

# ORIENTACIONES PARA LA SELECCIÓN Y EL CÁLCULO DEL TAMAÑO DE LA MUESTRA EN INVESTIGACIÓN



JORGE SUCASAIRE PILCO



## **Jorge Sucasaire Pilco**

Nacido en la ciudad de Lima. Es docente egresado de la Facultad de Educación de la Universidad Nacional Mayor de San Marcos con especialidad en Matemática y Física. Maestro y Doctor por la Universidad César Vallejo. Ha preparado y dictado cursos de Matemática y Estadística en el Instituto Superior Tecnológico Paul Müller, Instituto de Educación Superior Avansys (Cibertec), Instituto de Formación Bancaria IFB Certus, Instituto Superior IDAT, Centro Preuniversitario de la Universidad Autónoma del Perú, Universidad Privada César Vallejo, Universidad Tecnológica del Perú, Universidad Privada del Norte y la Universidad Le Cordon Bleu.

**ORIENTACIONES PARA LA  
SELECCIÓN Y EL CÁLCULO DEL  
TAMAÑO DE LA MUESTRA EN  
INVESTIGACIÓN**

©Jorge Sucasaire Pilco  
*Primera edición digital – marzo 2022*

Editado por:  
Jorge Sucasaire Pilco  
Jr. Los Lirios 380, Urb. La Primavera – El  
Agustino, Lima - Perú.  
E-mail: jorgesuca1982@gmail.com

**Libro electrónico disponible en:**  
<https://repositorio.concytec.gob.pe/>

Hecho el Depósito Legal en la Biblioteca  
Nacional del Perú N° 2022-02488

ISBN 978-612-00-7547-0

## ORIENTACIONES PARA LA SELECCIÓN Y EL CÁLCULO DEL TAMAÑO DE LA MUESTRA EN INVESTIGACIÓN

Diseño de portada: Jorge Sucasaire Pilco

No se permite reproducir, almacenar en sistemas de recuperación de la información ni transmitir parte o el total de esta publicación, cualquiera que sea el medio empleado, sin el permiso previo del autor o editor

Lima-Perú



## **Presentación**

Entre las primeras dificultades que se presentan en el trabajo del investigador se pueden considerar a la delimitación de su población y la selección de una muestra para su estudio. La delimitación de la población se consigue analizando los objetivos y el alcance de la investigación, pero para seleccionar la muestra se tiene que observar otros aspectos, en muchos casos, relacionados con la estadística inferencial. En esta etapa de la investigación surgen las preguntas: ¿Cuántos elementos deben integrar la muestra para mi estudio? y ¿Cómo se eligen los elementos para esta muestra? Con este libro se trata de brindar orientaciones para dar respuesta a estas preguntas con base en la comprensión del diseño de investigación y las reglas de la inferencia estadística.

En un segundo plano, se trata de evidenciar la noción incorrecta que tienen algunos investigadores con respecto a la cantidad de expresiones matemáticas que permiten determinar el tamaño muestral. En consecuencia, en los diferentes capítulos se esclarece que no existe una fórmula única o cuatro fórmulas únicas para lograr dicho propósito. Así, se desarrolla una gran variedad de expresiones que permiten realizar el cálculo del tamaño de la muestra, cuya elección depende de criterios que deben estar bien definidos en el marco metodológico de la investigación.

En la actualidad existe software que permite estimar el tamaño de la muestra para los diferentes tipos de investigación. La responsabilidad del investigador recae en hacer una elección adecuada del procedimiento para el cálculo del tamaño y la selección de la muestra. Por ello, el investigador debe tener cierto dominio de las nociones básicas de los diseños de investigación y de los términos necesarios del lenguaje de la inferencia estadística, de modo que se elabore el sustento adecuado para la metodología en el proceso de la toma de muestra.

En el presente texto se trata de utilizar la notación formal de las expresiones matemáticas en lo más mínimo posible para que el lector pueda comprender con facilidad los procesos y elementos del muestreo. Este trabajo es de utilidad para los profesionales que se inician en la actividad investigativa y para su comprensión no se requiere un entrenamiento previo especial en estadística o matemática.

La teoría del muestreo es una rama de la estadística cuyos contenidos son muy amplios. Por ello, no se tratará de explicar la teoría del muestreo, sino que se utilizarán algunas de sus fórmulas para derivar de ellas las expresiones que nos permitan realizar el cálculo del tamaño de la muestra. Sin embargo, se abarca el procedimiento básico para diferentes diseños de investigación que pueden aparecer en distintas ramas del conocimiento, de modo que, los contenidos que se abordan para el proceso de muestreo sirven al investigador que se adentra en campos de estudio como la ingeniería, la administración, la psicología, la gastronomía, las ciencias de los alimentos y las ciencias de la educación.

Para lograr una mejor comprensión del cálculo del tamaño de muestra, en muchos casos optaremos por prescindir de expresiones matemáticas complejas y se brindarán las rutas para el proceso de muestreo en programas como SPSS, Minitab, Excel o G\*Power. También se brindan tablas con tamaños de muestra sugeridos para diferentes situaciones de muestreo. Finalmente, se presentan las referencias bibliográficas completas para el lector que desee realizar un estudio más minucioso del siempre importante proceso de selección de la muestra.

*Jorge Sucasaire Pilco*

## Tabla de contenido

Capítulo 1: Población y muestra .....	12
1.1. Población .....	12
1.2. Delimitación de la población .....	13
1.3. Muestra .....	16
1.4. Notación para el tamaño de la población y de la muestra .....	17
1.5. Unidad de análisis .....	18
1.6. Marco muestral.....	19
Capítulo 2: Diseños de investigación .....	20
2.1. Diseños no experimentales u observacionales.....	21
2.1.1. Diseño transversal .....	22
2.1.2. Diseño longitudinal.....	23
2.1.3. Diseño retrospectivo .....	23
2.1.4. Diseño prospectivo .....	24
2.1.5. Diseño correlacional .....	25
2.1.6. Diseño correlacional-causal .....	26
2.1.7. Estudio de cohortes .....	27
2.1.8. Estudios de casos y controles.....	28
2.2. Diseños experimentales .....	29
Capítulo 3: Elementos básicos de estadística inferencial .....	32
3.1. La inferencia estadística .....	33
3.2. Estadígrafo .....	33
3.3. Parámetro .....	34
3.4. Clasificación de la estadística inferencial .....	34
3.4.1. Estimación de parámetros .....	34



