

CUINAP | Argentina

Año 1 • 2020 | Cuadernos del INAP

Manual de estadística básica para no estadísticos

Jorge Zappino

13

Capacitar e investigar para fortalecer las capacidades estatales

CUINAP | Argentina

**Manual de estadística básica para
no estadísticos**

Jorge Zappino

13



Autoridades

Dr. Alberto Ángel Fernández

Presidente de la Nación

Lic. Santiago Andrés Cafiero

Jefe de Gabinete de Ministros

Dra. Ana Gabriela Castellani

Secretaria de Gestión y Empleo Público

Dr. Alejandro Miguel Estévez

Director Institucional del INAP

Índice

Prólogo	6
1. Introducción	9
2. Organización de los datos	26
3. Elaboración de gráficos	36
4. Estadística descriptiva	47
5. Indicadores, relaciones entre variables y cifras	88
6. Referencias bibliográficas	105

Los Cuadernos del INAP

El Instituto Nacional de la Administración Pública es un organismo público con dos funciones centrales: la capacitación y la investigación para fortalecer las capacidades estatales. Con distintos vaivenes y matices, en general vinculados con el devenir del país, las tareas llevadas a cabo han permitido desarrollar ambas finalidades, aunque puede señalarse como falencia la limitada interacción histórica entre ellas.

A partir de esta situación, cobra especial relevancia el texto de Jorge Zappino, docente e investigador del INAP, quien redactó un manual de estadística especialmente diseñado para la enseñanza en el ámbito de los cursos de capacitación de la administración pública. La adaptación de los contenidos tradicionales de estadística para esta situación específica tiene la virtud, entre otras, de resultar más atractiva para los cursantes, lo cual facilita la transmisión de conocimientos.

El avance tecnológico ha generado cambios que afectan a todas las actividades socioeconómicas. La administración pública no constituye una excepción y, en la actualidad, se trabaja con elementos muy distintos con respecto al pasado. La disponibilidad de información ha crecido exponencialmente y resulta necesario, y hasta imprescindible, potenciar la capacidad de procesamiento

de datos a través de un mejoramiento del instrumental estadístico. Por ello este manual es un aporte importante para la formación en el sector público.

Por último, cabe mencionar que el alcance de esta publicación no se limitará a su aplicación en los cursos del INAP, sino que se presume que será utilizada en otros espacios del sector público (provincias, municipios) o, eventualmente, en el sector privado. Además, constituye un ejemplo de articulación entre la enseñanza y la investigación que promoverá iniciativas similares, en un efecto colateral positivo destacable.

En síntesis, resulta una satisfacción poder presentar un trabajo que impulse el vínculo entre la producción de conocimientos, la enseñanza y, en última instancia, la toma de decisiones mejor informada en las distintas esferas del aparato estatal.

Dr. Juan Ignacio Doberti
Director de Investigaciones y Publicaciones del INAP

Manual de estadística básica para no estadísticos

Jorge Zappino

Jorge Zappino es licenciado en Ciencia Política por la Universidad de Buenos Aires (UBA). Magíster en Historia Económica y de las Políticas Económicas (UBA) y Magíster en Generación y Análisis de Información Estadística (UNTREF). Actualmente, es investigador del Instituto Nacional de la Administración Pública (INAP). Ejerció como docente universitario en la UBA y desarrolló diversas actividades en otras universidades públicas y privadas del país.

Resumen

Se trata de un manual de estadística básica especialmente preparado para los/as empleados/as públicos.

Palabras clave

Manual, estadística básica, capacitación, empleo público.

Abstract

This is a basic statistics manual specially prepared for public employees.

Key words

Manual, basic statistics, training, public employment.

Introducción

1.1. Concepto de estadística

Podemos comenzar por lo que nos dice el *Diccionario de la Lengua Española* (2019): «Estudio de los datos cuantitativos de la población, de los recursos naturales e industriales, del tráfico o de cualquiera otra manifestación de las sociedades humanas». En otra acepción, explica que es la «Rama de la matemática que utiliza grandes conjuntos de datos numéricos para obtener inferencias basadas en el cálculo de probabilidades».

No obstante lo expuesto, no resulta simple obtener una única definición de «estadística» que conforme a todos sus adherentes. En la infinidad de libros dedicados a la materia, existen decenas de definiciones, y algo semejante sucede con otros términos relativos a esta disciplina, los que se irán incorporando paulatinamente al vocabulario del curso.

En resumen, y adoptando una definición propia, la estadística debe ser interpretada simplemente como una de las mejores herramientas para tomar decisiones en condiciones de incertidumbre. Es un conjunto de instrumentos que se utilizan para recopilar, presentar y caracterizar la información, para brindar soporte al análisis de datos y al proceso de

toma de decisiones. La estadística nos permite predecir fenómenos con mayor grado de exactitud.

1.2. Fuentes estadísticas

Las fuentes estadísticas son las que nos proveen los datos necesarios para que podamos procesarlos y obtener información útil. Diversos organismos brindan datos estadísticos, como, por ejemplo:

- INDEC (Instituto Nacional de Estadística y Censos).
- Direcciones provinciales de estadística.
- BCRA (Banco Central de la República Argentina).
- ONU (Organización de las Naciones Unidas).
- OMS (Organización Mundial de la Salud).
- FMI (Fondo Monetario Internacional).
- BM (Banco Mundial).
- BID (Banco Interamericano de Desarrollo).

Además, es necesario destacar que nosotros mismos podemos ser recolectores y creadores de datos a través de nuestras investigaciones, pero también es cierto que resulta más costoso y lleva mayor tiempo realizarlo.

1.3. Población o universo y muestra

En estadística, a menudo, hablamos de *población* y de *muestra*. Supongamos que necesitamos realizar un estudio sobre la proporción de personas entre 30 y 45 años de edad que habitan en la Ciudad Autónoma de Buenos Aires (CABA). Para hacerlo, podríamos interrogar a todas las personas de esa franja etaria (*población* o *universo*), pero

esto resulta sumamente complicado, en primer lugar, porque no sabríamos dónde encontrarlos a todos y, en segundo lugar, por el alto costo económico y el tiempo que demandaría dicho relevamiento. Por lo tanto, debemos encontrar una manera que, sin ser exacta, nos aproxime en mayor grado a la realidad de lo que está ocurriendo y así buscaremos a algunas personas (no a todas). Este relevamiento parcial es lo que llamamos *muestra*.

La *población* constituye el universo bajo estudio, es decir, el conjunto de elementos con ciertas características comunes, y puede presentarse de dos maneras:

- Población finita: se trata de una población cuyo número de elementos es numerable. Por ejemplo: la población de un municipio.
- Población infinita: se trata de una población cuyo número de elementos no se puede determinar por ser excesivamente grande. Por ejemplo: la cantidad de estrellas en el universo.

En estadística, frecuentemente, se recurre a muestras, es decir, se analiza una parte de la población, del universo bajo estudio. En nuestro caso, la muestra estaría conformada por las personas que seleccionamos al azar para hacerles las preguntas, en tanto que la población serían todas las personas entre 30 y 45 años de edad de la CABA.

Lo importante en una muestra es que sea representativa, es decir que los resultados que podamos obtener, luego de analizar los datos recogidos, sean representativos de la población.

Esquemáticamente, quedaría de la siguiente manera:



Cabe destacar que, generalmente, se simboliza con «n» el tamaño de la muestra y con «N», el tamaño de la población.

1.3.1. Componentes del muestreo

Marco de la muestra: es el registro de todos los ítems de la población (1, 2,..., n) de la que se extrae la muestra.

Unidad de la muestra: es el elemento al que habrá que dirigirse para obtener los datos de interés.

1.3.2. Clases de muestras

Básicamente, el muestreo puede realizarse de dos formas: probabilística o no probabilística.

- a) Muestreo probabilístico o al azar: los elementos de la muestra se escogen por métodos estadísticos basados en la teoría de la probabilidad, de forma que se pueda determinar la probabilidad

de que un elemento pertenezca a la muestra. Es el muestreo más usado y más científico.

Con este muestreo:

- Se puede realizar inferencia estadística.
- Se puede detectar el error.
- Se da la representatividad.
- Se pueden establecer intervalos de confianza.

A su vez, el muestreo probabilístico puede realizarse de dos maneras:

- Con reposición o reemplazo: al seleccionar las sucesivas unidades muestrales, las anteriores y ya seleccionadas vuelven a formar parte de la población de la que se extrae la muestra. Es decir que un mismo elemento puede ser seleccionado más de una vez.
 - Sin reposición o reemplazo: una vez que una unidad de muestra es seleccionada, ya no puede volver a serlo, debido a que queda fuera de la población sobre la cual se extrae la muestra.
- b) Muestreo no probabilístico: no es posible fijar la probabilidad de selección de cada individuo o ítem ni el nivel de confianza, al no tener los elementos de la población una probabilidad de elección definida.

1.4. Errores en la investigación

Hay dos tipos de error que se pueden cometer a la hora de realizar una investigación:

- Error aleatorio: es imputable al azar y se lo puede controlar. El tamaño de la muestra está relacionado con el error aleatorio, y tanto es así que, si deseamos disminuir el error aleatorio, debemos aumentar el tamaño de la muestra y, por el contrario, si queremos disminuir la muestra, ya sea por razones económicas o de tiempo, debemos tener presente que el margen de error a la hora de estimar será mayor.
- Error sistemático o ajeno al muestreo: es ajeno al azar y, por lo tanto, no es controlable, ya que depende de factores no estadísticos, como, por ejemplo, la elección de técnicas de investigación incoherentes, preguntas mal redactadas, ejecución incorrecta de los análisis.

1.5. Estimadores y parámetros

- Un *estimador* es una medida resumen que se calcula para describir una característica de una población con base en los datos extraídos de la muestra. Se utiliza para aproximarse al verdadero valor de los parámetros poblacionales.
- Un *parámetro* es una medida resumen que se calcula para describir una característica de una población.

1.6. Divisiones de la estadística

La estadística suele dividirse en dos grandes campos:

- **Estadística descriptiva:** se trata de la parte de la estadística que se encarga de describir los datos de un fenómeno ya ocurrido y cuyo comportamiento conocemos.
- **Estadística inferencial:** se trata de la parte de la estadística que permite obtener, a partir de los resultados provenientes de la estadística descriptiva y con cierto grado de confianza, conclusiones generales.

1.7. El método estadístico

El método por el cual se procesan los datos abarca las siguientes etapas:

- 1 **Recolección de datos:** es la etapa inicial. Estos datos brindan información sobre las características de los individuos observados.
- 2 **Agrupamiento de datos y gráficos:** los datos recogidos en la etapa 1 deben ser convenientemente ordenados y agrupados, es decir, dispuestos en tablas o gráficos para facilitar su lectura, de modo que se puede tener una rápida interpretación de los hechos.
- 3 **Medición de datos y análisis:** aquí comienza a actuar la matemática. Se observa si los datos tienden a centrarse en torno de algunos valores denominados *parámetros de posición* (media, mediana, moda). Luego, se analiza la dispersión de esos valores respecto de esos parámetros de posición y surgen así los *parámetros de dispersión* (desviación media, varianza, desvío estándar).

- 4 Inferencia estadística: mientras la estadística descriptiva trabaja con todos los individuos de la población, la estadística inferencial lo hace con muestras, subconjuntos formados por algunos individuos de la población. A partir del estudio de la muestra se pretende inferir aspectos relevantes de toda la población. Cómo se selecciona la muestra, cómo se realiza la inferencia y qué grado de confianza se puede tener en ella son aspectos fundamentales de la estadística inferencial, para cuyo estudio se requiere un alto nivel de conocimientos de estadística, probabilidad y matemática. La inferencia estadística puede permitir, en algunos casos, predecir el comportamiento futuro de la población observada.

1.8. Datos y variables

En el proceso de observación se registra para cada elemento una característica y esta constituye un *dato*. Una *variable* es cualquier característica que varía de una unidad a otra.

Por ejemplo: si se quiere conocer la nota que obtuvieron los participantes de un curso de estadística dictado por el INAP y se le pregunta a un participante por su calificación y responde 7, ese es un dato. A su vez, la variable será la nota del curso, que puede asumir valores desde el 0 hasta el 10.

1.8.1. Unidades estadísticas

Debe entenderse por unidades estadísticas al «objeto de un estudio estadístico». Son aquellas sobre las cuales se realizan las observaciones o mediciones, o que responden las consultas que se realizan. Por ejemplo:

- Las personas: en un censo de población, en una encuesta de opinión, etc.
- Las empresas: en un censo económico, en una encuesta sobre el sistema de riesgos de trabajo, en un estudio sobre insumos de producción, etc.
- Los países: en un estudio sobre exportaciones agropecuarias, en una encuesta sobre la natalidad de la población, etc.

1.8.2. Variables estadísticas

Se define como variable estadística a «todo aquello pasible de tomar valores diferentes de unidad en unidad».

Ejemplo

Unidades estadísticas	Variables
Personas	edad, estado civil, cantidad de hermanos, etc.
Empresas	producción (en pesos o toneladas), rama de actividad, tamaño, etc.
Países	exportaciones (en millones de dólares o toneladas), nivel de riesgo país, etc.

1.8.3. Clasificación de variables

Las variables que estudiamos en estadística pueden ser de dos tipos:

- Las variables *cuantitativas* son aquellas cuyos valores pueden expresarse como cantidades numéricas. Estas, a su vez, pueden ser:
 - Discretas: están relacionadas con números enteros y se asocian a procesos de conteo. Por ejemplo: la cantidad de trámites realizados por una oficina de ANSES durante un día determinado.

- Continuas: están vinculadas con los números reales o con procesos de medición. Por ejemplo: el monto total de salarios pagados por un organismo en un mes determinado.
- Las variables *cualitativas* son aquellas que pueden clasificarse, pero no medirse. Estas, a su vez, pueden ser:
 - Nominales: aquellas cuyos resultados no expresan dimensión ni orden entre las variables. Por ejemplo: estado civil, rama de actividad económica, rubro principal de exportaciones, etc.
 - Ordinales: aquellas cuyos resultados expresan dimensión u orden entre las variables. Por ejemplo: nivel de educación alcanzado, tamaño, nivel de riesgo país, etc.

1.9. La medición

Medir es asignar un valor numérico o un símbolo a un fenómeno observado mediante la utilización de ciertas técnicas o métodos. Hay casos en que esto no reviste mayor dificultad, como cuando se mide la cantidad de habitantes o de viviendas en una población. El problema surge cuando se quiere expresar numéricamente actitudes, opiniones, ideales o sentimientos de un determinado grupo social.

