^2 = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{n-1}$	$S = \sqrt{S^2} = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \overline{X}\right)^2}{n-1}}$

Si los datos están agrupados, el cálculo de la varianza se realiza mediante la siguiente fórmula:

	Varianza	Desviación estándar
Población	$\sigma^2 = \frac{\sum_{i=1}^{k}(MC_i - \mu)^2 f_i}{N}$	$\sigma = \sqrt{\sigma^2} =$
Muestra	$S^2 = \frac{\sum_{i=1}^{n}\left(MC_i - \overline{X}\right)^2 f_i}{n-1}$	$s = \sqrt{s} =$

donde:

k: número de clases.
f_i: frecuencia absoluta de cada clase, es decir, el número de elementos que pertenecen a dicha clase.
MC_i: marca de clase de cada IC. Es el punto medio del límite inferior y del límite superior.
σ2: varianza de la población.
σ: desviación estándar de la población.
μ: media de la población.
S^2: varianza de la muestra.
x_i: son las observaciones de la serie (i = 1,…, n);
$\overline{X}$: media aritmética de la serie (muestra)
n: total de observaciones.
s: desviación estándar de la muestra.

Esta medida logra describir adecuadamente la dispersión del conjunto de datos, pero tiene un inconveniente: su resultado se expresa en unidades cuadradas, algo harto engorroso y difícil de entender en la mayoría de las situaciones prácticas, y por demás disonante en relación con la medida de tendencia central utilizada. Sería algo así como años cuadrados, o pesos cuadrados (¿?).

A fin de eliminar este aparente escollo, puedes hallar la raíz cuadrada positiva del número obtenido, con lo que tendrás de vuelta a las unidades originales, obteniendo así una medida denominada desviación típica o estándar, y es la medida de variación más ampliamente utilizada en el mundo de las estadísticas. Su símbolo es S (por ser la raíz cuadrada de la varianza), aunque se utiliza también DS (desviación standard) o SD (standard desviation). Tiene, además, la ventaja de que hasta las calculadoras de bolsillo las científicas, claro está la calculan, y casi la totalidad de los paquetes estadísticos existentes en el mercado del software.

La desviación estándar es la raíz cuadrada positiva de la varianza

Vamos a calcular la desviación estándar de la serie del ejemplo de los pesos (en libras) del acápite anterior.

150.5, 180.8, 145.3, 127.9, 130.5

$\overline{X}$ = 147 libras

Xi	(Xi -$\overline{X}$)2
150.5	12.25

180.8	1142.44
145.3	2.89
127.9	364.81
130.5	272.25
	1794.64

S2 = 1794.64 ÷ 5 = 358.93 libras^2

$S = \sqrt{358.93} = 18.94$ libras

Con el resultado obtenido puedes decir que, en promedio, la mayoría de los datos se desvían de la media en casi 19 libras.

En gran parte de las situaciones del mundo biomédico, y basándose en elementos de la Estadística Inferencial, se puede utilizar la desviación estándar y la media para construir intervalos en los que se mueve la mayor parte de los datos. Por ejemplo, en el intervalo cuyo extremo inferior sea – SD y el superior sea + SD, o sea, [– SD; + SD], se encuentra cerca del 68% del total de las observaciones. Si construyes el intervalo con el duplo de la SD: [– 2·SD; + 2·SD], entonces ahí estará cerca del 95% de los datos; y si utilizas el triplo de la SD, el intervalo contendrá entonces a más del 99% (99.73%) de las observaciones.

Los elementos utilizados en el cálculo son los mismos que ya dijimos. El cálculo de la desviación estándar se limita a extraer la raíz cuadrada del número obtenido arriba.

Calculemos la desviación estándar para el ejemplo de las frecuencias cardiacas (media = 76.3 años).

FC	Pacientes	(MC - $\overline{X}$)^2 fi
21 – 40	1	$(30.5 – 76.3)^2$ = 2097.64
41 – 60	17	$(50.5 – 76.3)^2$ = 11315.88
61 – 80	38	$(70.5 – 76.3)^2$ = 1278.32
81 – 100	40	$(90.5 – 76.3)^2$ = 8065.60
101 – 120	4	$(110.5 – 76.3)^2$ = 4678.56
Total	100	27436.00

S^2 = 27 436 ÷ 100 = 274.36

S = 16.5

Esto significa que, en promedio, la mayoría de los pacientes se aleja de la media en 16 latidos por minuto.

Suponiendo que a esta población se le pueden aplicar los porcentajes mencionados, entonces podríamos decir que aproximadamente el 95% de los pacientes tenían una frecuencia cardiaca que oscilaba entre 44 y 109 latidos por minuto.

Si la serie posee valores aberrantes, te viste obligado a utilizar la mediana, por lo que ahora debes sustituir la media por la mediana en la fórmula para calcular la varianza y la desviación típica.

6.4.2. Medidas de variabilidad relativa

En muchas ocasiones es necesario comparar la dispersión entre dos o más conjuntos de datos, y sucede que las variables tienen diferentes unidades de medida. Con las medidas de dispersión estudiadas no podrás llegar a una conclusión válida acerca de las desviaciones de los datos. Incluso, aun cuando se trate de una sola unidad de medida, las mediciones pueden variar considerablemente: si comparas la desviación típica de la estatura de los niños de 5-14 años de tu área con la de los estudiantes de preuniversitario, es muy probable que esta última sea mayor que la primera, debido a que las tallas sean mayores per se, y no porque la variabilidad sea mayor precisamente.