3.4.2. Prueba de hipótesis .....	36
3.5. La distribución normal .....	37
Capítulo 4: Selección de los elementos para la muestra .....	42
4.1. Muestreo .....	42
4.2. Tipos de muestreo.....	43
4.2.1. Muestreo probabilístico.....	44
4.2.1.1. Muestreo aleatorio simple .....	45
4.2.1.2. Muestreo sistemático .....	58
4.2.1.3. Muestreo estratificado .....	62
4.2.1.4. Muestreo por conglomerados.....	73
4.2.1.5. ¿Qué tipo de muestreo probabilístico debo utilizar? .....	77
4.2.2. Muestreo no probabilístico.....	79
4.2.2.1. Por conveniencia o accidental .....	80
4.2.2.2. Juicio de expertos o de criterio .....	80
4.2.2.3. Por cuotas.....	81
4.2.2.4. Bola de nieve .....	82
4.2.2.5. ¿Qué tipo de muestreo no probabilístico debo utilizar? ....	83
Capítulo 5. Cálculo del tamaño de la muestra.....	84
5.1. Variable cualitativa y variable cuantitativa.....	88
5.2. Población finita y población infinita.....	89
5.2.1. Población finita.....	89
5.2.2. Población infinita.....	90
5.3. Tamaño de la muestra para variable cualitativa.....	91
5.3.1. Tamaño de la muestra para estimar la proporción .....	91
5.3.1.1. Para una población infinita .....	92
Proporción $p$ y $q$ .....	93
¿Qué hacer cuando no se conoce $p$ ?.....	95

Error de estimación de $p$ .....	97
¿Qué sucede con el tamaño de muestra cuándo $e$ toma un valor pequeño? .....	98
El valor $Z$ .....	99
Ejemplos sobre cálculo del tamaño de muestra .....	103
5.3.1.2. Para población finita .....	106
El tamaño de población $N$ .....	107
¿Qué sucede cuando se trabaja con un error pequeño?.....	108
Ejemplos sobre cálculo del tamaño de muestra .....	108
5.3.2. Tamaño de la muestra para la diferencia de proporciones .....	111
5.3.3. Tamaño de la muestra para estudios de casos y controles .....	113
5.4. Tamaño de la muestra para variable cuantitativa .....	115
5.4.1. Tamaño de la muestra para estimar la media .....	115
5.4.1.1. Para una población infinita .....	115
Desviación estándar poblacional $\sigma$ .....	115
Ejemplos sobre cálculo del tamaño de muestra .....	119
5.4.1.2. Para población finita .....	121
Ejemplos sobre cálculo del tamaño de muestra .....	121
5.4.2. Tamaño de muestra para la diferencia de dos medias en grupos independientes .....	124
La diferencia entre medias $D$ .....	127
5.4.3. Tamaño de muestra para la diferencia de dos medias en muestras relacionadas .....	129
5.4.4. Tamaño de la muestra para asociar dos variables cuantitativas utilizando el coeficiente de correlación de Pearson .....	130
5.5. Tamaño de la muestra para diseños experimentales .....	133
5.5.1. Tamaño de muestra para una regresión lineal simple .....	138
5.5.1. Tamaño de muestra para un ANOVA .....	140

5.6. Diseño, objetivo, método estadístico y tamaño de la muestra .....	142
Capítulo 6: Preguntas frecuentes sobre la muestra .....	145
¿Se debe incluir el proceso del cálculo del tamaño de muestra en la investigación o tesis? .....	145
¿El tamaño de la muestra puede ser igual al tamaño de la población?.	146
¿Qué sucede cuando en una muestra se quieren analizar características cuantitativas y cualitativas? .....	147
¿Qué hacer si no se puede completar el tamaño de la muestra? .....	148
¿Qué dificultades se deben considerar en experimentos con seres vivos? .....	149
¿Se deben considerar muestras de reserva?.....	150
¿Cómo se realiza el muestreo en alimentos?.....	152
¿Cómo considerar el tamaño del efecto en el cálculo del tamaño de la muestra? .....	155
Capítulo 7: Ejemplos adicionales .....	161
Referencias.....	170

## Capítulo 1: Población y muestra

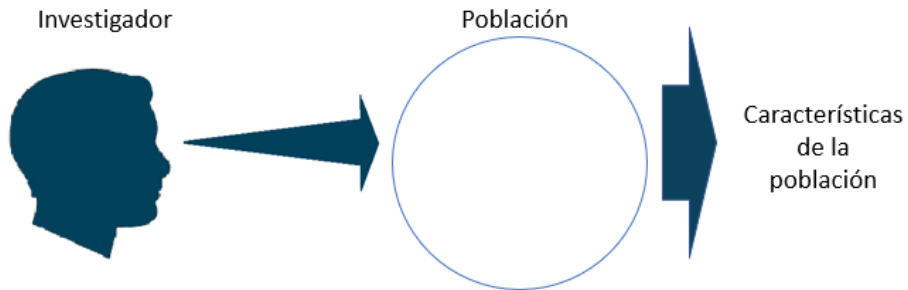
### 1.1. Población

En investigación se trata de conocer ciertas características de la realidad y la relación que existe entre ellas, estas características se denominan variables y se estudian en un grupo específico de elementos. Entonces, una vez establecidas las variables que se desean analizar, la población se define como el conjunto de todos los elementos en los cuales se analizarán dichas variables.



**Figura 1.1** En una población se pueden analizar diferentes variables.

Los resultados y conclusiones que se obtienen en una investigación se hacen con respecto a la población. Sin embargo, no siempre se analizan todos los elementos de la población, sino que se generalizan los resultados a partir de una cierta cantidad de observaciones.