1.9.1. Escalas de medición

Según las operaciones matemáticas que se pueden realizar con los valores asignados a los fenómenos observados, es posible distinguir cuatro niveles o escalas de medición: nominal, ordinal, por intervalos o proporciones y por razón.

Esto permite, a su vez, clasificar los datos sociales.

1.9.2. Escala nominal

Es el nivel más bajo de medición y consiste en nombrar con lenguaje común los objetos con el único fin de clasificarlos. Por ejemplo, se describe una población en un determinado momento histórico: había 2000 habitantes en una superficie de 5 km², 150 casas, 1 iglesia, 1 plaza, 2 escuelas, etc. En este caso, se inspeccionan los objetos físicos, se les adjudican nombres y se suma el número existente de cada clase.

Una clasificación nominal tiene las siguientes características: el orden en que aparece la información no implica jerarquía de importancia. Puede aparecer primero la cantidad de iglesias o el número de habitantes, lo que no significa que un dato sea más importante que otro. Las categorías bajo las cuales se ordenan los conceptos son excluyentes. Un objeto no puede pertenecer a dos categorías: no se podría hablar de casas, iglesias e inmuebles, porque tanto las casas como las iglesias entrarían en la categoría de inmuebles. Todo individuo del ámbito analizado debe caer bajo alguno de los conceptos de la investigación. No puede haber un elemento que no responda a ninguna categoría.

1.9.3. Escala ordinal

Podemos establecer un orden entre las distintas categorías; indicar, por ejemplo, que una categoría está formada por partes que son más grandes, más antiguas, más importantes que las partes comprendidas en otras.

Si se puede establecer una relación de orden entre las categorías, decimos que la escala de medición es ordinal. Esta escala tiene un nivel superior al nominal. Por ejemplo, si tomamos una población y establecemos tres

categorías por niveles salariales, clase alta, clase media y clase baja, podemos tener el siguiente detalle:

Clase alta: 15.000 personas

Clase media: 50.000 personas

Clase baja: 70.000 personas

Primero, hay una clasificación nominal (contar cuántos ganan un determinado salario y agruparlos en clases) y, segundo, una clasificación ordinal que establece una jerarquía en la que ordenamos las personas de clase alta, las de la clase media y las de la clase baja.

Así como vimos que en la clasificación nominal no importa el orden, en la ordinal este es fundamental.

1.9.4. Escala por intervalos o proporciones

Cuando se tienen datos históricos en forma ordinal, a veces, también es posible clasificarlos por intervalos o proporciones. Esto se puede hacer si, además del orden, se conoce el tamaño de los intervalos entre ellos, como, por ejemplo, las cifras electorales, la composición de una cámara, las estadísticas de población.

Analicemos el siguiente caso de la conformación de la Honorable Cámara de Diputados de la Nación en 2020:

Juntos por el Cambio: 119

Frente de Todos: 109

Consenso Federal: 7

Frente Cívico por Santiago: 7

Córdoba Federal: 4

Unidad Justicialista: 3
 Partido Renovador de la Concordia: 3
 Frente de Izquierda: 2
 Juntos Somos Río Negro: 1
 Movimiento Popular Neuquino: 1
 Somos Mendoza: 1

Esto permite determinar qué partido político, coalición o alianza tiene más legisladores, pero además saber cuántos más. Tenemos una unidad fija —número de legisladores— para medir la fuerza de los partidos políticos. Cuando existe la posibilidad de que esa unidad de medida tenga un punto cero —caso del partido sin legisladores—, se tiene una proporción. Cuando no existe el punto cero para la unidad de medida, los datos son intervalos y no existe una escala de proporciones. Un ejemplo es la medición de la riqueza, ya que es arbitrario establecer el punto cero, es decir, no tener riqueza.

Gran parte de los datos sociales son proporciones. Si bien hemos hecho la distinción entre intervalos y proporción, muchas veces, se los utiliza indistintamente en la práctica.

1.9.5. Escala por razón

La escala por razón es la más rica en cuanto a la información que proporciona, porque además de poseer todas las propiedades de las escalas anteriores —igualdad-desigualdad (escala nominal), orden (escala ordinal) y distancia (escala por intervalo)—, tiene la ventaja de tener un cero absoluto que indica carencia absoluta del atributo medido.

Esta es la escala más habitual para medir propiedades físicas, como el peso, la altura, el volumen, pero es difícil encontrar variables psicológicas

que puedan medirse con esta escala. Por eso, podemos afirmar que el que mide 180 cm es el «doble» de alto que el que mide 90 cm, pero nunca que un CI de 180 es el doble de otro CI de 90, a menos que la variable inteligencia la podamos medir con escala por razón, en la que el punto cero represente al que carece «absolutamente» de inteligencia.

En consecuencia, la escala por razón incorpora las características del sistema numérico de las escalas anteriores y les añade la propia, que es la igualdad de razones. Al tener un origen de escala absoluto, la única transformación admisible para la escala por razón es del tipo: $g(x) = bx$ siendo $b > 0$. Esta transformación supone simplemente un cambio en la unidad de medida de la escala, preservando siempre su origen.

Los datos de escala por razón se definen como un tipo de datos cuantitativos que se caracterizan por un punto de cero absoluto, lo que significa que no hay ningún valor numérico negativo. Los números se comparan en múltiplos uno. Los datos de escala por razón tienen todas las propiedades de los datos de la escala de intervalo, por ejemplo, los datos deben tener valores numéricos, la distancia entre los dos puntos es igual, etc.; sin embargo, a diferencia de los datos de intervalo en la que el cero es arbitrario, en los datos de una escala por razón, el cero es absoluto.

Un excelente ejemplo de los datos de escala por razón es la medición de alturas. La altura puede medirse en centímetros, metros, pulgadas o pies. No es posible tener una altura negativa. Los datos de escala por razón pueden ser multiplicados y divididos, esta es una de las principales diferencias entre los datos de escala por razón y los datos de escala de intervalo, los cuales solo pueden ser sumados y restados¹.

1 Extraído de <https://glosarios.servidor-alicante.com/terminos-estadistica/escala-de-razon>.

1.10. El agrupamiento de datos

Por último, podemos mencionar que una vez que tenemos los datos, estos pueden agruparse según las necesidades de la investigación. Si, por ejemplo, tenemos un listado de profesiones en un determinado período histórico: granjeros, abogados, ingenieros, almaceneros, libreros, militares, cada uno con sus respectivas cantidades, podemos agrupar dichas profesiones en una cantidad menor de clases: militares, profesionales, comerciantes y trabajadores rurales.

De esta manera, estamos agrupando los datos y si bien perdemos precisión, resulta más fácil manipular la información. Este es el tema que veremos en el siguiente capítulo.

1.11. Práctica²



Ejercicio 1

Un organismo de la APN recopila información proveniente de las fichas personales de sus empleados. Los datos que se recolectan son los siguientes:

- a) Sexo.
- b) Edad.
- c) Cantidad de integrantes del grupo familiar.
- d) Estado civil.
- e) Nivel de educación alcanzado.

² Para el resultado de las prácticas de todos los capítulos, contactar al autor por correo electrónico a zappinoj@jefatura.gob.ar.

- f) Tipo de estudios secundarios cursados.
- g) Cantidad de años de estudio (todos los niveles).
- h) Distancia en kilómetros del domicilio al lugar de trabajo.
- i) Cantidad de medios de transporte utilizados para llegar al lugar de trabajo.

Analizar a qué tipo de variable corresponde cada uno de los datos recopilados.



Ejercicio 2

Explique las diferencias entre variable cualitativa y cuantitativa, y brinde tres ejemplos de cada una.



Ejercicio 3

Explique las diferencias entre variables discretas y continuas, y brinde tres ejemplos de cada una.



Ejercicio 4

Determine si cada una de las siguientes variables cuantitativas es discreta o continua.

- a) Monto de sueldo de un empleado de un organismo de la APN.
- b) Vida útil de un foco eléctrico de 100 watts.
- c) Cantidad de empleados de una universidad nacional.
- d) Número de trámites realizados por una oficina de ANSES en un mes.



Ejercicio 5

Describa las diferencias entre las distintas escalas de medición y brinde dos ejemplos de cada una.

2

Organización de los datos

2.1. Sistematización de los datos

Existen tres formas de sistematizar los datos: serie simple, distribución de frecuencias e intervalos de clase.

2.1.1. Serie simple

Los datos no agrupados pueden ser una serie de valores recolectados por cualquiera de los métodos de recolección, en la cual figuran todos y cada uno de ellos.

Por ejemplo, tenemos las calificaciones finales de los 30 participantes de un curso de INAP:

Cuadro 1

5	5	3
8	7	4
7	7	2
9	5	1
9	6	3
8	1	1
10	4	4
6	3	5
8	2	4
7	3	5

Si miramos los datos, podemos ver fácilmente cuál/les son la/s nota/s más alta/s, la/s más baja/s, la/s que se repiten. Sin embargo, conviene ordenar los datos:

Cuadro 1A

1	5	8
2	5	8
3	6	8
4	6	8
4	7	8
4	7	9
4	7	9
5	7	9
5	7	9
5	8	10

El contenido del cuadro 1A representa una serie simple.

Los datos así presentados tienen una ventaja y un inconveniente: si estos son trabajados en una planilla de cálculo, es muy fácil obtener los estadísticos más importantes (medidas de tendencia central, de dispersión, de asimetría, etc.); sin embargo, si el análisis se hace manualmente, el proceso se torna engorroso, por lo cual se sugiere trabajar con el concepto de frecuencia, es decir, agrupar los datos en función de la cantidad de veces que se repiten.

2.1.2. Distribución de frecuencias

El cuadro de frecuencias tiene como finalidad presentar de manera ordenada los valores que toman las diferentes características, de forma que permitan al lector tener una visión de conjunto. Los datos se clasifican y ordenan de acuerdo con ciertas características cualitativas o

cuantitativas, indicándose el número de veces que se repite el valor de la variable o atributo.

En el caso de atributos o aspectos cualitativos, se indica el número de veces que el valor del atributo se repite y es lo que se denomina «frecuencia de ocurrencia». En términos de la proporción que esta frecuencia de ocurrencia representa en el número total observado, se la conoce con el nombre de «frecuencia relativa de ocurrencia» (Biondolillo, 2017).

Es importante destacar, en el caso de atributos, que la característica puede ser analizada, en parte, mediante el cálculo de razones o porcentajes y, al igual que las variables cuantitativas, se puede representar gráficamente (Biondolillo, 2017).

Siguiendo con el ejemplo del curso de INAP, agruparemos las notas de los participantes en frecuencias.

Cuadro 2

Nota final				
x_i	f_i	f_r	F_a	F_{ra}
1	1	0,03	1	0,03
2	1	0,03	2	0,07
3	1	0,03	3	0,10
4	4	0,13	7	0,23
5	5	0,17	12	0,40
6	2	0,07	14	0,47
7	5	0,17	19	0,63
8	6	0,20	25	0,83
9	4	0,13	29	0,97
10	1	0,03	30	1,00
Total	30	1,00		

donde:

x_i : valores que toma la variable.

f_i : frecuencia absoluta. Indica las veces que se repite.

f_r : frecuencia relativa. Indica la distribución porcentual. Resulta de dividir la frecuencia absoluta de cada intervalo por el total de casos (total de f_i). El total de f_r debe sumar 1.

F_a : frecuencia absoluta acumulada. Surge de sumar la frecuencia absoluta de cada caso con la frecuencia absoluta del caso siguiente. El resultado del último caso debe coincidir con el total de los casos.

F_r : frecuencia relativa acumulada. Resulta de dividir la frecuencia absoluta acumulada de cada caso por la frecuencia acumulada del último caso (que coincide con el total de casos).

2.1.3. Intervalos de clase

Para estudiar un hecho en el que la amplitud de la población o muestra es grande, se definen los intervalos de clase, que son subconjuntos del conjunto de valores que puede tomar la variable. Estas clases pueden tener una amplitud constante o variable.

Siguiendo con el ejemplo del curso de INAP, podríamos dividir las notas en intervalos de clase:

Cuadro 3

Intervalo	Marca de clase				
	x_i	f_i	f_r	F_a	F_r
1 - 2	1,5	2	0,07	2	0,07
3 - 4	3,5	5	0,17	7	0,23
5 - 6	5,5	7	0,23	14	0,47
7 - 8	7,5	11	0,37	25	0,83
9 - 10	9,5	5	0,17	30	1,00
Total		30	1,00		

donde:

xi: marca de clase (valor medio del intervalo). Se calcula sumando el límite inferior y el límite superior y, luego, dividirlo por dos.

fi: frecuencia absoluta (veces que se repite).

fr: frecuencia relativa. Indica la distribución porcentual. Resulta de dividir la frecuencia absoluta de cada intervalo por el total de casos (total de fi). El total de fr debe sumar 1.

Fa: frecuencia absoluta acumulada. Surge de sumar la frecuencia absoluta de cada caso con la frecuencia absoluta del caso siguiente. El resultado del último caso debe coincidir con el total de los casos.

Fra: frecuencia relativa acumulada. Resulta de dividir la frecuencia absoluta acumulada de cada caso por la frecuencia acumulada del último caso (que coincide con el total de casos).

Se puede observar que las clases tienen una amplitud constante de 2 puntos. Los extremos son 2, 4, 6, etc. y las marcas de clase o punto medio del intervalo (xi), 1,5, 3,5, etc., las cuales se calculan sumando el límite inferior y el límite superior del intervalo y, luego, dividirlo por dos.

Se llama «límites del intervalo» a los valores extremos de dicho intervalo. Por ejemplo, dado el intervalo 1-2, diremos que 1 es el límite inferior (Linf.) y que 2 es el límite o extremo superior (Lsup.).

Los criterios a seguir para elegir la amplitud y el número de intervalos de clase son los siguientes:

- en cuanto a la amplitud de cada clase, en general, se eligen clases con amplitud constante, por simplicidad. Para calcular la amplitud del intervalo (c), se realiza la siguiente operación:

$$c = \text{rango} / m$$

donde:

m = cantidad de clases

- en cuanto al número de clases a adoptar, dependerá de la precisión de las medidas que se pretende alcanzar, la finalidad del estudio, el grado de variabilidad de los datos, la necesidad de efectuar comparaciones, entre otros criterios. Es común la utilización de la fórmula de Sturges para obtener la cantidad de intervalos de clase (m):

$$m = 1 + 3,3 * \log(n)$$

Con respecto a la agrupación de los datos en clases o intervalos, hay que decir que lleva implícito un cierto error en el cálculo posterior de las distintas medidas de la distribución, y cuanto mayor sea la amplitud de los intervalos elegidos, este error será mayor. Pero si esta amplitud es muy pequeña y, por consiguiente, el número de intervalos es muy grande, entonces, la serie estadística pierde claridad y los cálculos resultan más engorrosos. Por lo tanto, se debe tender al máximo equilibrio entre estos dos efectos contrarios:

- mayor número de clases implica menos claridad, pero más exactitud y precisión;
- menos clases y más amplias significa ganar en claridad de exposición y facilidad de cálculo, en detrimento de exactitud y precisión a conseguir en los cálculos.