6.4.2.1. El coeficiente de variación.

Ante casos como el descrito con anterioridad, es imprescindible contar con una medida de variabilidad relativa, como el coeficiente de variación (CV), que expresa a la desviación típica como porcentaje de la media, y su cálculo se realiza mediante:

$$\mathrm{CV} = \frac{\mathrm{SD}}{\overline{X}} 100$$

Observa que, por tener la desviación estándar y la media las mismas unidades de medida, quedan canceladas dichas unidades, de ahí que el coeficiente de variación no tenga unidades propias15, lo que facilita la comparación.

En el ejemplo siguiente, si comparas las desviaciones estándares de los dos grupos, pudieras creer que ambos tienen igual dispersión:

Grupo 1: media = 60 cm; SD = 4 cm

Grupo 2: media = 170 cm; SD = 4 cm

Sin embargo, si calculas la medida recién conocida, entonces: $CV_1 = 6.6$, $CV_2 = 2.3$. Al contrastarlos, ves algo bien diferente, pues en realidad el grupo 1 tiene casi tres veces más dispersión que el grupo 2.

[15] Con otras palabras, el coeficiente de variación es adimensional.

Resumen

En este tema aprendiste que:

1. Las medidas de resumen para variables cuantitativas suelen dividirse en: de Tendencia Central, de Posición y de Variabilidad; resultando conveniente no utilizar las primeras de forma aislada, sino acompañadas de alguna medida de la dispersión.
2. La media aritmética es la medida de tendencia central más utilizada; en tanto que la desviación estándar sobresale por su uso entre las medidas de variabilidad.
3. Cuando las distribuciones son asimétricas o hay intervalos abiertos, se prefiere el cálculo de la mediana al de la media.
4. Las medidas de posición son los cuartiles, deciles percentiles, sirven para encontrar puntos específicos en los datos no agrupados o agrupados que estemos manejando en nuestra investigación.
5. Cuando las distribuciones son simétricas, coinciden la media, la mediana, la moda, el segundo cuartil, el quinto decil y el percentil cincuenta.
6. Se prefiere utilizar como medida de dispersión a la desviación media cuando se usa como medida de tendencia central la mediana. Si utilizaste la media, generalmente se elige la desviación estándar.

Bibliografía

1. Rubén Quesada M, Torres Delgado JA, Bayarre Vea H. Informática Médica La Habana: Ciencias Medicas; 2004.

2. Moya Calderon R. Estadistica Descriptiva Lima: San Marcos S.A.; 1999.

3. Buckland WR, Kendall MG. Diccionario de Estadistica Hardcover , editor. Madrid: Piramide; 1980.

4. Jhonson R, Kuby P. Estadística elemntal. 10th ed. Perez FdJC, editor. México: CENGAGE Learning; 2008.

5. Spiegel MR, Stephens LJ. Estadística. 4th ed. Vázquez PER, editor. México: Mc Graw Hill; 2009.

6. Bayarre H, Hersford R. Estadística Descriptiva y Estadística de Salud Ecimed , editor. La Habana: Ciencias Medicas; 2005.

7. Jesus CdlRAd. Bioestadística. 2nd ed. México: El Manual Moderno; 2008.

Anexos

MAPA MENTAL DE VARIABLES

Cualitativa

Las variables cualitativas se refieren a características o cualidades que no puedan ser medidas con números. Podemos distinguir de dos tipos

Nominal

Presenta modalidades no numéricas que no admiten un criterio de orden

Ejemplo

El estado civil: soltero, unión libre, casado, divorciado, y viudo

Ordinal

Presenta modalidades no numéricas, en las que existe un orden

Ejemplo

Nivel de escolaridad: primaria, secundaria, universitaria

Cuantitativa

Las variables cuantitativas es la que se expresa mediante un número, por tanto, se pueden realizar operaciones aritméticas con ella podemos distinguir de dos tipos

Discreta

Es aquella que solo puede tomar un numero finito de valores entre dos valores cualesquiera de una característica

Ejemplo

Numero de hijo de una familia: 0,1,3,4,5

Continua

Es aquella que puede tomar un número infinito de valores entre dos valores cualesquiera de una característica

Ejemplo

La estatura en pacientes tan tendidos en hospital X: 1.71,1.77, 1.69 …

MAPA MENTAL: ESCALAS DE CLASIFICACIÓN Y COMPUTO DE DATOS

MAPA MENTAL: TIPOS DE FUENTES Y FUENTES DE INFORMACIÓN

Las Fuentes de information son todos los conocimientos que de una forma u otra difunden los conocimientos propios de un área

Primarias

Informes técnicos

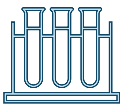
Diarios

Experimentos

Poemas

Personal

Discursos



Terciarias

Almanaques

Cronologías

Directorios

Manuales

Libros de instrucciones

Secundarios

Artículos de revista

Crítica literaria y cometarios

Enciclopedias

Diccionarios

Anuarios

Bibliografías

MAPA MENTAL: TIPOS DE ENTREVISTA

- Entrevista
  - Por el número de personas quien se dirige
    - Individual
    - Grupo Focal
  - Por el nivel de estructuración
    - No estructurada: se realiza haciendo una o dos preguntas generales
    - Semiestructurada: Se elabora una guía de preguntas abiertas que orientan la entrevista.
    - Estructurada: Se elabora una guía de preguntas muy puntuales sobre el tema