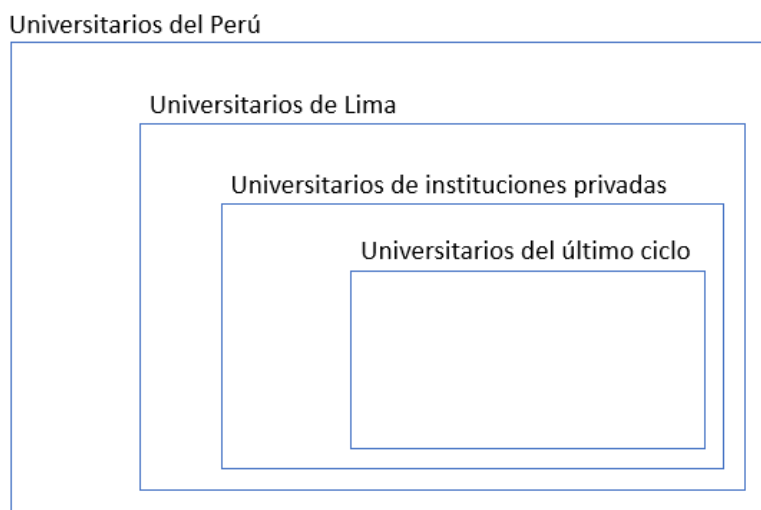


**Figura 1.2** El interés del investigador se encuentra en las características de la población.

## 1.2. Delimitación de la población

Vargas (2007) señala que la población se define de acuerdo con los objetivos de la investigación. Es decir, primero se debe determinar qué es lo que se quiere analizar y luego se establece en que grupo se va a realizar dicho análisis.

Hernández et al. (2014) resaltan que muchos trabajos de investigación presentan como deficiencia el no describir adecuadamente las características de la población donde se analizarán las variables. Esta deficiencia implica el riesgo de que la investigación se realice sobre una población heterogénea compuesta de elementos muy diferenciados. Finalmente, la falta de una adecuada delimitación de la población conlleva a establecer conclusiones incorrectas sobre la población en estudio.



**Figura 1.3** Delimitación de una población a diferentes niveles.

Con la figura 1.3 se pretende ejemplificar la delimitación de una población desde el caso más general hasta casos más particulares. Realizar una investigación sobre la población de estudiantes universitarios del Perú requiere contar con muchos recursos, por ello, se puede optar por realizar la investigación en los estudiantes universitarios de la ciudad de Lima, en este caso nuestras conclusiones limitan su alcance a esta población. También existe la opción de realizar la investigación en los universitarios limeños que estudian en universidades privadas, ahora, las conclusiones de la investigación reducen, aún más, su alcance a esta nueva población. Otra alternativa consiste en realizar la investigación en los estudiantes limeños de universidades privadas que cursan el último ciclo de estudio, así las conclusiones tienen un alcance mucho menor en comparación con las que se obtienen con las poblaciones anteriores.

Los elementos que componen las poblaciones son de diferente naturaleza. Es decir, pueden ser individuos, instituciones, artículos, etc. Por ello, es importante definir adecuadamente la población que será la base para el desarrollo de la investigación. La población puede estar compuesta de los siguientes elementos:

## Personas

- Estudiantes universitarios de la ciudad de Lima.
- Padres de familia de una escuela “X”.
- Abogados de la ciudad de Cusco.
- Personal médico del hospital arzobispo Loayza.
- Chefs de una cadena de restaurantes.

## Instituciones

- Universidades del Perú.
- Colegios profesionales del Perú.
- Colegios particulares del distrito de Breña.
- Micro y pequeñas empresas de la ciudad de Piura.
- Hospitales del sur del país.

## Plantas

- Árboles de las zonas urbanas de Arequipa.
- Frutas que se expenden en los mercados de la capital.
- Mazorcas de maíz de una cosecha.
- Tomates de una plantación.
- Vegetales que se expenden en los mercados de una ciudad.

## Animales

- Conejos de una granja.
- Peces capturados por una empresa pesquera.
- Larvas de insectos en las parcelas de cultivo.
- Pollos enfermos en una empresa avícola.
- Mascotas en las zonas residenciales de una ciudad.

## Objetos

- Panes elaborados con harina de maíz.

- Pantalones fabricados en el último mes.
- Viviendas del distrito de Magdalena.
- Máquinas de una fábrica.
- Televisores vendidos en diciembre del 2020.

Lohr (1999) diferencia entre población objetivo y población muestreada. El primer concepto hace referencia al conjunto total de elementos que se desea analizar y el segundo se refiere al conjunto de elementos de donde se puede extraer la muestra de manera efectiva. Por ejemplo, cuando se quiere realizar una encuesta online, a los padres de familia de la ciudad de La Paz, por medio de la red social Facebook; la población objetivo se compone de todos los padres de familia de la ciudad de La Paz. Sin embargo, no todos los padres de familia de la ciudad de La Paz tienen acceso a Facebook, por ello, la población muestreada se compone de todos los padres de familia de la ciudad de La Paz que tienen acceso a Facebook y están dispuestos a responder a la encuesta.

### 1.3. Muestra

En muchas investigaciones no es posible examinar a todos los elementos de la población en estudio. Esto se debe a limitaciones en el control de diferentes factores como el tiempo y otros recursos. Frente a esta dificultad se opta por realizar el estudio solo con una parte de la población, es decir, se utiliza una muestra.

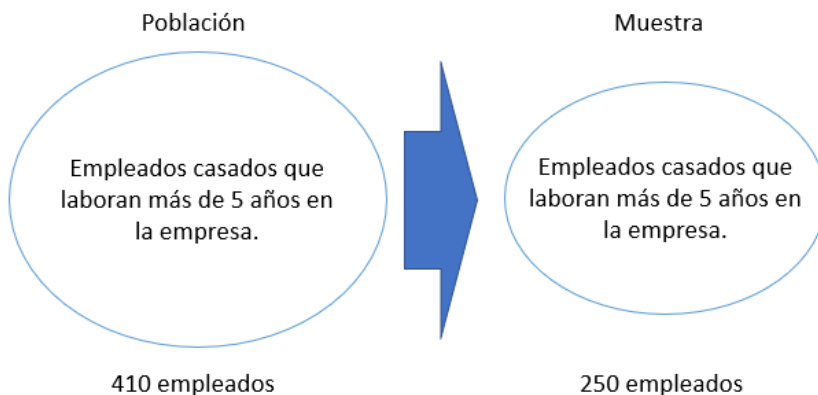
Cochran (1980) sostiene que existen ventajas al trabajar con muestras tales como el costo reducido, la mayor rapidez de recolección y resumen de información, la mayor posibilidad de obtener información y la mayor exactitud que se puede lograr.

Existe una condición importante para el uso de muestras: la muestra debe ser representativa. La muestra tiene que ser un reflejo de la población, es decir, debe poseer las mismas características, o en todo caso ser muy similar a la población. Esta condición asegura que las



estimaciones que se obtienen a partir de la muestra sean significativas también para la población. Si una muestra no representa adecuadamente a la población, los resultados que se obtienen del análisis solo le corresponden a la muestra y no se pueden generalizar para la población.