2.2. Práctica



Ejercicio 1

Los siguientes valores representan la cantidad de empleados públicos distribuidos en 50 distintas dependencias de la Municipalidad de Rosario, provincia de Santa Fe:

Cuadro 3A

29	58	80	35	30	23	88	49	35	97
12	73	54	91	45	28	61	61	45	84
83	23	71	63	47	87	36	8	94	26
95	63	86	42	22	44	88	27	20	33
28	91	87	15	67	10	45	67	26	19

Con esa información, construir un cuadro de intervalos con 10 clases de igual distancia entre el valor mínimo y el valor máximo de cada clase. Luego, agregarle las siguientes columnas: intervalo de clase, marca de clase, frecuencia absoluta, frecuencia relativa absoluta, frecuencia acumulada, frecuencia relativa acumulada.



Ejercicio 2

En un estudio relativo a una peste que afecta a las plantaciones de cítricos, el INTA subdivide el terreno de una explotación en San Pedro, provincia de Buenos Aires, en 42 parcelas de similar tamaño, y contabiliza la cantidad de naranjos apestados en cada una de ellas. Los resultados obtenidos se muestran en el siguiente cuadro:

Cuadro 3B

2	3	3	14	15	16	17	17	19	19	19	22	22	23
25	25	25	28	28	29	32	34	34	35	36	36	36	40
40	41	42	46	47	51	52	52	53	54	60	65	81	84

Con esa información, construir un cuadro de frecuencias con las siguientes columnas: valor de la variable, frecuencia absoluta, frecuencia relativa absoluta, frecuencia acumulada, frecuencia relativa acumulada.



Ejercicio 3

Dada la siguiente serie simple, correspondiente a la cantidad de trámites realizados por una oficina de ANSES durante dos meses (44 días hábiles), crear intervalos de clase iguales y las siguientes columnas: intervalo de clase, marca de clase, frecuencia absoluta, frecuencia relativa absoluta, frecuencia acumulada, frecuencia relativa acumulada.

Cuadro 3C

5	80	24	30
73	68	5	94
55	97	78	42
24	79	15	10
22	39	52	76
68	15	52	78
49	20	49	5
77	94	68	80
49	60	30	22
27	29	27	92
22	39	52	76



Ejercicio 4

Se quiere conocer si los 24 propietarios de los departamentos de un edificio de CABA están a favor o en contra del reemplazo de tubos de iluminación fluorescente por iluminación LED, para lo cual se han recogido los siguientes datos:

Cuadro 3D

2	5	5	5	3	4
5	5	4	2	4	4
3	4	3	4	4	3
5	2	4	1	4	5

Regla de codificación

- 1** En contra
- 2** Bastante en contra
- 3** Indiferente
- 4** Bastante a favor
- 5** A favor

Construir un cuadro de frecuencias con las siguientes columnas: valor de la variable, frecuencia absoluta, frecuencia relativa absoluta, frecuencia acumulada, frecuencia relativa acumulada.



Ejercicio 5

Los siguientes datos se refieren al puntaje obtenido por 40 participantes de un curso de INAP:

Cuadro 3E

6,75	6,50	6,00	6,75
7,00	6,25	6,75	7,10
7,00	6,25	6,75	7,25
6,75	6,50	7,10	6,75
6,50	6,65	7,00	6,25
6,50	7,00	6,70	6,25
7,15	7,25	6,50	7,00
7,00	6,70	6,75	6,75
6,50	6,00	6,25	7,00
6,50	6,75	6,65	7,15

Crear un cuadro de intervalos de clase iguales con las siguientes columnas: intervalo de clase, marca de clase, frecuencia absoluta, frecuencia relativa absoluta, frecuencia acumulada, frecuencia relativa acumulada.

3

Elaboración de gráficos

3.1. Formas de presentación

La presentación de las informaciones obtenidas se puede realizar de varias maneras:

- textual (en forma de texto).
- cuadros (como analizamos en el capítulo 2).
- gráficos.

En un informe textual complementado con cuadros, estos sirven de ayuda al lector para entender o ampliar lo que dice el texto. Sin embargo, la utilización de texto y cuadros no es suficiente para la presentación de un buen informe, ya que la lectura del texto implicará un tiempo precioso para comprender el contenido. Ahora bien, si el lector se remite a los cuadros, encontrará cifras que, en algunos casos, no le aclaran lo suficiente y, en otros, no le dicen mucho sobre el problema estudiado, de ahí que sea necesario que el informe vaya acompañado de gráficos, no como sustituto de las formas anteriores de presentación, sino como la mejor manera de visualizar la información contenida en los cuadros y en el texto.

De todas las representaciones de datos estadísticos, los gráficos muestran en forma extensiva y con la necesaria profundidad las características del fenómeno en estudio.

El gráfico es una representación de datos ilustrada, construida principalmente mediante la asignación de valores a un conjunto de líneas horizontales y verticales llamadas «ejes».

Generalmente, los gráficos usan como sostén un sistema de ejes cartesianos X e Y, donde se presentan dos variables asociadas. En el eje «Y», se colocan las frecuencias que presentan los distintos valores de la variable X, que se representa sobre el eje de abscisas o «X».

3.2. Reglas en la elaboración de gráficos

La forma del gráfico depende, en gran medida, del gusto de la persona que lo elabora; sin embargo, hay ciertos principios generales que se deben tener en cuenta en el logro de un buen gráfico:

- El mejor gráfico es el más *simple*. Las líneas y símbolos usados deben ser los estrictamente indispensables para una mejor visualización de la información.
- Si hay más de dos gráficos, deben estar *numerados*.
- Todo gráfico debe llevar un *título* que aclare el contenido. Debe responder a los interrogantes: qué, cómo, cuándo y dónde.
- El *título* debe ser colocado en la parte superior.
- Las *líneas* que llevan escalas deben dibujarse más gruesas que las demás coordenadas; a su vez, las líneas que representan los datos o la variable deben ser más gruesas que las escalas.

- La línea vertical, denominada *ordenada*, se utiliza para representar la frecuencia.
- Las características cualitativas y cuantitativas van en la línea horizontal o *abscisa*.
- La *lectura* de la escala del eje horizontal se hace de izquierda a derecha y la lectura del eje vertical se realiza desde abajo hacia arriba.
- En todo gráfico se debe explicar la *fuentes* de donde fueron obtenidos los datos; además, aclarar las escalas, leyendas, notas y convenciones que ayuden a identificar las características presentadas.
- Los gráficos deben ser lo bastante *sencillos* para procurar una idea clara y ser comprensibles sin la ayuda de las descripciones del texto.
- Los datos numéricos sobre los que se basa el gráfico deben presentarse en *cuadro adjunto*, si no son incluidos en el mismo gráfico.
- Los gráficos deben seguir y nunca preceder la *exposición del texto*.
- Cuando los valores representados en un gráfico son elevados y varían poco, se pierde un gran espacio por debajo de la curva, siendo preciso adoptar una escala bastante pequeña que desdibuja las variaciones. Por ello, es aconsejable interrumpir el cuadrículado mediante un *corte* horizontal o vertical según el eje en cuestión.
- Cuando el gráfico presenta más de una variable, debe hacerse una muy clara *diferenciación* por medio de leyendas, notas o signos convencionales.

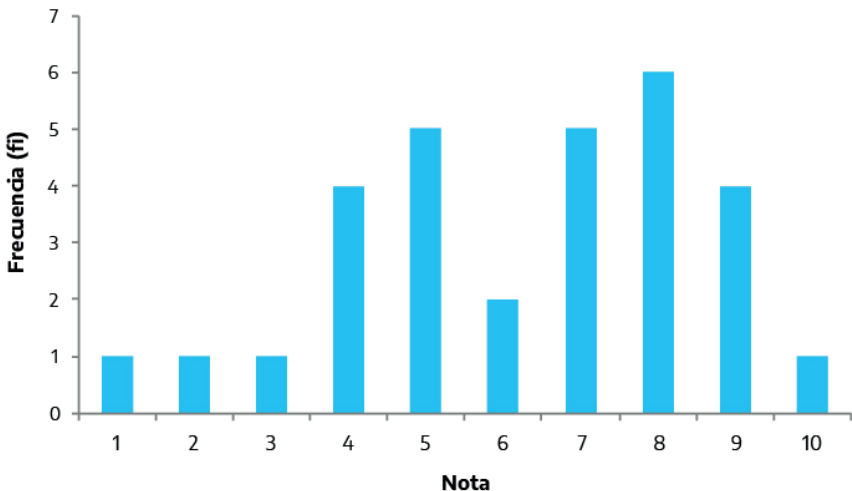
3.3. Tipos de gráficos

Hay numerosos tipos de gráficos, pero veremos detenidamente los más usuales.

3.3.1. Gráfico de frecuencias

Se utiliza para la representación de la variable discreta. En el eje horizontal, se colocan los valores que toma la variable x_i y, en el eje vertical, las frecuencias absolutas o relativas. También se puede hacer la representación de la variable utilizando las frecuencias absolutas o relativas acumuladas; por tal razón, se denominan «diagramas de frecuencias acumuladas». En este caso, se observará que las líneas que representan la variable van en sentido horizontal, a diferencia del caso anterior cuyas barras son verticales. Los gráficos elaborados con frecuencias absolutas son exactamente iguales a aquellos elaborados con frecuencias relativas, y se diferencian únicamente en los valores de las escalas correspondientes a las frecuencias.

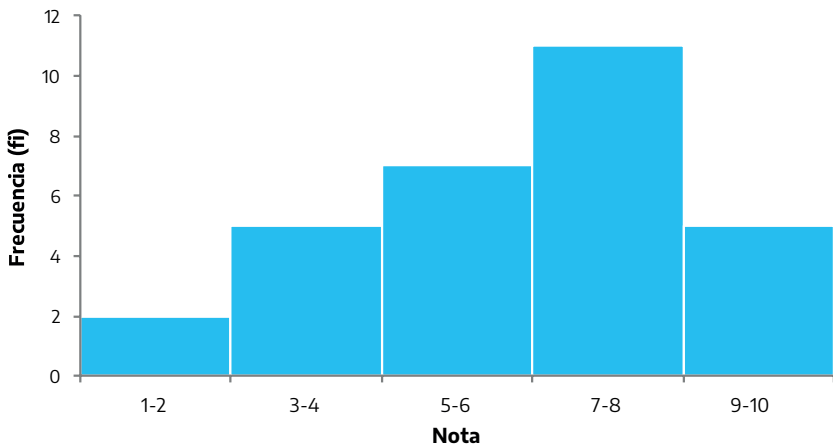
Gráfico 1. Notas de participantes curso INAP



3.3.2. Histograma

Está formado por un conjunto de rectángulos, cada uno de ellos levantado en cada intervalo, de manera que la base será igual a la amplitud del intervalo (c) y la altura estará dada por la frecuencia absoluta o relativa. Se utiliza para la representación de la variable continua. Si elaboramos el histograma teniendo en cuenta las frecuencias relativas, será un gráfico de gran utilidad al comparar dos distribuciones.

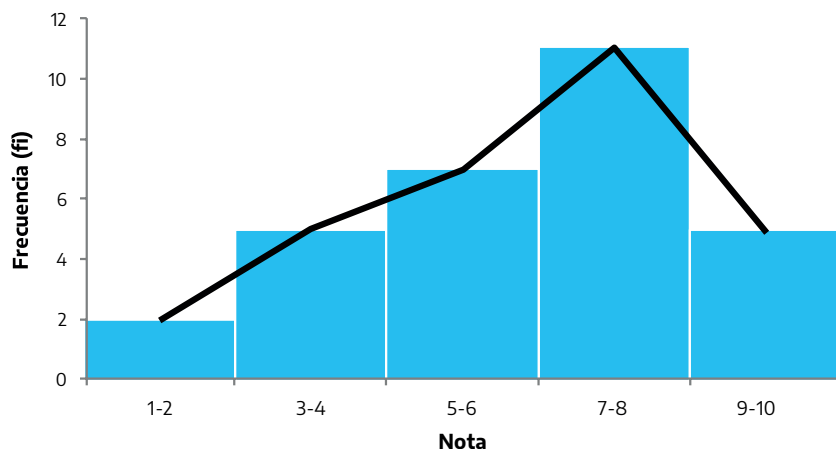
Gráfico 2. Notas de participantes curso INAP



3.3.3. Polígonos de frecuencias

Con la misma información utilizada para la elaboración del histograma, se puede dibujar el polígono de frecuencia. Primero, se establecen los puntos medios en la parte superior de cada rectángulo y, luego, se unen dichos puntos mediante una línea que se prolonga desde el primer y último intervalo hasta el eje de abscisas y hasta el punto medio del intervalo anterior y posterior a los de la variable en estudio.

Gráfico 3. Notas de participantes curso INAP

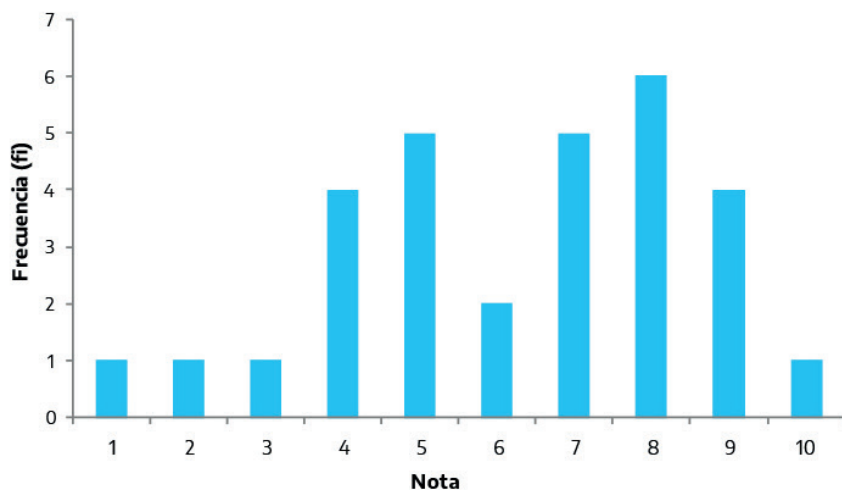


3.3.4. Gráfico de barras

Es uno de los gráficos más utilizados por su sencillez y por la facilidad que ofrece para representar las características cuantitativas y cualitativas. Pueden ser verticales u horizontales.

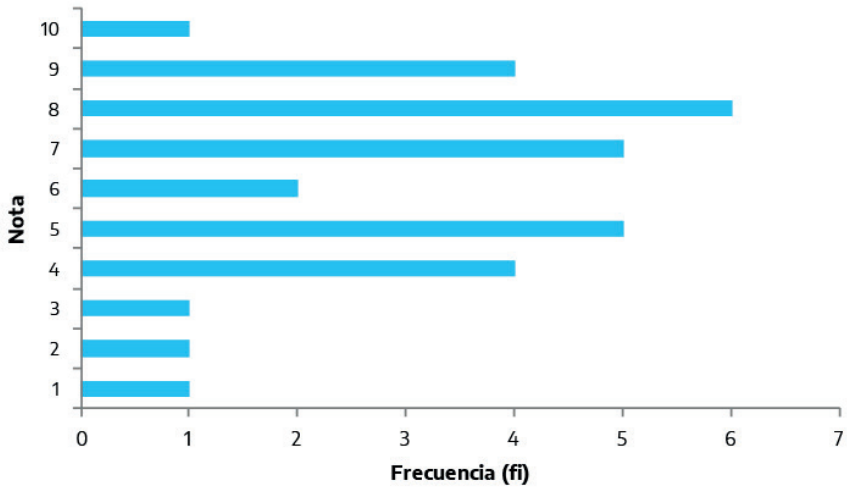
3.3.4.1. Verticales

Gráfico 4. Notas de participantes curso INAP



3.3.4.2. Horizontales

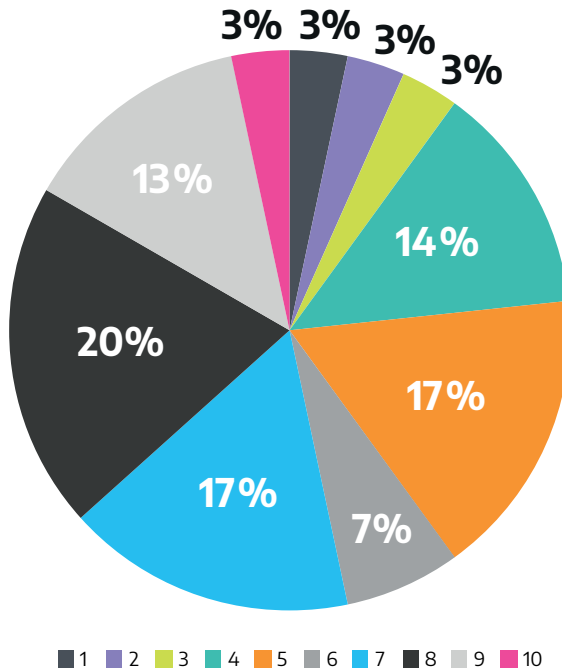
Gráfico 5. Notas de participantes curso INAP



3.3.5. Gráfico circular

Se utiliza con mucha frecuencia para representar características cualitativas y sirve para hacer notar las diferencias en las proporciones o porcentajes en que está dada la distribución. Este tipo de comparación es relativamente efectivo siempre que los segmentos sean lo suficientemente grandes para permitir comparaciones. El proceso que se sigue en la confección del gráfico circular o de pastel consiste en subdividir los 360 grados de la circunferencia proporcionalmente al porcentaje de cada una de las clases que ha tomado la característica.

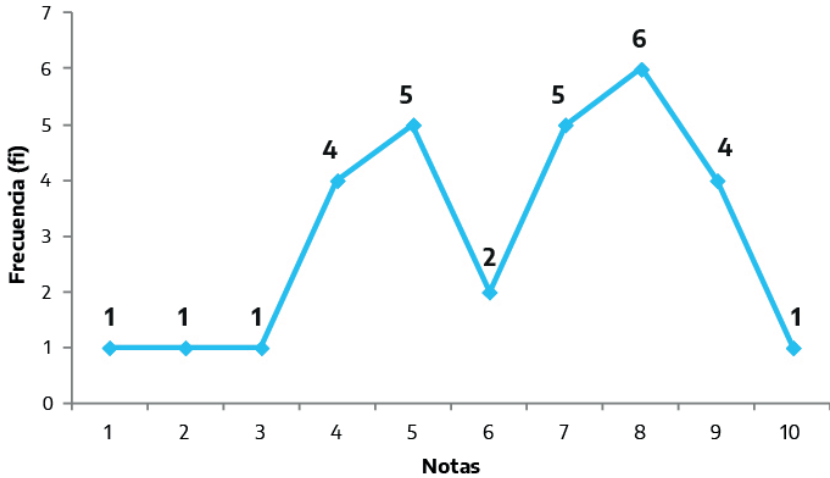
Gráfico 6. Notas de participantes curso INAP



3.3.6. Diagramas de líneas

Aunque es otro de los gráficos muy utilizados, es el que presenta mayores dificultades en la visualización de los datos, lo que da lugar, algunas veces, a conclusiones erróneas debido a la mala confección de las escalas de los ejes.

Gráfico 7. Notas de participantes curso INAP



3.4. Práctica



Ejercicio 1

Con base en el cuadro creado en la práctica 3 de la unidad 2, realizar:

- a) Gráfico de frecuencias simples.
- b) Histograma de frecuencias simples.



Ejercicio 2

Con base en el cuadro creado en la práctica 4 de la unidad 2, realizar:

- a) Polígono de frecuencias simples.
- b) Gráfico de barras verticales de la frecuencia acumulada.



Ejercicio 3

Con base en el cuadro creado en la práctica 5 de la unidad 2, realizar:

- Gráfico de barras verticales de la frecuencia simple.
- Gráfico circular de la frecuencia simple.
- Gráfico de líneas de la frecuencia simple.



Ejercicio 4

Con base en el siguiente cuadro, que indica la frecuencia de acceso a medios de transporte para 114 pasajeros, realizar:

Cuadro 7A

Medio de transporte	f_i	f_{ri}	F_a	F_{ra}
Subterráneo	53	46,49	53	46,49
Colectivo	29	25,44	82	71,93
Tren	13	11,40	95	83,33
Automóvil	11	9,65	106	92,98
Motovehículo	3	2,63	109	95,61
Otros	5	4,39	114	100,00
Total	114	100,00		

- Gráfico de frecuencias acumuladas.
- Gráfico de líneas de la frecuencia acumulada.



Ejercicio 5

Con base en el siguiente cuadro, que muestra la cantidad de días pernoctados en alojamientos de tipo rural por 254 turistas, realizar:

Cuadro 7B

Intervalo	Marca de clase				
	x_i	f_i	f_{ri}	F_a	F_{ra}
1 - 3	2	378	72,14	378	72,14
3 - 7	5	102	19,47	480	91,60
7 -15	11	25	4,77	505	96,37
15 - 50	32,5	19	3,63	524	100,00
Total	50,5	524	100,00		

- Histograma de frecuencias simples.
- Gráfico de barras verticales de la frecuencia simple.

Estadística descriptiva

4.1. Definición y clasificación de las medidas

En este capítulo explicaremos que las medidas son los valores que identifican o caracterizan a ciertas particularidades de las distribuciones de frecuencias. Estos valores son usualmente funciones algebraicas de las cifras que toma la variable analizada al ser medida u observada, o cumplen ciertas condiciones estipuladas.

Estas medidas sirven para poder resumir la información de un conjunto grande de datos y, de esta forma, poder tomar decisiones sin necesidad de conocer todos y cada uno de ellos.

A grandes rasgos, a las medidas que caracterizan una distribución de frecuencias se las puede clasificar de la siguiente forma:

- Medidas de posición o de tendencia central: son aquellas que sugieren o indican el lugar de ubicación o posición en que se encuentra una distribución de frecuencias dentro del campo de definición de la variable. En este caso, tenemos la media aritmética, la mediana y la moda.
- Medidas de dispersión o de variabilidad: son aquellas que determinan el grado de dispersión (o concentración) de una

distribución de frecuencias. En este caso, tenemos el rango, la varianza y el desvío estándar.

- Medidas de tendencia no central: son aquellas que permiten conocer ciertos puntos de una serie de datos que no siempre son valores centrales como la mediana. Se podría dividir la serie en cuatro, diez o cien grupos que contengan la misma cantidad de elementos. En este caso, tenemos los percentiles, los deciles, los cuartiles y el rango intercuartílico.

4.2. Medidas de posición o de tendencia central

4.2.1. Media aritmética (\bar{X})

Es el centro de gravedad de la distribución. Es un tipo de promedio que depende de todos los valores de la serie. Cuando se habla del promedio, generalmente, se hace referencia a la media aritmética.

La media aritmética es la medida de tendencia central que más se utiliza y viene definida de acuerdo a si los datos están o no agrupados.

a) Cálculo de la media aritmética para datos no agrupados

$$\bar{X} = \frac{\sum_{i=1}^n x_i}{n}$$

donde:

$\sum_{i=1}^n x_i$: suma de todos los valores de la variable.

n: número total de valores.

Ejemplo

Volvamos al caso del curso de INAP que usamos en el capítulo 2:

Cuadro 4

5	5	3
8	7	4
7	7	2
9	5	1
9	6	3
8	1	1
10	4	4
6	3	5
8	2	4
7	3	5

el cálculo de la media aritmética es como sigue:

$$\bar{X} = \frac{5+8+7+9+9+8+10+6+8+7+5+7+7+5+6+1+4+3+2+3+3+4+2+1+3+1+4+5+4+5}{30} = \frac{152}{30} = 5,07$$

Entonces, la media aritmética o el promedio de las notas del curso es 5,07.

b) Cálculo de la media aritmética para datos agrupados en frecuencias simples

$$\bar{X} = \frac{\sum_{i=1}^n (x_i * f_i)}{n}$$

donde:

$\sum_{i=1}^n (x_i * f_i)$: suma de todos los valores de la variable multiplicados por su frecuencia simple.

n: suma de frecuencias simples.

Ejemplo

Si tomamos el mismo curso de INAP, pero agrupamos las notas por cantidad de veces que se repiten, tenemos:

Cuadro 5

Nota final		
x_i	f_i	$x_i * f_i$
1	3	3
2	2	4
3	4	12
4	4	16
5	5	25
6	2	12
7	4	28
8	3	24
9	2	18
10	1	10
Total	30	152

En este caso, el cálculo de la media aritmética es como sigue:

$$\bar{X} = \frac{\sum_{i=1}^n x_i * f_i}{30} = \frac{152}{30} = 5,07$$

De esta manera, llegamos al mismo resultado que antes, que es 5,07.

c) Cálculo de la media aritmética para datos agrupados en intervalos de clase

$$\bar{X} = \frac{\sum_{i=1}^n (x_i * f_i)}{n}$$

donde:

$\sum_{i=1}^n (x_i * f_i)$: suma de todos los valores de la variable multiplicados por su frecuencia simple.

x_i : marca de clase o valor medio del intervalo.

n : suma de frecuencias simples.

Ejemplo

Siguiendo con el ejemplo del curso de INAP:

Cuadro 6

Intervalo	Marca de clase		
	x_i	f_i	$x_i * f_i$
1 - 2	1,5	5	7,50
3 - 4	3,5	8	28,00
5 - 6	5,5	7	38,50
7 - 8	7,5	7	52,50
9 - 10	9,5	3	28,50
Total		30	155,00

el cálculo de la media aritmética es como sigue:

$$\bar{X} = \frac{155}{30} = 5,17$$

Es decir, el promedio de notas es 5,17.

Nota 1: la diferencia con los dos casos anteriores, aunque pequeña, se debe a que no se toma el valor original de la variable, sino el valor medio del intervalo.

Nota 2: si la media aritmética no proviene de un conjunto de datos homogéneos, su importancia no será significativa. Además, si el límite superior o inferior de la variable no están limitados, tampoco será posible el cálculo.

4.2.2. Mediana (Me)

La mediana es un valor que separa la serie ordenada en dos grupos que contienen la misma cantidad de datos, de manera que en el primer grupo se encuentran todos los valores menores o iguales al 50 % de los datos y en el segundo grupo, el resto de los valores.

No es propiamente un promedio, ya que no depende de los valores particulares de la serie de datos observada, sino de su ordenación y de su concentración. Tampoco está influida por un valor extremo de la variable, como ocurre con la media aritmética.

Para su cálculo, se debe tener en cuenta si los datos no están agrupados o se trata de una distribución de frecuencias simples y si están agrupados en intervalos de clase. Para el caso de datos no agrupados y frecuencias simples, primero, se procede a su ordenamiento de menor a mayor y, luego, se analiza si la cantidad de datos es par o impar. Si es impar, la mediana coincide con el valor del medio; si es par, se toman los dos valores centrales y se calcula el promedio aritmético. El valor obtenido es la mediana.

La mediana es la medida de centralización que se utiliza cuando la variable se mide con una escala numérica u ordinal. Es la medida más apropiada en el caso de que la distribución de frecuencia de la variable sea asimétrica, ya que es menos sensible que la media aritmética a valores extremos.

a) Cálculo de la mediana para datos no agrupados

Se ordenan los datos de manera creciente y, luego, se calcula el orden de la mediana. Con número impar de observaciones, el valor de la mediana está dado por el valor de la variable cuya ubicación en el arreglo de datos coincide con el orden de la mediana. Con número par de observaciones,

el valor de la mediana está dado por la media aritmética de los dos valores centrales del arreglo de datos. La fórmula es la siguiente:

$$Me = \frac{n+1}{2}$$

Ejemplo

Si tenemos una serie con cantidad par de números:

5	9	5	7	7	1	2	4	8	5	8	9	6	5	5	4	8	7	9	4	7	8	8	7	6	3	8	9	4	5
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

se ordenan en forma creciente

1	2	3	4	4	4	4	5	5	5	5	6	6	7	7	7	7	7	8	8	8	8	8	8	8	9	9	9	9	10
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	----

resultando $Me = (7 + 7) / 2 = 7$

Si tenemos una serie con cantidad impar de números:

5	9	5	7	7	1	2	4	8	5	8	9	6	5	5	4	8	7	9	4	7	8	8	7	6	3	8	9	4
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

se ordenan en forma creciente

1	2	3	4	4	4	4	5	5	5	5	6	6	7	7	7	7	7	8	8	8	8	8	8	8	9	9	9	9
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

resultando $Me = 7$

b) Cálculo de la mediana para datos agrupados en frecuencias simples

Se calcula el orden de la mediana como $n/2$ y, con ese valor, ubicamos la clase mediana según la columna de F_i (frecuencia absoluta acumulada) teniendo en cuenta que el valor de F_i debe ser igual o superior al valor dado por el orden de la mediana. El valor de la mediana es el valor de la variable que corresponde a la F_i indicada.