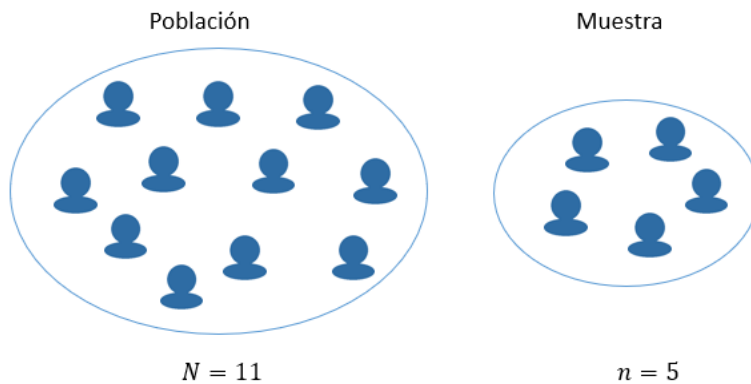
Leyton (2018) considera que toda muestra debe cumplir con dos requisitos para ser considerada como representativa: En primer lugar, la representatividad cuantitativa referente a la cantidad de elementos de la muestra, con respecto al tamaño de la población y estratos, que son necesarios para establecer las inferencias. En segundo lugar, está la representatividad cualitativa que considera que los elementos de la muestra deben contar con las características generales y específicas de la población en estudio.



**Figura 1.4** La muestra debe ser representativa en cantidad y calidad.

#### 1.4. Notación para el tamaño de la población y de la muestra

Para hacer más sencilla la explicación de los temas posteriores, en referencia a los tamaños de la población y de la muestra, consideraremos que a la cantidad de elementos que conforman la población se le llamará tamaño de la población y su símbolo será la letra  $N$  y a la cantidad de elementos que conforman la muestra se le llamará tamaño de la muestra y se le simbolizará con la letra  $n$ .



**Figura 1.5** El tamaño de la muestra  $n$  se puede obtener a partir de cálculos estadísticos usando  $N$ , si es que se conoce dicho valor.

### 1.5. Unidad de análisis

Son los elementos que componen la muestra y por tanto también a la población de interés. Lohr (1999) denomina unidad de análisis o unidad de observación al objeto sobre el cual se realiza una medición.

**Ejemplo 1:** En una investigación donde se quiere conocer la relación entre el consumo de alimentos ultra procesados y la actividad física en los estudiantes de la Universidad Nacional de Ingeniería en el año 2021. La unidad de análisis está representada por un estudiante activo de la Universidad Nacional de Ingeniería en el año 2021.

**Ejemplo 2:** En un estudio donde se quiere conocer el efecto que tiene el consumo de cierta proteína sobre la masa corporal de las ratas de tres meses de vida. La unidad de análisis está representada por una rata de tres meses de vida.

En investigación experimental los elementos a partir de los cuales se obtienen los datos se denominan unidades experimentales. Esta unidad experimental también se puede describir como el elemento utilizado para generar un valor que sea representativo del resultado del experimento (Gutiérrez y de la Vara, 2008).

## 1.6. Marco muestral

Es una lista que contiene el registro de todos los elementos de la población (Anderson et al., 2012).

Por ejemplo:

- El registro de todos los usuarios de una empresa de telefonía fija.
- La lista de todos los estudiantes matriculados en el año 2020 en la universidad “X”.
- El registro de todas las empresas de servicios en la ciudad de Lima.
- La lista de todos los deportistas participantes de los juegos panamericanos Lima 2019.
- La planilla de todos los trabajadores de una empresa.

Entonces, dada una población, el marco muestral facilita el acceso a los elementos que la componen. Sin embargo, En muchas ocasiones este registro no se conoce. En el capítulo 4 se establece la importancia de contar con el marco muestral para seleccionar el proceso de muestreo adecuado.

## Capítulo 2: Diseños de investigación

En este capítulo se realiza una introducción a las definiciones de los diseños de investigación que son de uso frecuente en diferentes estudios. Es necesario conocer los tipos de diseños porque a partir de estos se establece la expresión matemática para el cálculo del tamaño muestral, es decir, se define con cierto grado de probabilidad la cantidad mínima de elementos que componen la muestra. Las definiciones que se presentan son básicas, pero cumplen con la función de orientar al investigador en la selección del diseño de acuerdo con los objetivos de su investigación.

Lwanga y Lemeshov (1991) indican que no es posible establecer el tamaño de la muestra si previamente no se han especificado las características del estudio y sus objetivos. Entonces, la elección del procedimiento adecuado para la selección de una muestra también depende del diseño de la investigación. Hernández et al. (2010) indican que el diseño es la planificación elaborada para obtener la información necesaria para el estudio. Además, clasifican a los diseños en dos grandes grupos: diseños experimentales y diseños no experimentales.

Pintado (2014) señala que además de establecer adecuadamente los objetivos de la investigación se debe tener en cuenta el diseño. Es decir, se debe establecer si se analizará una sola población, si se analizarán diferentes poblaciones, si se realizará una sola medición de la variable o será una recolección de información en diferentes momentos.

En la literatura sobre metodología de la investigación científica se han establecido diferentes formas de clasificar los tipos de investigación. En este texto se considera como diseño a la planificación que se elabora

previamente a la etapa de recolección de datos y que responde a los objetivos de la investigación. Así, una investigación puede ser descriptiva, correlacional, experimental, longitudinal, etc. Por el momento se están mencionando algunos tipos de diseño, en el desarrollo de este capítulo se presentará una explicación más clara de los diseños de investigación de uso más frecuente.

Respecto a la clasificación de los tipos de investigación Tamayo (2003) resalta que casi nunca se presentan estudios con un tipo de investigación único; generalmente se combinan entre sí y se adecúan a los requerimientos de la investigación, esto mismo sucede con los diseños. Por ejemplo, existen investigaciones descriptivas correlacionales, investigaciones descriptivas transversales, investigaciones descriptivas longitudinales, etc.

De lo expuesto, se entiende que el proceso de definir el diseño de investigación no es algo complementario, sino que es de suma importancia para la elección apropiada del proceso para el cálculo del tamaño de la muestra. Más adelante se pondrá en evidencia que existen diferencias notables entre las fórmulas para calcular el tamaño muestral según el diseño de investigación.