Ejemplo

Si tomamos nuevamente las notas de los participantes del curso de INAP:

Cuadro 7

Nota final		
x_i	f_i	F_a
1	3	3
2	2	5
3	4	9
4	4	13
5	5	18
6	2	20
7	4	24
8	3	27
9	2	29
10	1	30
Total	30	

$$n/2 = 15$$

El 15 se encuentra comprendido en el valor $x_i=6$, ya que $F_a=20$.

La mediana, entonces, es 6 (línea sombreada).

c) Cálculo de la mediana para datos agrupados en intervalos de clase

Se realiza sobre el intervalo en el que se acumula el 50 % de las frecuencias, y la fórmula es la siguiente:

$$Me = L_{inf} + \frac{\frac{n}{2} - Fi_{-1}}{f_i} * W_i$$

donde:

Fi_{-1} : frecuencia absoluta acumulada de la clase anterior a la clase mediana.

L_{mf} : límite inferior de la clase mediana.

$n/2$: orden de la mediana.

fi : frecuencia absoluta de la clase mediana.

W_i : amplitud de la clase mediana.

Ejemplo

Si tenemos el siguiente cuadro, donde se muestra la cantidad de agentes pertenecientes a la APN según rango etario (datos al 1/9/2019):

Cuadro 8

Rango etario (años)	Marca de clase		Cantidad de agentes		
	x_i	f_i	f_{ri}	F_a	F_{ra}
20-24	22	40.481	0,11	40.481	0,11
25-29	27	54.269	0,14	94.750	0,25
30-34	32	54.484	0,14	149.234	0,40
35-39	37	56.358	0,15	205.592	0,55
40-44	42	52.097	0,14	257.689	0,68
45-49	47	41.718	0,11	299.407	0,79
50-54	52	34.239	0,09	333.646	0,89
55-59	57	26.139	0,07	359.785	0,96
60-64	62	16.842	0,04	376.627	1,00
Total		376.627	1,00		

Fuente: elaboración propia con base en BIEP, <https://www.argentina.gob.ar/modernizacion/empleopublico/biep/informes-estadisticos>

$n/2 = 188.313,5$. Este valor se encuentra comprendido en el intervalo 35-39. Es decir, este es el intervalo de la mediana.

Por consiguiente, el cálculo se realiza de la siguiente manera:

$$Me = 35 + \frac{\frac{376.627}{2} - 149.234}{56.358} * 5 = 38,47$$

Entonces, la mediana es igual a 38,47 años.

4.2.3. Moda (M_0)

Es el valor de variable que más observaciones presenta. En general, cuando la serie de datos está conformada por pocos valores, la moda no es relevante. Este parámetro solo se aplica a datos agrupados.

Nota: en el caso de que existiera más de un valor entre los que más se repiten, tendríamos una serie con más de una moda. Si hubiera dos, la serie sería bimodal; si hubiera 3, trimodal, etc.

a) Cálculo de la moda para datos no agrupados

En este caso es simplemente el valor de la variable que más veces se repite.

Cuadro 9

1	5	5
7	5	6
8	9	4
8	9	4
8	9	8
2	10	8
3	6	7
4	7	7
4	7	8
5	5	9

La nota que más se repite es 8 (seis veces).

b) Cálculo de la moda para datos agrupados en frecuencias simples

Cuadro 9A

Nota final	
xi	fi
1	1
2	1
3	1
4	4
5	5
6	2
7	5
8	6
9	4
10	1
Total	30

El valor con mayor frecuencia absoluta (fi) es 8, que representa el valor modal.

c) Cálculo de la moda para datos agrupados en intervalos de clase

La forma de cálculo viene dada por la fórmula:

$$Mo = L_{\text{Inf}} + \frac{d_1}{d_1 + d_2} * w_1$$

donde:

L_{Inf} : límite inferior del intervalo donde se encuentra el modo —surge de la simple observación de la serie de datos—.

d_1 : diferencia entre la frecuencia del intervalo donde se encuentra el modo y la frecuencia del intervalo anterior.

d_2 : diferencia entre la frecuencia del intervalo donde se encuentra el modo y la frecuencia del intervalo posterior.

w : amplitud del intervalo, que debe ser constante.

Ejemplo

Volvamos al caso de los agentes de la APN:

Cuadro 10

Rango etario (años)	Marca de clase x_i	Cantidad de agentes f_i
20-24	22	40.481
25-29	27	54.269
30-34	32	54.484
35-39	37	56.358
40-44	42	52.097
45-49	47	41.718
50-54	52	34.239
55-59	57	26.139
60-64	62	16.842
Total		376.627

Fuente: elaboración propia con base en BIEP, <https://www.argentina.gob.ar/modernizacion/empleopublico/biep/infomres-estadisticos>

El intervalo modal es 35-39, ya que es el que más frecuencia absoluta (f_i) presenta. Entonces, la fórmula es la siguiente:

$$Mo = 35 + \frac{56.358 - 54.484}{(56.358 - 54.484) + (56.358 - 52.097)} * 5 = 36,53$$

Es decir, la moda es 36,53 años.

4.3. Medidas de dispersión o de variabilidad

Estas medidas nos proporcionan información adicional a los efectos de juzgar sobre la confiabilidad de nuestra medida de tendencia central, en especial, de la media aritmética. Para los casos en que los datos se encuentren altamente dispersos, estas medidas nos informarán que la media aritmética no es representativa; para los casos en que los datos estén fuertemente concentrados, estas medidas nos informarán que la media es altamente representativa de todos ellos. Tengamos presente que la variabilidad es de capital importancia en el control de la producción de petróleo por la empresa estatal YPF, en el tiempo de cancelación de la deuda por parte del país, en el déficit y el superávit de las cuentas públicas, etc.

El concepto de las medidas de dispersión (MD) sostiene que, a partir del valor cero (cuando no existen diferencias entre los valores de la variable, es decir que no hay dispersión), toman mayores resultados (positivos) cuanto más dispersos se encuentran los valores de la variable. La unidad de medida resultante depende de la forma en que se las define, aunque la mayoría queda expresada en la misma unidad de medida que la variable.

Si se identifica genéricamente a las medidas de dispersión con MD, se deberá cumplir que:

$$0 < MD < xMáx$$

donde:

si $MD = 0$, significa que la distribución de frecuencias no tiene dispersión, o sea que todos los valores de la variable de todas las unidades en estudio son iguales entre sí.

si $MD = x_{Máx}$, significa la máxima dispersión que puede darse en la distribución de frecuencias.

4.3.1. Rango (R)

Es la distancia entre el mayor y menor de los valores observados. Nos permite analizar la extensión de las variaciones observadas. Es muy sensible a la presentación de valores atípicos. Tiene como desventaja que su cálculo se basa en solo dos valores que, por ser los extremos, pueden ser muy atípicos. No brinda información acerca de cómo se dispersan los datos dentro del intervalo limitado por el menor y el mayor valor.

Ejemplo

Siguiendo con el ejemplo de la cantidad de agentes de la APN según rango etario (ver cuadro 10), la menor edad observada es 20 años y la mayor 64 años, con lo cual el rango es $R = 64 - 20 = 44$ años.

4.3.2. Varianza (s^2)

Es una medida de variabilidad basada en las desviaciones respecto de la media aritmética de la variable. Nos da una distancia promedio—en unidades cuadráticas— entre cualquier observación del conjunto de datos y la media.

a) Cálculo para datos no agrupados

Para el caso de datos no agrupados, donde «n» es la cantidad de observaciones.

$$s^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}$$

Ejemplo

Siguiendo con el caso del curso de INAP, tenemos la serie de notas cuya media aritmética, recordamos, es 6,27.

Procedemos al cálculo de la varianza:

Cuadro 11

Nota final		
x_i	Media aritmética	$(x_i - \text{Media aritmética})^2$
5		1,61
9		7,45
10		13,91
7		0,53
7		0,53
1		27,77
2		18,23
4		5,15
8		2,99
5		1,61
8		2,99
9		7,45
6		0,07
5		1,61
5		1,61
4	6,27	5,15
8		2,99
7		0,53
9		7,45
4		5,15
7		0,53
8		2,99
8		2,99
7		0,53
6		0,07
3		10,69
8		2,99
9		7,45
4		5,15
5		1,61
Total		149,87

Según el cuadro, $\sum_{i=1}^n (X_i - \bar{X})^2 = 149,87$

entonces: $s^2 = \frac{149,87}{30} = 5^2$

Es decir, la varianza es 5^2 .

b) Cálculo para datos agrupados en frecuencias simples

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{X})^2 * f_i}{N}$$

Ejemplo

Si tomamos el mismo curso de INAP, pero agrupamos las notas por cantidad de veces que se repiten, tenemos:

Cuadro 12

Nota final		Media aritmética	$((x_i - \text{Media aritmética})^2) * f_i$
xi	fi		
1	1	6,27	27,77
2	1		18,23
3	1		10,69
4	4		20,61
5	5		8,06
6	2		0,15
7	5		2,66
8	6		17,96
9	4		29,81
10	1		13,91
Total	30		149,87

Según el cuadro, $\sum_{i=1}^n \left(X_i - \bar{X} \right)^2 * f_i = 149,87$

entonces: $s^2 = \frac{149,87}{30} = 5^2$

Es decir, la varianza es 5^2 , que coincide con el resultado para los datos no agrupados.

c) Cálculo para datos agrupados en intervalos de clase

Para el caso de datos agrupados en intervalos de clase, donde «n» es la cantidad de intervalos y «N» es la cantidad de observaciones, tenemos:

$$s^2 = \frac{\sum_{i=1}^n \left(x_i - \bar{X} \right)^2 * f_i}{N}$$

Ejemplo

En el siguiente cuadro, se muestra la cantidad de agentes pertenecientes a la APN según rango etario (datos al 1/9/2019):

Cuadro 13

Rango etario (años)	Marca de clase x_i	Cantidad de agentes f_i	$x_i * f_i$	Media aritmética	$((x_i - \text{Media aritmética})^2) * f_i$
20-24	22	40.481	890.582	38,89	11.551.476,64
25-29	27	54.269	1.465.263		7.675.309,38
30-34	32	54.484	1.743.488		2.588.324,08
35-39	37	56.358	2.085.246		201.842,78
40-44	42	52.097	2.188.074		503.087,58
45-49	47	41.718	1.960.746		2.742.209,89
50-54	52	34.239	1.780.428		5.882.512,32
55-59	57	26.139	1.489.923		8.570.525,15
60-64	62	16.842	1.044.204		8.992.920,29
Total		376.627	14.647.954		

Fuente: elaboración propia con base en BIEP, <https://www.argentina.gov.ar/modernizacion/empleopublico/biep/informes-estadisticos>

$$s^2 = \frac{48.708.208,11}{376.627} = 129,33^2$$

Es decir, la varianza es de 129,33².

4.3.3. Desvío estándar (s)

Como consecuencia de la modalidad de definición de la varianza, esta medida viene dada en unidades cuadráticas de la variable; por ello, es conveniente generar otra medida —que se desprenda de aquella— que esté dada en las mismas unidades de la variable. Esta nueva medida de

dispersión es el desvío estándar y no es más que el valor positivo de la raíz cuadrada de la varianza.

El desvío estándar mide la dispersión de los datos alrededor de la media aritmética. Por lo tanto, su uso es correcto siempre que la media sea la medida de centralización adecuada:

$$s = +\sqrt{s^2}$$

Ejemplo

En los tres casos anteriores:

$$s = +\sqrt{5^2} = 2,24$$

y

$$s = +\sqrt{129,33^2} = 11,37$$

4.3.4. Coeficiente de variación (CV)

Se utiliza para relacionar el tamaño de la media aritmética con la variabilidad de la variable. Su fórmula expresa la desviación estándar como porcentaje de la media aritmética, mostrando una interpretación relativa del grado de variabilidad independiente de la escala de la variable, a diferencia del desvío estándar.

A diferencia de la desviación típica, este coeficiente es fuertemente sensible ante cambios de origen en la variable. Por ello, es importante que todos los valores sean positivos y su media dé, por lo tanto, un valor positivo. A mayor valor del coeficiente de variación, mayor heterogeneidad de los valores de la variable; a menor CV, mayor homogeneidad de los valores de la variable.

Por ejemplo, si el CV es menor o igual al 20 %, la media aritmética es representativa del conjunto de datos, por ende, el conjunto de datos es «homogéneo». Por el contrario, si el CV es mayor al 20 %, el promedio no será representativo del conjunto de datos, por lo que resultará «heterogéneo».

Además, si se tienen dos o más muestras, utilizando la media aritmética y el desvío estándar, se puede calcular el coeficiente de variación de cada una. El menor CV indicará cuál es la muestra más representativa.

Igualmente, todo dependerá de qué es lo que se está midiendo con la variable en estudio. Es claro que si, por ejemplo, la variable es mortalidad por una enfermedad determinada, un porcentaje de 20 % sería altísimo. Lo mejor en estos casos es analizar el CV según la variable estudiada.

Se calcula con la siguiente fórmula:

$$CV = \frac{s}{\bar{X}}$$

donde:

s: desvío estándar.

\bar{X} : media aritmética.

Se puede mostrar en porcentaje calculando:

$$CV = \frac{s}{\bar{X}} * 100$$

4.3.5. Cómo interpretar el desvío estándar

Suponiendo que los datos muestren un histograma en forma de campana, es decir, cuando presenta una disminución suave y simétrica respecto del pico hacia los extremos, se cumplen estas condiciones:

$\bar{X} \pm s$: Este intervalo contiene aproximadamente el 68 % de los valores observados.

$\bar{X} \pm 2*s$: Este intervalo contiene aproximadamente el 95 % de los valores observados.

$\bar{X} \pm 3*s$: Este intervalo contiene aproximadamente el 99 % de los valores observados.

Ejemplo

Si tenemos los siguientes datos:

Cuadro 14

Ministerio	Puestos de trabajo		
	xi	Media aritmética	(xi - media aritmética) ²
Defensa	109.089	32.101,50	5.927.075.156,25
Seguridad	98.637		4.426.972.760,25
Hacienda	30.849		1.568.756,25
Justicia y Derechos Humanos	17.527		212.416.050,25
Relaciones Exteriores y Culto	2.724		863.037.506,25
Salud y Desarrollo Social	50.566		340.937.760,25
Producción y Trabajo	22.199		98.059.506,25
Educación, Cultura y Ciencia y Tecnología	21.268		117.364.722,25
Transporte	10.002		488.387.900,25
Interior, Obras Públicas y Vivienda	9.457		512.773.380,25
Presidencia de la Nación	7.896		585.906.230,25
Jefatura de Gabinete de Ministros	5.004		734.274.506,25
Total	385.218		

Fuente: elaboración propia a partir de Informe Trimestral de Empleo Público (ITEP), Año 1, N.º 2, marzo 2019.

cuya media aritmética es 32.101,50 y su desvío estándar es 192,73, la distribución sería la siguiente:

$$\bar{X} \pm s = 32.101,50 \pm 192,73$$

En el intervalo 31.908,77–32.294,23, se encuentra aproximadamente el 68 % de los valores observados.

$$\bar{X} \pm s = 32.101,50 \pm 2 * 192,73$$

En el intervalo 31.716,04–32.486,96, se encuentra aproximadamente el 95 % de los valores observados.

$$\bar{X} \pm s = 32.101,50 \pm 3 * 192,73$$

En el intervalo 31.523,31–32.679,69, se encuentra aproximadamente el 99 % de los valores observados.

Ahora bien, si tenemos dos o más muestras, podemos utilizar el CV para verificar cuál es la más representativa.

Ejemplo

Supongamos que hay tres muestras distintas de 15 números cada una. Se presentan, además, la media aritmética y el desvío estándar de cada una.

Cuadro 15

	Muestra 1	Muestra 2	Muestra 3
	29	89	60
	24	2	25
	60	40	84
	31	37	7
	71	100	93
	16	57	40
	10	45	25
	33	88	92
	55	71	91
	37	40	90

	65	29	95
	59	74	25
	51	44	54
	5	21	18
	31	18	47
Media aritmética	38,47	50,33	56,40
Mediana	33	44	54
Moda	31	40	25
Desvío estándar	20,63	28,75	32,11
CV	53,62	57,11	56,93

¿En cuál de las tres muestras la media aritmética es más representativa?

Como el menor CV corresponde a la muestra 1, esta es la que tiene la media aritmética más representativa, aunque su valor dependa, como se dijo más arriba, de qué es lo que esté midiendo la variable utilizada.

4.4. Medidas de tendencia no central

La mediana permite dividir a los términos ordenados de la serie en dos grupos que contienen la misma cantidad de elementos, pero también se podría dividir la serie en cuatro, diez o cien grupos, que contengan la misma cantidad de elementos. Así, surge la idea de cuartiles —generación de cuatro subgrupos—, deciles —diez subgrupos— y centiles —cien subgrupos—, para los cuales la modalidad de cálculo sigue la misma lógica que para la mediana —tengamos presente que el cuartil 2 coincide con la mediana, lo mismo sucede con el decil 5, etc.—.

Por su parte, el rango intercuartílico (RI) surge para solucionar en parte el inconveniente que presenta el rango, ya que el RI es una medida simple, que no se basa en valores extremos, sino que calcula el intervalo entre el

primer y el tercer cuartil. Esta medida se define como la distancia entre los valores entre los cuales se encuentra el 50 % central de los datos.

4.4.1. Cuartiles (Q_i), rango intercuartílico (RI), deciles (D_i) y percentiles (P_i)

a) Cálculo para datos no agrupados

Cantidad par de datos:

$$Q_i = \frac{i \cdot n}{4}$$

$$RI = \frac{Q_3 - Q_1}{2}$$

$$D_i = \frac{i \cdot n}{10}$$

$$P_i = \frac{i \cdot n}{100}$$

Cantidad impar de datos:

$$Q_i = \frac{i \cdot (n + 1)}{4}$$

$$RI = \frac{Q_3 - Q_1}{2}$$

$$D_i = \frac{i \cdot (n + 1)}{10}$$

$$P_i = \frac{i \cdot (n + 1)}{100}$$

donde:

i : orden del cuartil, decil o percentil.

n : cantidad de casos.

Ejemplo

Si tenemos una serie con cantidad par de números:

5	9	10	7	7	1	2	4	8	5	8	9	6	5	5	4	8	7	9	4	7	8	8	7	6	3	8	9	4	5
5	6	7	7	8	9	10	4	7	8	8	7	6	10	7	7	1	2	4	8	5	8	9	6	5	5	4	8	7	9

Primero, ordenamos en forma creciente:

1	1	2	2	3	4	4	4	4	4	4	4	5	5	5	5	5	5	5	5	5	6	6	6	6	6	7	7	7	7	
7	7	7	7	7	7	7	7	8	8	8	8	8	8	8	8	8	8	8	8	8	9	9	9	9	9	9	9	10	10	10

y calculamos el cuartil 1 y el cuartil 3, para luego poder calcular el rango intercuartílico:

$$Q_1 = \frac{1 \cdot 60}{4} = 15$$

$$Q_3 = \frac{3 \cdot 60}{4} = 45$$

Buscamos, en la lista ordenada, el valor que se encuentra en la posición 15. Entonces, el cuartil 1 es:

$$Q_1 = 5$$

Buscamos, en la lista ordenada, el valor que se encuentra en la posición 45. Entonces, el cuartil 3 es:

$$Q_3 = 8$$

Entonces, el rango intercuartílico es:

$$RI = \frac{8 - 5}{2} = 1,5$$

Calculamos ahora el decil 2:

$$D_2 = \frac{2 \cdot 60}{10} = 12$$

Buscamos, en la lista ordenada, el valor que se encuentra en la posición 12. Entonces, el decil 2 es:

$$D_2 = 4$$

Calculamos, finalmente, el percentil 60:

$$P_{60} = \frac{60 * 60}{100} = 36$$

Buscamos, en la lista ordenada, el valor que se encuentra en la posición 36. Entonces, el percentil 60 es:

$$P_{60} = 7$$

Para una cantidad impar de datos:

78	59	3	11	95	42	14	68	89
----	----	---	----	----	----	----	----	----

Primero, ordenamos en forma creciente:

3	11	14	42	59	68	78	89	95
---	----	----	----	----	----	----	----	----

Calculamos el cuartil 1 y el cuartil 3, para luego poder calcular el rango intercuartílico:

$$Q_1 = \frac{1 * (9 + 1)}{4} = 2,5$$

$$Q_3 = \frac{3 * (9 + 1)}{4} = 7,5$$

En este caso, los resultados no son enteros, por lo cual seguiremos el siguiente procedimiento:

Para el cuartil 1, buscamos, en la lista ordenada, el valor que se encuentra en la posición 2. Evidentemente, el cuartil 2 está entre la posición 2 y la posición 3. Restamos la posición 3 menos la posición 2, multiplicamos el resultado por la parte decimal de 2,5 (0,5) y a ese resultado lo sumamos al número que figura en la posición 2:

$$14 - 11 = 3$$

$$3 * 0,5 = 1,5$$

$$11 + 1,5 = 12,5$$

Entonces, el cuartil 1 es:

$$Q_1 = 12,5$$

Para el cuartil 3, buscamos, en la lista ordenada, el valor que se encuentra en la posición 7. Evidentemente, el cuartil 3 está entre la posición 7 y la posición 8. Restamos la posición 8 menos la posición 7, multiplicamos el resultado por la parte decimal de 7,5 (0,5) y a ese resultado lo sumamos al número que figura en la posición 7:

$$89 - 78 = 11$$

$$11 * 0,5 = 5,5$$

$$78 + 5,5 = 83,5$$

Entonces, el cuartil 3 es:

$$Q_3 = 83,5$$

Y el rango intercuartílico es:

$$RI = \frac{83,5 - 12,5}{2} = 48$$

Calculamos ahora el decil 2:

$$D_2 = \frac{2 * (9+1)}{10} = 2$$

Buscamos, en la lista ordenada, el valor que se encuentra en la posición 2. Entonces, el decil 2 es:

$$D_2 = 11$$

$$P_{73} = \frac{73 * (9+1)}{100} = 7,3$$

Como en el caso del cuartil 3, tenemos que el resultado no es un entero. Realizamos la misma operación:

$$89 - 78 = 11$$

$$11 * 0,3 = 3,3$$

$$78 + 3,3 = 81,3$$

Entonces, el percentil 73 es:

$$P_{73} = 81,3$$

b) Cálculo para datos agrupados en intervalos de clase

$$Q_i = L_{\text{inf}} + \frac{i * \frac{n}{4} - Fi_{-1}}{fi} * W_i$$

$$RI = \frac{Q_3 - Q_1}{2}$$

$$D_i = L_{\text{inf}} + \frac{i * \frac{n}{10} - Fi_{-1}}{fi} * W_i$$

$$P_i = L_{inf} + \frac{i * \frac{n}{100} - Fi_{-1}}{fi} * W_i$$

donde:

Fi_{-1} : frecuencia absoluta acumulada de la clase anterior a la clase mediana.

L_{inf} : límite inferior de la clase mediana.

fi : frecuencia absoluta de la clase mediana.

W_i : amplitud de la clase mediana.

i : orden del cuartil, del decil o del percentil.

Ejemplo

Tomamos nuevamente el caso de la cantidad de agentes pertenecientes a la APN según rango etario y calculamos el cuartil 1, el cuartil 3, el decil 3 y el percentil 65.

Cuadro 16

Rango etario (años)	Marca de clase	Cantidad de agentes			
	x_i	f_i	f_{ri}	F_a	F_{ra}
20-24	22	40.481	0,11	40.481	0,11
25-29	27	54.269	0,14	94.750	0,25
30-34	32	54.484	0,14	149.234	0,40
35-39	37	56.358	0,15	205.592	0,55
40-44	42	52.097	0,14	257.689	0,68
45-49	47	41.718	0,11	299.407	0,79
50-54	52	34.239	0,09	333.646	0,89
55-59	57	26.139	0,07	359.785	0,96
60-64	62	16.842	0,04	376.627	1,00
Total		376.627	1,00		

Fuente: elaboración propia con base en BIEP, <https://www.argentina.gov.ar/modernizacion/empleopublico/biep/informes-estadisticos>

El intervalo de la mediana es 35-39. Entonces,

$$Q_1 = 35 + \frac{1 * \frac{376.627}{4} - 149.234}{56.358} * 5 = 30,11$$

Es decir que el 75 % de los agentes tienen más de 30,11 años, o lo que es lo mismo, el 25 % de los agentes tienen 30,11 años o menos.

$$Q_3 = 35 + \frac{3 * \frac{376.627}{4} - 149.234}{56.358} * 5 = 46,82$$

Es decir que el 25 % de los agentes tienen más de 46,82 años, o lo que es lo mismo, el 75 % de los agentes tienen 46,82 años o menos.

Entonces, el rango intercuartílico es:

$$RI = \frac{46,82 - 30,11}{2} = 8,36$$

Para calcular el decil 3:

$$D_3 = 35 + \frac{3 * \frac{376627}{10} - 149.234}{56.358} * 5 = 31,78$$

Es decir que el 70 % de los agentes tienen más de 31,78 años, o lo que es lo mismo, el 30 % de los agentes tienen 31,78 años o menos.

$$P_{65} = 35 + \frac{65 * \frac{376627}{100} - 149.234}{56.358} * 5 = 43,48$$

Es decir que el 35 % de los agentes tienen más de 43,48 años, o lo que es lo mismo, el 65 % de los agentes tienen 43,48 años o menos.

4.5. Análisis de simetría

Cuando se desea analizar si los datos están distribuidos simétricamente respecto de la media aritmética, existen tres maneras de hacerlo: la comparación de la media aritmética con la mediana y la moda, la comparación relativa de la media aritmética con la mediana respecto del desvío estándar y el coeficiente de asimetría.

4.5.1. Comparación de la media aritmética con la mediana y la moda

Las tres medidas de tendencia central estudiadas pueden combinarse y compararse a fin de conocer la simetría o asimetría de una serie de datos.

4.5.1.1. Para una distribución que presenta un solo máximo

Si $\bar{X} = Me = Mo$ la serie es simétrica. De otra manera, es asimétrica.

4.5.1.2. Para una distribución que presenta dos o más máximos

Si $\bar{X} = Me$ y ambos máximos (si son dos) son simétricos respecto a la media aritmética, la serie es simétrica. De otra manera, es asimétrica.

Ejemplo

Retomando el ejemplo de los/las agentes de la APN distribuidos según rango etario, calcularemos la media aritmética, la mediana y la moda:

Cuadro 16A

Rango etario (años)	Marca de clase x_i	Cantidad de agentes f_i	$x_i * f_i$	Fa	Media aritmética	$((x_i - \text{Media aritmética})^2) * f_i$
20-24	22	40.481	890.582	40.481		11.548.100
25-29	27	54.269	1.465.263	94.750		7.672.122
30-34	32	54.484	1.743.488	149.234		2.586.470
35-39	37	56.358	2.085.246	205.592		201.316
40-44	42	52.097	2.188.074	257.689	38,89	503.887
45-49	47	41.718	1.960.746	299.407		2.743.880
50-54	52	34.239	1.780.428	333.646		5.884.729
55-59	57	26.139	1.489.923	359.785		8.572.863
60-64	62	16.842	1.044.204	376.627		8.994.842
Total		376.627	14.647.954			48.708.210

Fuente: elaboración propia con base en BIEP, <https://www.argentina.gob.ar/modernizacion/empleopublico/biep/informes-estadisticos>

Calculamos la media aritmética:

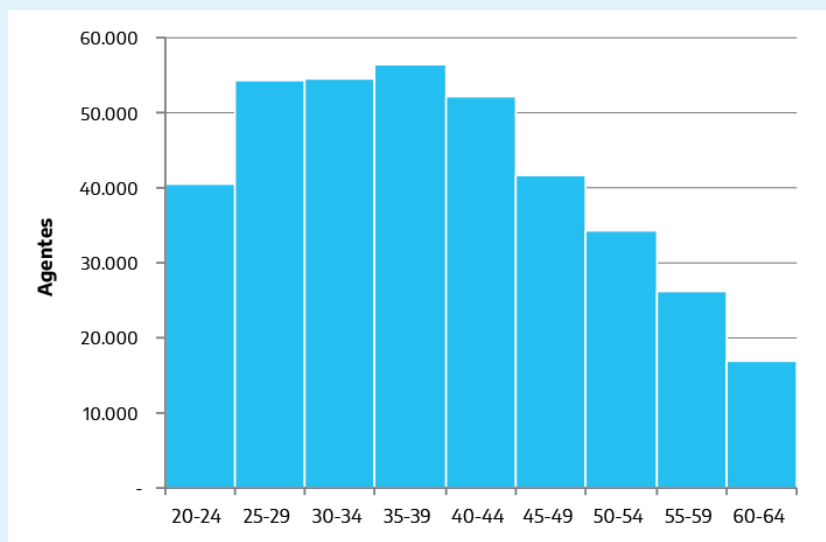
$$\bar{X} = \frac{\sum_{i=1}^n (x_i * f_i)}{n} = \frac{14.647.964}{376.627} = 38,89$$

$$Me = L_{inf} + \frac{\frac{n}{2} - Fi_{-1}}{fi} * W_i = 35 + \frac{\frac{376.627}{2} - 149.234}{376.627} * 5 = 38,47$$

$$Mo = L_{inf} + \frac{d_1}{d_1 + d_2} * w_i = 35 + \frac{56.358 - 54.484}{(56.358 - 54.484) + (56.358 - 52.097)} * 5 = 36,53$$

Puede verse que las tres medidas no son iguales, lo cual indica que la serie no es simétrica con respecto a la media aritmética. Visto gráficamente mediante un histograma de las frecuencias absolutas:

Gráfico 8



Fuente: elaboración propia con base en BIEP, <https://www.argentina.gob.ar/modernizacion/empleopublico/biep/informes-estadisticos>

El gráfico muestra una asimetría positiva, lo cual significa que hay más valores alejados de la media en la cola derecha.