## 2.1. Diseños no experimentales u observacionales

La característica principal de los diseños observacionales es que en todo el proceso de investigación no se manipulan variables. Es decir, se analizan las variables tal y como estas se manifiestan en la realidad. Manterola y Otzen (2014) señalan que la clasificación de los diseños que se usa a menudo, no hace referencia a tipos de diseños propiamente dichos, sino que se deben considerar como características de los diseños.

Por ello, a continuación, se presentan como características de las investigaciones y se muestra una breve descripción de ellas:

**Tabla 2.1** Características de los diseños según Manterola y Otzen

Característica	Descripción
Transversal	Una sola medición de la variable.
Longitudinal	Mas de una medición de la variable.
Retrospectivo	Hechos pasados.
Prospectivo	Estudio de los hechos hacia el futuro.

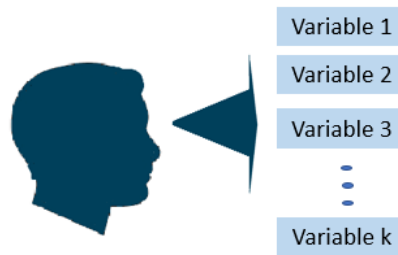
Hernández et al. (2014) clasifican los diseños no experimentales en dos grandes grupos: trasversal y longitudinal o transeccional y evolutivo respectivamente.

**Tabla 2.2** Diseños no experimentales según Hernández, Fernández y Baptista

Diseño	Tipo
Transversal	Exploratorio
	Descriptivo
	Correlacional-causal
Longitudinal	Tendencia
	Cohortes
	Panel

### 2.1.1. Diseño trasversal

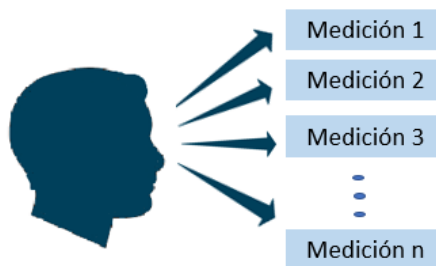
En un diseño trasversal, la medición de la variable se realiza en un solo momento. Esto no quiere decir que se analiza a todos los elementos de la muestra en un mismo momento, sino que se refiere a que solo se realiza una sola medición de la variable. Por ejemplo, para investigar sobre los hábitos de estudio en estudiantes universitarios, se puede realizar una encuesta a diferentes estudiantes, no necesariamente a la misma hora, pero si en circunstancias similares. El caso señalado representa un diseño trasversal dado que la variable hábitos de estudio se mide una sola vez en los estudiantes.



**Figura 2.1** La variable o las variables se miden una sola vez.

### 2.1.2. Diseño longitudinal

En este tipo de diseño se trata de recolectar información en diferentes momentos para conocer los cambios y establecer conclusiones sobre ellos. Por ejemplo, una investigación cuyo objetivo principal es conocer la evolución del consumo de bebidas alcohólicas en los distritos del sur de Lima, representa un diseño longitudinal puesto que se debe observar la variable cada año para identificar los periodos con cambios más relevantes.

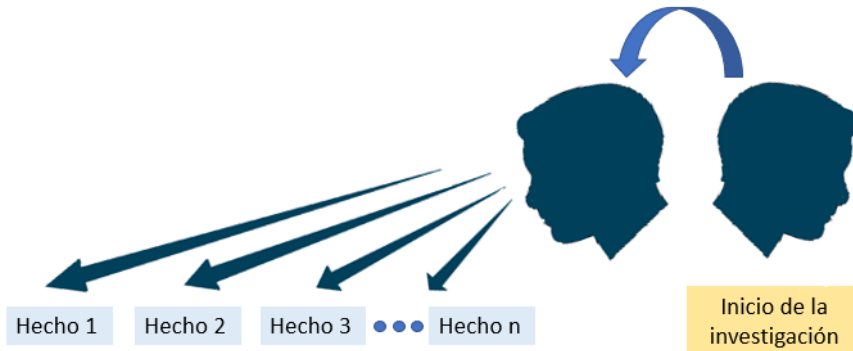


**Figura 2.2** La variable o las variables se miden en diferentes momentos.

### 2.1.3. Diseño retrospectivo

En este tipo de diseño el inicio de la investigación ocurre en un tiempo posterior a la realización de los hechos que se quiere analizar. Para recolectar la información se hace uso de archivos e informes elaborados sobre el hecho ya realizado. Por ejemplo, una investigación donde se

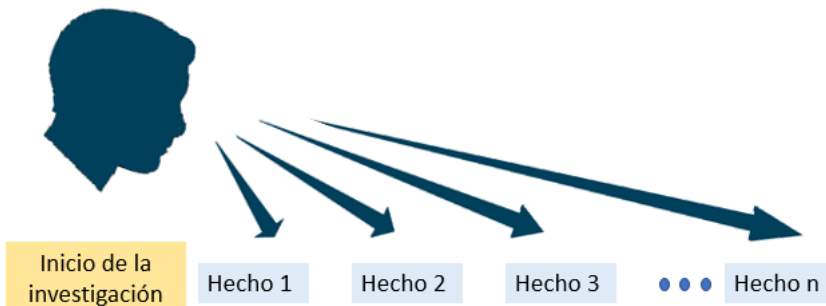
quiere conocer los distintos métodos de enseñanza usados por los docentes de educación básica antes de la pandemia por el COVID 19.



**Figura 2.3** Los hechos se presentan ordenados en el tiempo.

#### 2.1.4. Diseño prospectivo

En este tipo de diseño el inicio de la investigación ocurre en un tiempo anterior a la realización de los hechos que se quiere analizar. Por ejemplo, una investigación donde se quiere conocer los distintos métodos de enseñanza (híbridos) usados por los docentes de educación básica después de la pandemia por el COVID 19.



**Figura 2.4** Los hechos se presentan ordenados en el tiempo.

Manterola y Otzen (2014) también resaltan la clasificación de los estudios observacionales en dos grandes grupos:



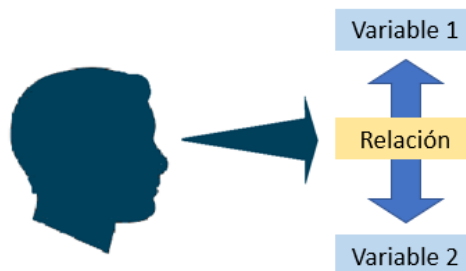
**Tabla 2.3** Diseños no experimentales según Manterola y Otzen.

Tipo	Descripción
Descriptivos	Descripción de variables en un grupo de sujetos.
Analíticos	Comparación de grupos de sujetos.

Dentro de los estudios observacionales analíticos encontramos a los correlacionales, los estudios de cohortes y los estudios de casos y controles.

### 2.1.5. Diseño correlacional

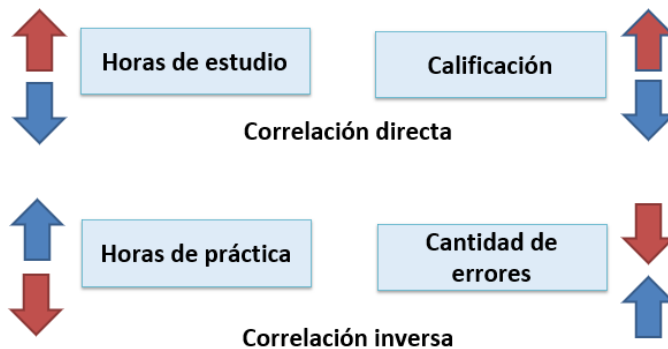
En este tipo de diseño se analiza la relación entre dos variables sin que esto implique establecer la noción de causalidad. Es decir, en este tipo de diseños no se considera variable independiente o dependiente porque no se establece la relación de causa y efecto, las variables deben considerarse como variable 1 y variable 2.



**Figura 2.4** El interés principal del investigador está en conocer la relación entre las variables.

En un análisis de correlación se pueden obtener dos resultados: las variables presentan una correlación significativa o las variables no presentan una correlación significativa. De obtenerse una correlación significativa entre las variables la correlación puede ser directa cuando a mayores valores de la variable 1 se observan mayores valores para la

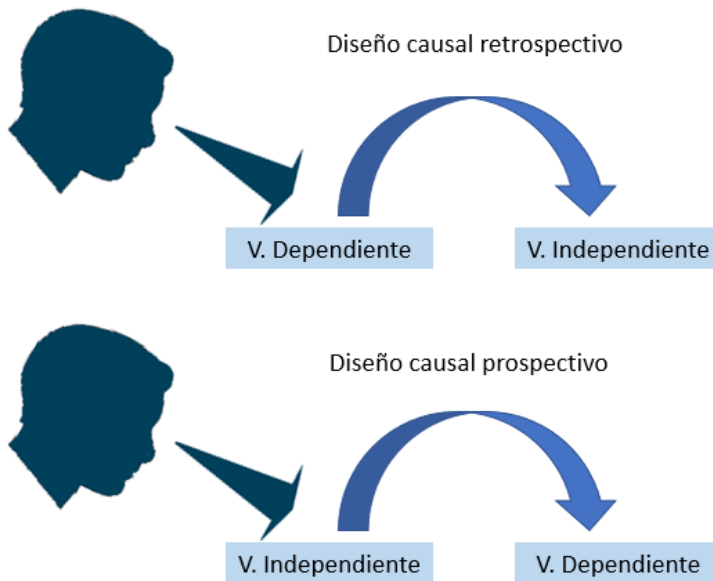
variable 2. Cuando a mayores valores en la variable 1 se asocian menores valores en la variable 2 la correlación es inversa. De no observarse ninguna de las opciones anteriores se dice que las variables presentan correlación nula o simplemente no están relacionadas.



**Figura 2.5** La correlación directa y la correlación inversa también son denominadas correlación positiva y negativa respectivamente.

### 2.1.6. Diseño correlacional-causal

En este tipo de diseños, además de establecer la correlación se trata de identificar una relación de causalidad. Es decir, se trata de distinguir variables independientes y dependientes en la población observada. Hernández et al. (2014) manifiestan que las relaciones de causalidad se utilizan en los diseños experimentales, pero en un diseño correlacional-causal las causas y los efectos ya han ocurrido, por tanto, no se efectúa la manipulación de variables.

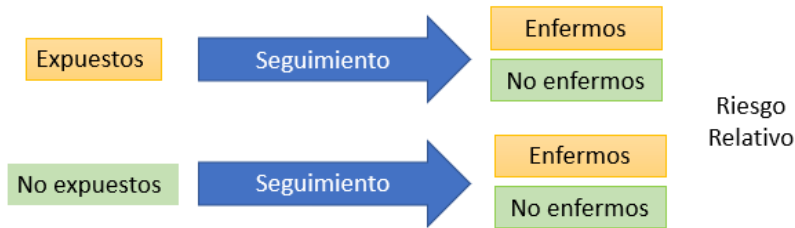


**Figura 2.6** La relación de causalidad se establece a partir de la variable dependiente o independiente según León y Montero (2003, citados por Hernández, et al., 2014).

### 2.1.7. Estudio de cohortes

Hernández et al. (2014) señalan que en un estudio de cohortes se examinan los cambios en grupos de individuos a través del tiempo. Manterola y Otzen (2014) precisan que una cohorte es un grupo de sujetos que se observa en diferentes momentos esperando la aparición de una variable resultado, y por otro lado un factor que pueda predecir la variable resultado. En este tipo de estudios se realiza una comparación entre grupos de sujetos, pero no se tiene control sobre las variables.

Un ejemplo bastante utilizado hace referencia a la observación de la exposición de un grupo de individuos a cierto contexto y el desarrollo de una enfermedad. De este modo, en un primer momento, la observación se divide en expuestos y no expuestos. Luego, dentro de los grupos mencionados se observa la cantidad de individuos que han desarrollado la enfermedad y lo que no han desarrollado la enfermedad.

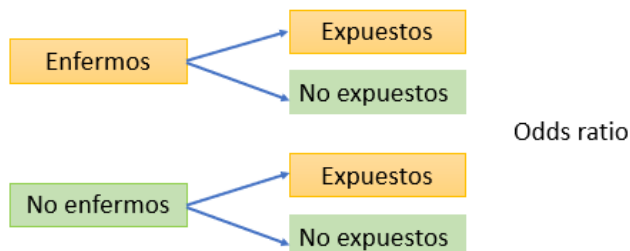


**Figura 2.7** Se analizan dos grupos: los expuestos y los no expuestos.

En la figura 2.7 el riesgo relativo hace referencia a la razón de dos razones: la cantidad de enfermos que han sido expuestos entre la cantidad total de expuestos y la cantidad de enfermos que no han sido expuestos entre la cantidad total de no expuestos. De este modo se define si la exposición es determinante para el desarrollo de la enfermedad.