4.5.2. Comparación relativa de la media aritmética con la mediana respecto del desvío estándar

Otra medida de la asimetría que permite ver el tipo de sesgo que presenta la serie —en el caso de ser asimétrica— es la que compara relativamente a la media aritmética con la mediana, de acuerdo a la siguiente fórmula:

$$A_s = \frac{\bar{X} - M_e}{s}$$

El denominador de esta relación siempre es positivo, por lo tanto, el signo de A_s viene dado por el numerador —si A_s es positivo, los datos están sesgados hacia la derecha, o sea, la cola se alarga a la derecha; si A_s es negativo, los datos están sesgados hacia la izquierda—. Esto se debe a que, en una distribución asimétrica con una cola muy larga en una dirección y un único pico, la media se ve desplazada en la dirección de la cola más larga. Tengamos presente que la media se ve influenciada por los valores extremos de la variable y por las frecuencias de aparición de cada uno de los valores de la variable.

Utilizando el mismo ejemplo del punto anterior y calculando el desvío estándar:

$$\bar{X} = 38,89$$

$$M_e = 38,47$$

Para calcular el desvío estándar, primero calculamos la varianza:

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{X})^2 * f_i}{N} = \frac{48.708.210}{376.627} = 129,33^2$$

Luego, calculamos el desvío estándar:

$$s = +\sqrt{s^2} = \sqrt{129,33^2} = 11,37$$

para finalmente calcular:

$$A_s = \frac{38,89 - 38,47}{11,37} = 0,04$$

Como el resultado es positivo aunque pequeño, la distribución se sesga levemente hacia la derecha.

4.5.3. Coeficiente de asimetría

Viene dado por el cociente entre el momento tercero de la variable (que está representado por el promedio de los desvíos de la variable con respecto a la media aritmética elevados al cubo) y el cubo del coeficiente de dispersión. Tengamos presente que el cubo de un número muy grande es mucho más grande que el cubo de un número pequeño y que, en datos sesgados a derecha, hay algunas desviaciones positivas relativamente grandes. Si los datos tienen distribución aproximada en forma de campana, este coeficiente se asemeja al valor cero:

$$CA_s = \frac{\frac{\sum (X_i - \bar{X})^3}{N}}{\left[\sqrt{\frac{\sum (X_i - \bar{X})^2}{N}} \right]^3}$$

Si $CA_s < 0$ datos sesgados hacia la izquierda.

Si $CA_s > 0$ datos sesgados hacia la derecha.

Si $CA_s = 0$ datos no sesgados.

Ejemplo

Cuadro 17

Rango etario (años)	Marca de clase xi	Cantidad de agentes fi	xi * fi	Media aritmética	(xi - Media aritmética)^2	(xi - Media aritmética)^3
20-24	22	40.481	890.582	38,89	285,36	-4.820,36
25-29	27	54.269	1.465.263		729,00	19.683,00
30-34	32	54.484	1.743.488		1.024,00	32.768,00
35-39	37	56.358	2.085.246		1.369,00	50.653,00
40-44	42	52.097	2.188.074		1.764,00	74.088,00
45-49	47	41.718	1.960.746		2.209,00	103.823,00
50-54	52	34.239	1.780.428		2.704,00	140.608,00
55-59	57	26.139	1.489.923		3.249,00	185.193,00
60-64	62	16.842	1.044.204		3.844,00	238.328,00
Total		376.627	14.647.954			17.177,36

Fuente: elaboración propia con base en BIEP, <https://www.argentina.gob.ar/modernizacion/empleopublico/biep/informes-estadisticos>

$$CA_s = \frac{840.323,64}{\left[\frac{376.627}{\sqrt{\frac{17.177,36}{376.627}}} \right]^3} = \frac{2,23}{0,60} = 3,73$$

Se confirma, entonces, que la serie está sesgada a la derecha.

4.6. Práctica



Ejercicio 1

Un estudiante se encuentra promediando la carrera y pretende revisar su analítico, el cual consta de las siguientes notas:

Cuadro 17A

4	4	4	6	7
8	4	4	2	5
10	4	6	7	8
2	10	4	6	7

- ¿Qué tipo de variable se está analizando?
- Hallar la moda y explicar su significado.
- Hallar la mediana y explicar su significado.
- Hallar la media aritmética y el desvío estándar
- Hallar el rango
- Hallar el decil 3
- Hallar el percentil 60
- Hallar el cuartil 1
- Realice una comparación de la media aritmética, la mediana y la moda y elabore un análisis como el realizado en el punto 4.5.1.



Ejercicio 2

Según los registros de una fábrica, la cantidad mensual de llegadas tarde de los operarios fue la siguiente:

Cuadro 17B

Llegadas tarde	Cantidad de empleados
x_i	f_i
0	13
1	18
2	6
3	2
4	1
Total	40

- ¿Cuál es la cantidad mensual de llegadas tarde que más veces se presenta?
- ¿Cuál es el promedio de llegadas tarde?
- Calcular el decil 4.
- Calcular el percentil 80.



Ejercicio 3

Los resultados de un test de aptitud tomado a un grupo de 100 personas se vuelcan en la siguiente tabla:

Cuadro 17C

Intervalo		fi
0,50	5,50	7
5,50	10,50	12
10,50	15,50	21
15,50	20,50	32
20,50	25,50	28
Total		100

- ¿Cuál es el intervalo modal?
- ¿En qué intervalo se encuentra la mediana?
- Calcular la media aritmética.
- Calcular la varianza y el desvío estándar.



Ejercicio 4

El número de horas extras realizadas por semana en una dependencia del PAMI es un factor que influye en la mejora de la atención a los jubilados. Una muestra aleatoria de 25 empleados produjo las siguientes estimaciones del tiempo (medido en horas):

Cuadro 17D

Intervalo	
3	6
7,5	15
12	6,5
8	4
5,5	6
5	12
1	3,5
3	7,5
5	10
8	3,5
9	2
6,5	1
5	

Calcular:

- La media aritmética.
- La varianza y el desvío estándar.
- El coeficiente de variación e interpretar.
- El percentil 65 y explicar.
- El decil 7 y explicar.
- El cuartil 3 y explicar.
- El rango intercuartílico.



Ejercicio 5

El cuadro siguiente representa los conceptos fijos liquidados brutos promedio del personal civil de la Administración Nacional y de los entes públicos en escalafones seleccionados a octubre de 2018:

Cuadro 17E

Escalafón	Conceptos fijos
AFIP	136.384
PAMI	68.025
ANSES	65.107
CONICET	53.420
INTA	49.291
PECIFA	39.250
SINEP	36.382

Fuente: Informe Trimestral de Empleo Público (ITEP), Año 1, N.º 2, marzo 2019.

- Calcular la media aritmética.
- Calcular la mediana y el modo.
- Calcular la varianza y el desvío estándar. Interpretar el desvío estándar.
- Calcular el percentil 84 y explicar.
- Calcular la asimetría (punto 4.5.2) y analizar.
- Calcular el coeficiente de asimetría (punto 4.5.3) y analizar.

5

Indicadores, relaciones entre variables y cifras

5.1. Introducción³

En los capítulos anteriores, los ejemplos presentados involucraban a unidades y usualmente a una sola variable por vez. Sin embargo, en la práctica es muy común que un mismo estudio combine más de una variable o cifras pertenecientes a las mismas unidades o que varios conjuntos de unidades se combinen en diversas relaciones.

Este capítulo se refiere a los casos en los cuales en un mismo problema se combinan variables o cifras diversas, provenientes de uno o varios conjuntos de unidades, mediante las cuales se confeccionan «indicadores» o «relaciones», que condensan la información básica y enriquecen el análisis.

5.2. Concepto de indicador

Suele denominarse así a «todo tipo de información que muestra (indica) la situación en un momento dado, o la evolución o modificaciones que registra una unidad a través del tiempo».

³ Parte del contenido está basado en SEDRONAR (2013).

En estadística, es usual denominar «indicador» a «las relaciones lógicas entre variables o cifras que resumen información diversa en unos pocos números, que se utilizan para mostrar el estado de una unidad de análisis o su evolución a través del tiempo, o la comparación entre varias unidades».

Ejemplos de indicador:

- La *población* (cantidad de personas que habitan en una región) de un país es un indicador de su tamaño o importancia con relación a otros países o regiones.
- El *Producto Bruto Interno* (PBI) (suma de los resultados de las actividades económicas de una región) es otro indicador de su tamaño o importancia con relación a otras regiones.
- El *PBI por habitante* es un indicador del grado de desarrollo o del nivel económico de las personas que habitan en un determinado país.
- La *composición de las exportaciones* de un país sugiere el tipo de actividades a que se dedica o, por lo menos, aquellas con las que comercia.
- El *porcentaje de participación del sector agropecuario en el PBI* es un indicador de la importancia de ese sector en la actividad económica del país.

En todos estos ejemplos, la unidad de análisis es el/los país/es.

5.3. Tipo de relaciones entre variables o cifras

En términos generales, los indicadores o relaciones entre variables o cifras que se pueden utilizar dentro de la estadística descriptiva son los siguientes:

- Proporciones
- Promedios
- Índices
- Tasas
- Otros

A continuación, veremos un resumen de cada uno de los indicadores citados.

5.3.1. Proporciones

En el capítulo 2, analizamos un caso particular de proporciones que son las frecuencias relativas (f_{ri}) que, entre las distribuciones de frecuencias, constituyen una proporción, ya que representan la importancia de cada uno de los valores de la variable (a través de su frecuencia absoluta, f_i), con relación al total de las unidades consideradas (n).

Lo usual en la práctica es que las relaciones entre la parte y el total se encuentren multiplicadas por algún número redondo (10, 100, 1000, etc.), a fin de obtener resultados mayores que 1 y más fáciles de interpretar.

La fórmula de cálculo para las proporciones es:

$$P_i = (C_i / T) * N^\circ$$

donde:

C_i : cifra correspondiente a la variable x_i (mide la importancia o presencia o cantidad de veces que se repite ese valor de la variable y, en el caso de las distribuciones de frecuencias, se simboliza con f_i).

T: total de unidades consideradas, dentro de las que forma parte C_i (en las distribuciones de frecuencias es n).

k: resultados diferentes de la variable x_i , cantidad de alternativas distintas que se dan entre las T unidades.

N° : puede ser 100, 10, 1000, etc.

Algunos ejemplos de proporciones son:

- Distribución en porcentajes entre los diversos niveles educativos alcanzados de la población de una región.
- Participación o penetración de una empresa en el mercado.
- Proporción de unidades defectuosas en un proceso productivo (con relación al total de unidades producidas).

Ejemplo

En el siguiente cuadro se muestra la cantidad de puestos de trabajo de 12 ministerios de la APN a marzo de 2019:

Cuadro 18

Ministerio	Puestos de trabajo
Defensa	109.089
Seguridad	98.637
Hacienda	30.849
Justicia y Derechos Humanos	17.527
Relaciones Exteriores y Culto	2.724
Salud y Desarrollo Social	50.566
Producción y Trabajo	22.199
Educación, Cultura y Ciencia y Tecnología	21.268
Transporte	10.002
Interior, Obras Públicas y Vivienda	9.457
Presidencia de la Nación	7.896
Jefatura de Gabinete de Ministros	5.004
Total	385.218

Fuente: Informe Trimestral de Empleo Público (ITEP), Año 1, N.º 2, marzo 2019.

Determinar el porcentaje de cada uno dentro del total:

Cuadro 19

Ministerio	Puestos de trabajo	%
Defensa	109.089	28,3
Seguridad	98.637	25,6
Hacienda	30.849	8,0
Justicia y Derechos Humanos	17.527	4,5
Relaciones Exteriores y Culto	2.724	0,7
Salud y Desarrollo Social	50.566	13,1
Producción y Trabajo	22.199	5,8
Educación, Cultura y Ciencia y Tecnología	21.268	5,5
Transporte	10.002	2,6
Interior, Obras Públicas y Vivienda	9.457	2,5
Presidencia de la Nación	7.896	2,0
Jefatura de Gabinete de Ministros	5.004	1,3
Total	385.218	100,0

Fuente: elaboración propia a partir de Informe Trimestral de Empleo Público (ITEP), Año 1, N.º 2, marzo 2019.

5.3.2. Promedios

Un promedio queda siempre expresado en unidades de medidas compuestas, dependiendo de las unidades de medida de las variables que se relacionan.

La fórmula de cálculo para los promedios es:

$$P_{Ri} = (CV_{1i} / CV_{2i})$$

donde:

$i = 1 - n$ (cantidad total de unidades consideradas).

CV_{1i} : cifra correspondiente a la variable 1 en la unidad i -ésima.

CV_{2i} : cifra correspondiente a la variable 2 en la unidad i -ésima.