### 2.1.8. Estudios de casos y controles

Los estudios de casos y controles son un tipo de diseño usado generalmente en investigaciones clínicas cuyo objetivo es reconocer factores de riesgo. En este tipo de estudio se toma una muestra de individuos que presentan cierta característica y se les denomina casos, se toma una muestra de individuos que no presentan la característica antes mencionada y se les denomina controles. Finalmente se trata de establecer si los individuos, en general, que han sido expuestos a un determinado factor aparecen con mayor frecuencia en el grupo de casos o controles para establecer la relación causa y efecto.



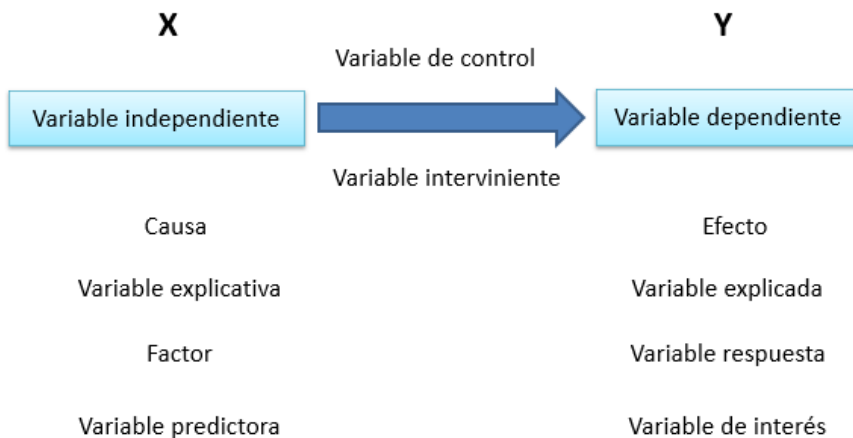
**Figura 2.8** Se analiza dos grupos de individuos; enfermos y no enfermos.

En la figura 2.8 el Odds Ratio hace referencia a la razón de dos razones: la cantidad de enfermos que han sido expuestos entre la cantidad de enfermos que no han sido expuestos y la cantidad de no enfermos que han sido expuestos entre la cantidad de no enfermos que no han sido expuestos. De este modo se define si la exposición es determinante para el desarrollo de la enfermedad.

## 2.2. Diseños experimentales

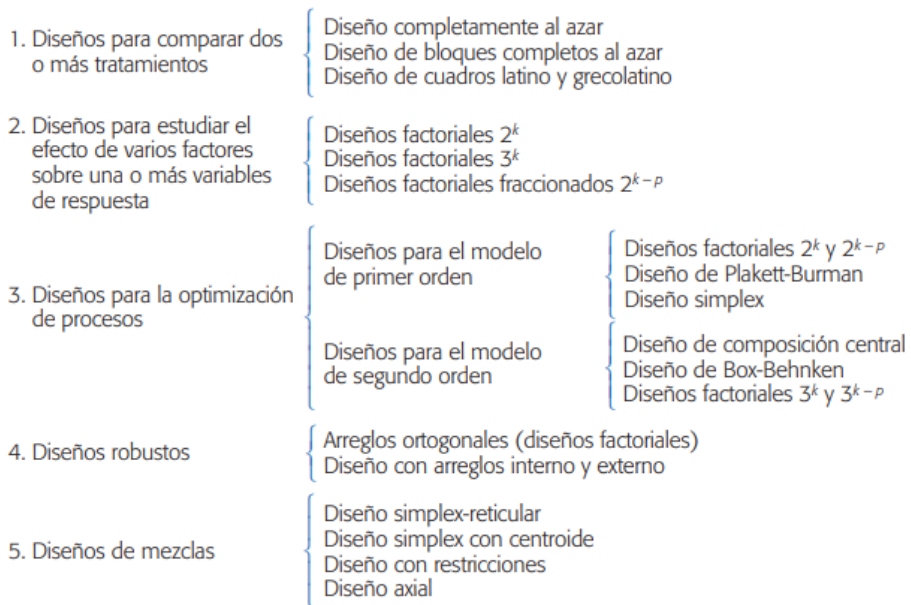
La característica principal de los diseños experimentales es que durante el proceso investigativo se manipulan variables para observar su efecto sobre otras variables.

Pintado (2014) indica que un diseño experimental es el planeamiento previo que se realiza con la finalidad de analizar los datos obtenidos de manera objetiva y llegar a conclusiones válidas sobre el problema en estudio. Durante este planteamiento se trata de responder a interrogantes como: ¿En quienes se llevará a cabo el experimento? ¿Dónde se realizará el experimento? ¿Bajo qué condiciones se realizará el experimento?, etc.



**Figura 2.9** Diferentes denominaciones para los elementos del diseño experimental.

La clasificación de los diseños experimentales es más compleja que la clasificación de los diseños no experimentales y maneja una terminología y simbología propia. Así, en un diseño experimental se menciona el grupo de control, grupo experimental, factores, niveles, tratamientos, réplicas y otros elementos que se abordarán más adelante para entender la metodología para el cálculo del tamaño de la muestra en este tipo de diseños.



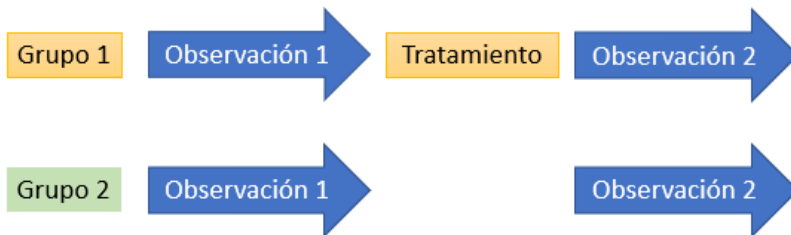
**Figura 2.10** Clasificación de los diseños experimentales.

Fuente: Gutiérrez, H. y de la Vara, R. (2008). *Análisis y diseño de experimentos*.

En el capítulo 5 se abordan solo algunos de estos diseños experimentales y se muestra el camino para obtener el tamaño muestral adecuado con el soporte de software estadístico.

En los diseños experimentales existen grupos de comparación y en este tipo de diseños es donde se establece, de forma rigurosa, la relación de

causalidad. A continuación, se muestra un diseño donde se comparan dos grupos, uno de control y uno experimental, a los que se les ha realizado dos observaciones en distintos momentos.