Algunos ejemplos de promedio son:

- Medición del producto bruto interno por habitante en un conjunto de países (en dólares o miles de dólares por persona).
- Productividad de un conjunto de empresas (medida según la producción con relación a la cantidad de personas ocupadas en cada una en pesos o kilogramos por persona).
- Producción de trigo por hectárea cultivada en los establecimientos agropecuarios de la provincia de Buenos Aires (en toneladas o bolsas por hectárea).
- Monto de las deudas de las empresas con relación al patrimonio con que cuentan para enfrentarlas (pesos por pesos).

Ejemplo

La Secretaría de Energía realiza un estudio comparativo sobre la productividad de las estaciones de servicio pertenecientes a tres empresas: una estatal (C) y dos privadas (A y B). Para realizar el estudio se recopila la siguiente información:

Cuadro 20

Empresa	Cantidad de		
	Estaciones	Empleados	Combustible vendido (m^3)
A	10	80	1.200
B	15	90	1.080
C	25	300	6.000

Determinar:

- Cuál de las tres empresas tiene mayor productividad en sus ventas.
- En cuál de las empresas las estaciones tienen un menor tamaño relativo de personal.

Sabiendo que (i) = empresas, (Esi) = cantidad de estaciones de servicio, (Emi) = cantidad de empleados, (Ci) = venta de combustible, podemos armar el siguiente cuadro:

(i)	Pri = Ci / Emi (m ³ por empleado)	Pri = Emi / Esi (empleados por estación)
A	1.200 / 80 = 15	80/10 = 8
B	1.080/90 = 12	90/15 = 6
C	6.000/300 = 20	300/25 = 12

- La empresa más productiva es la estatal (C) (20 m³ por empleado).
- La empresa con estaciones más pequeñas es la (B), pues tiene un promedio de 6 personas por estación.

5.3.3. Índices (simples)

Desde el punto de vista de la aplicación práctica, se los puede clasificar en cronológicos o sectoriales.

- Índices cronológicos: son «aquellos en los que se relacionan cifras de una variable medidas sobre una misma unidad (o conjunto de ellas), a través del tiempo, tomando un cierto período como comparación o base de referencia».

Algunos ejemplos de índices son:

- Índice del precio de un artículo producido por una empresa, tomando como comparación el precio que el 30 de junio de 1998.
 - Índice de mano de obra ocupada de una ciudad, comparando la población ocupada en la Encuesta Permanente de Hogares que realiza el INDEC, con respecto a la registrada en la toma realizada en octubre de 1996.
 - Índice de los juicios que se registran en el mercado asegurador a través del tiempo, utilizando como comparación los juicios existentes en un cierto momento o año.
- Índices sectoriales/geográficos/etc.: son «aquellos en los que se relacionan cifras de una misma variable en un mismo período de tiempo pertenecientes a distintas unidades (sectoriales, regionales etc.), tomando una de estas como comparación o base de referencia».

Algunos ejemplos de índices son:

- Índice de Precios al Consumidor (IPC).
- Productividad de las ramas económicas del Censo Económico Nacional, utilizando una de esas ramas como base de comparación.

La fórmula de cálculo para los índices es:

$$I_{i/o} = (C_i / C_o) * 100$$

donde:

$$i = 1 - n$$

I_i : número índice (simple), correspondiente al momento (en el caso de los cronológicos), sector, región, etc. (en el caso de los restantes), i -ésimo, sobre la base del momento, sector, región, etc. Y 0 se toma como referencia o punto de comparación.

C_i : cifra correspondiente al momento, sector, región, etc., i -ésimo.

C_0 : cifra correspondiente al momento, sector, región, etc. Y 0 se toma como base o referencia.

Puede observarse en la fórmula que, a cualquiera de ambos tipos de índices, se los suele presentar en porcentaje.

Ejemplo

Tomando el mismo caso de las estaciones de servicio, calculamos los índices de productividad y tamaño con relación a la empresa A, que se transforma en la base 100:

Cuadro 21

Empresa (i)	$li/a = (PR_i/PR_a)*100$	$li/a = (PR_i/PR_a)*100$
A	$(15/15)*100 = 100,0$	$(8/8)*100 = 100,0$
B	$(12/15)*100 = 133,3$	$(6/8)*100 = 75,0$
C	$(20/15)*100 = 80,0$	$(12/8)*100 = 150,0$

Con respecto a la empresa A, los índices sectoriales indican:

- La empresa C tiene una productividad 33,3 % mayor, mientras que la B muestra una productividad 20 % menor.
- La empresa C es un 50 % mayor que la A en lo que respecta al promedio de empleados por estación, mientras que la empresa B es un 25 % más chica.

5.3.4. Tasas

Este indicador puede ser clasificado en dos tipos conceptualmente diferentes: simples o de incremento o evolución.

- Tasas simples: «son relaciones entre dos variables medidas sobre una misma unidad, con la particularidad de que el denominador es mayor que el numerador, y, por lo tanto, para facilitar su lectura y comprensión, su resultado se multiplica por algún número redondo (100, 1000, etc.).»

Técnicamente, las tasas son promedios o proporciones que cumplen las condiciones particulares enunciadas en la definición.

El caso de los promedios es el más frecuente entre las tasas y son aquellos cuyos resultados directos dan menor que uno y cuando se los multiplica por algún número redondo quedan expresados «en porcentajes», «por miles», etc.

Podemos citar los siguientes ejemplos de tasas simples:

- Tasa de médicos por habitantes (usualmente se expresa por 10.000).
- Tasa de mortalidad (muertos cada 1000 habitantes).

También, hay ejemplos de tasas que se confunden con las proporciones:

- Tasa de desempleo: total de desocupados sobre el total de la población económicamente activa (se lo presenta en porcentajes).
- Tasa de analfabetismo: cantidad de analfabetos sobre el total de la población en edad escolar (se lo expresa en porcentajes).

La fórmula de cálculo de la tasa simple es:

$$T_i = (Y_i / X_i) * N^\circ$$

donde:

$$i = 1 - n.$$

T_i : tasa correspondiente al momento o lugar i -ésimo.

Y_i : valor de la variable y_i en el momento o lugar i -ésimo.

X_i : valor de la variable x_i en el momento o lugar i -ésimo.

N° : número redondo en el cual quedan expresados los resultados.

- Tasas de incremento o evolución: «es la relación entre el incremento producido en una variable entre dos momentos del tiempo, y la cifra correspondiente al momento base o inicial».

Este es un tipo de relación muy particular, que puede ser asociada a los índices cronológicos (ya que ambos miden cambios a través del tiempo).

Podemos plantear como ejemplos de tasas de incremento o evolución:

- Crecimiento del precio de los productos básicos de un mes al otro, tomando como comparación el mes con que se compara cada precio.
- Incremento (o disminución) de la producción automotriz de un año al siguiente (expresado con relación al año anterior).

La fórmula de cálculo de la tasa de incremento es:

$$t_i = [(V_i / V_{i-1}) / V_{i-1}] * N^\circ$$

donde:

V_i : valor de la variable en análisis en el momento i -ésimo.

V_{i-1} : valor de la variable en el momento $i-1$ ésimo.

N° : número redondo en el cual quedan expresados los resultados.

Si es necesario calcular las tasas de evolución en las que se comparen períodos no contiguos, debe tenerse en cuenta la forma en que puede evolucionar la variable en análisis. Según lo que se suponga o compruebe, se podrá tener:

$$t_{i+k} = 1 / k \left[(V_{i+k} / V_i) / V_i \right] * N^\circ$$

Si se adoptara que la evolución de la variable es del tipo lineal o rectilínea, puede calcularse la tasa de evolución dividiendo la diferencia entre las variables por la cantidad de períodos transcurridos.

Si se pensara o comprobara que la evolución no es lineal, deberían hacerse supuestos de posibles evoluciones antes de calcular las tasas de incremento.

Ejemplo

En un trabajo de tipo demográfico para tres países, se dispone de la siguiente información:

Cuadro 22

País (i)	Población total (en miles) (P_{ti})	Población Económicamente Activa (PEAI) (en miles)
1	25.000	10.000
2	150.000	45.000
3	8.000	2.800

Se desea conocer la importancia de la PEA con respecto a la población total del país.

Cuadro 23

País (i)	Población total (en miles) (Pti)	Población Económicamente Activa (PEAi) (en miles)	$T_i = (PEA_i / P_{ti}) * 100$ (en %)
1	25.000	10.000	$(10.000/25.000)*100 = 40,0$
2	150.000	45.000	$(45.000/150.000)*100 = 30,0$
3	8.000	2.800	$(2.800/8.000)*100 = 35,0$

Estos resultados se leen de la siguiente manera:

En el país 1, la PEA representa el 40 % de la población total.

En el país 2, la PEA representa el 30 % de la población total.

En el país 3, la PEA representa el 35 % de la población total.

5.3.5. Otros indicadores

Cualquier otro tipo de indicador o relación no incluido en los anteriores quedaría «por diferencia» comprendido en esta categoría «residual». En particular, dentro de la metodología estadística, existen algunas medidas especiales que pueden quedar comprendidas en esta categoría, como, por ejemplo:

- El coeficiente de variación utilizado como medida de dispersión.
- Los coeficientes de la recta de regresión y de correlación (que corresponden al curso de estadística avanzada).

5.4. Práctica



Ejercicio 1

Partiendo de los datos de la cantidad de puestos de trabajo de 12 ministerios de la APN a marzo de 2019, expresar la importancia relativa de cada uno con relación al Ministerio de Defensa. Explicar los índices del Ministerio de Transporte, de la Jefatura de Gabinete de Ministros y del Ministerio de Justicia y Derechos Humanos respecto al Ministerio de Defensa.

Cuadro 23A

Ministerio	Puestos de trabajo
Defensa	109.089
Seguridad	98.637
Hacienda	30.849
Justicia y Derechos Humanos	17.527
Relaciones Exteriores y Culto	2.724
Salud y Desarrollo Social	50.566
Producción y Trabajo	22.199
Educación, Cultura y Ciencia y Tecnología	21.268
Transporte	10.002
Interior, Obras Públicas y Vivienda	9.457
Presidencia de la Nación	7.896
Jefatura de Gabinete de Ministros	5.004
Total	385.218

Fuente: Informe Trimestral de Empleo Público (ITEP), Año 1, N.º 2, marzo 2019.



Ejercicio 2

En el último cuatrimestre del año anterior, la producción mensual de una empresa laminadora de acero fue la siguiente:

Cuadro 24

Meses	Producción (en toneladas)
Septiembre	62
Octubre	80
Noviembre	75
Diciembre	112

Calcular el índice de producción con respecto a octubre.



Ejercicio 3

Para un conjunto de empresas asociadas a la Cámara del Vidrio, se recopiló la siguiente información:

Cuadro 25

Empresa	Producción		Insumos	Exportación		Personal		Superficie
	Toneladas	Miles \$	Miles \$	Toneladas	Miles \$	Total	Femenino	m ²
i	P _{ti}	P _i	I _i	E _{ti}	E _i	P _{Ti}	P _{Fi}	S _i
1	1	9.000	1.800	0,2	3.000	50	40	600
2	6	15.000	4.800	0,6	1.500	150	30	2.400
3	30	24.000	10.800	10,2	6.000	400	200	10.000
4	13	22.000	6.600			200	50	3.000
5	300	70.000	35.000			2.000	1.650	16.000
6	150	60.000	24.000	6,0	3.000	1.200	30	12.000

Calcular:

- El porcentaje de participación de cada empresa (en toneladas y en pesos) en la producción total del sector del vidrio.
- La productividad (valor de producción por persona ocupada) de cada empresa y del total del sector.
- El índice de personal ocupado con relación a la empresa 3.
- La participación del personal femenino en el total del personal de cada empresa y en el total del sector.
- La participación de las exportaciones en la producción de cada empresa (en toneladas y en pesos).

Indicar qué tipo de indicador se calcula en cada caso.



Ejercicio 4

En un informe de población para cuatro países, se dispone de la siguiente información:

Cuadro 26

País (i)	Población total (en miles)	Mujeres (en miles)	Varones (en miles)
A	25.000	13.250	11.750
B	150.000	72.000	78.000
C	8.000	3.760	4.240
D	44.000	21.560	22.440

Calcular la proporción de mujeres y varones para cada país.



Ejercicio 5

El consumo anual de combustible de una empresa (en miles de litros) se muestra en el siguiente cuadro:

Cuadro 27

Año	Consumo
2000	60
2001	70
2002	90
2003	95
2004	120
2005	80
2006	95
2007	125
2008	130
2009	122
2010	95
2011	98
2012	110
2013	80
2014	85

Calcular el índice de consumo anual con base en el año 2004.

Referencias bibliográficas

- Biondolillo, G. (2017). «Estadística descriptiva. Guía completa». Recuperado de <https://www.academia.edu/33639267/>. [Fecha de consulta: 17/4/2020].
- SEDRONAR. (2013). «Manual de Estadística Básica». Recuperado de <http://observatorio.gob.ar/media/k2/attachments/>. [Fecha de consulta: 17/4/2020].
- Real Academia Española. *Diccionario de la Lengua Española* (2019). Recuperado de <https://dle.rae.es/estadístico>. [Fecha de consulta: 17/4/2020].

CUINAP | Argentina, Cuadernos del INAP

Año 1 - N.º 13 - 2020

Instituto Nacional de la Administración Pública

Av. Roque Sáenz Peña 511, Ciudad Autónoma de Buenos Aires, Argentina

C. P.: C1035AAA - Tel.: 4343-9001

Correo electrónico: cuinap@jefatura.gob.ar

ISSN 2683-9644

Editor responsable

Alejandro M. Estévez

Idea original

Carlos Desbouts

Edición y corrección

Patricia Iacovone

Arte de tapa

Roxana Pierri

Federico Cannone

Diseño y diagramación

Roxana Pierri

Las ideas y planteamientos contenidos en la presente edición son de exclusiva responsabilidad de sus autores y no comprometen la posición oficial del INAP.

INAP no asume responsabilidad por la continuidad o exactitud de los URL de páginas web externas o de terceros referidas en esta publicación y no garantiza que el contenido de esas páginas web sea, o continúe siendo, exacta o apropiada.

Los Cuadernos del INAP y su contenido se brindan bajo una Licencia Creative Commons Atribución-No Comercial 2.5 Argentina. Es posible copiar, comunicar y distribuir públicamente su contenido siempre que se cite a los autores individuales y el nombre de esta publicación, así como la institución editorial.

El contenido de los Cuadernos del INAP no puede utilizarse con fines comerciales.

Esta publicación se encuentra disponible en forma libre y gratuita en: publicaciones.inap.gob.ar

Junio 2020

Secretaría de
Gestión y Empleo Público



Jefatura de
Gabinete de Ministros
Argentina