**Figura 2.11** El grupo de control no recibe ningún tratamiento.

A continuación, se bosqueja el proceso para la elección del diseño adecuado en base a preguntas básicas que se requieren para la etapa de la recolección de datos.

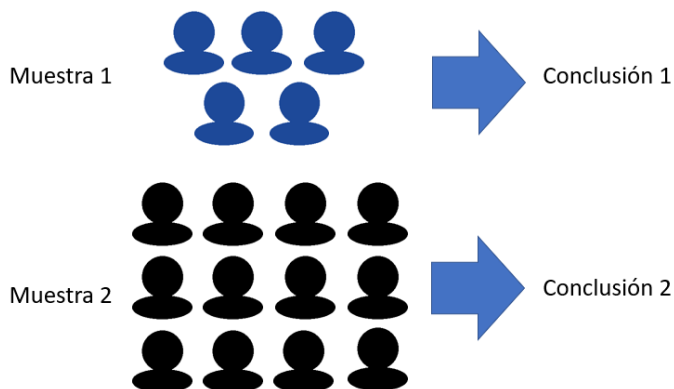


**Figura 2.12** Preguntas que el investigador puede hacerse para establecer el diseño a utilizar.

## Capítulo 3: Elementos básicos de estadística inferencial

En este capítulo se trata de establecer algunas nociones básicas de inferencia estadística que más adelante ayudarán a entender mejor la relación entre la muestra y la población en términos de probabilidad. Con la finalidad de no distraer la atención del objetivo principal de este texto se hará el uso necesario de la simbología estadística.

Para establecer el sustento matemático que permite el cálculo del tamaño de la muestra, es necesario comprender el lenguaje de la estadística inferencial puesto que todas las expresiones matemáticas que utilizaremos en el capítulo 5 para el cálculo de dicho tamaño muestral contienen términos específicos de la teoría del muestreo. La inferencia estadística es importante porque permite generalizar los resultados a partir de una muestra.



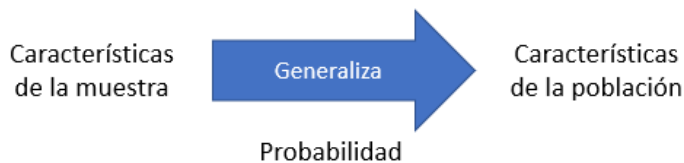
**Figura 3.1** ¿Qué tamaño de muestra nos brindará mejores conclusiones?



En la figura 3.1 se trata de representar la relación entre el tamaño de muestra y la calidad de las conclusiones que se pueden obtener a partir de esta. En los siguientes capítulos se desarrollará esta relación con mayor detalle. Se puede pensar que un mayor tamaño de muestra puede brindar mejores conclusiones. Esto último es totalmente aceptable si se verifica que ambas muestras presentan la misma variabilidad, en caso contrario no se puede dar una respuesta categórica.

### 3.1. La inferencia estadística

Una inferencia estadística consiste en hacer conclusiones acerca de las características de una población, pero sin la necesidad de estudiar la característica en cada uno de los elementos de dicha población, sino que podemos basarnos en una muestra. La inferencia estadística también se denomina estadística inferencial.



**Figura 3.2** A partir de una muestra se pueden estimar las características de la población.

La generalización de los resultados o características de la muestra hacia la población está sujeta a cierto grado de probabilidad. Esta probabilidad determina la precisión de las conclusiones que se generalizan y depende de factores como el tamaño y la variabilidad de la muestra.

### 3.2. Estadígrafo

Lind et al. (2016) indican que cualquier medición basada en una muestra de datos recibe el nombre de estadístico o estadígrafo. Podemos agregar que un estadígrafo es un valor que resume alguna característica

de la muestra. Como ejemplos de estadígrafos podemos mencionar a la media, la mediana, la moda, la desviación estándar, etc.

### 3.3. Parámetro

Blair y Taylor (2008) definen los parámetros como aquellos valores que resumen las características de una población. Los parámetros son valores estimados y su valor se determina a partir del valor de los estadígrafos.

**Tabla 3.1** Símbolos de algunos estadígrafos y parámetros

	Estadígrafo	Parámetro
Media	$\bar{X}$	$\mu$
Proporción	$p$	$\pi$
Desviación estándar	$S$	$\sigma$
Varianza	$S^2$	$\sigma^2$

### 3.4. Clasificación de la estadística inferencial

Nolberto y Ponce (2008) señalan que la inferencia estadística se clasifica en dos ramas: Estimación de Parámetros y Prueba de Hipótesis.

#### 3.4.1. Estimación de parámetros

Moya y Saravia (2004) indican que la estimación estadística es un proceso que implica la aproximación al valor del parámetro a partir de la información de la muestra. Existen dos formas de realizar la estimación: la estimación puntual y la estimación por intervalos de confianza. La estimación por intervalos de confianza es el proceso más utilizado debido a que se define el margen de error de la estimación.

**Tabla 3.2** Tipos de estimación del parámetro media o promedio.

Estadígrafo	Estimación	Parámetro
$\bar{X} = 48$	Puntual	$\mu = 48$
$\bar{X} = 48$	Intervalo de confianza	$\mu \in [45; 51]$

Observe que un estadígrafo es un estimador puntual del parámetro, sin embargo, no se indica que tan precisa es la estimación. En cambio, en la estimación por intervalo de confianza se establece la probabilidad de que el intervalo contenga al parámetro. Esto último se explicará mejor en el siguiente apartado.

A continuación, se muestra la expresión matemática para determinar el intervalo de confianza para la media poblacional a partir de los datos recolectados de una muestra.

$$\mu \in \left[ \bar{X} - \frac{Z_{\alpha/2} \sigma}{\sqrt{n}}; \bar{X} + \frac{Z_{\alpha/2} \sigma}{\sqrt{n}} \right]$$

Donde:

$\mu$ : media poblacional

$\bar{X}$ : media muestral

$\sigma$ : desviación estándar poblacional

$Z_{\alpha/2}$ : valor asociado al nivel de confianza

$n$ : tamaño de la muestra

La siguiente es la expresión matemática para determinar el intervalo de confianza para la proporción poblacional a partir de los datos recolectados de una muestra.

$$\pi \in \left[ p - Z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}}; p + Z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}} \right]$$

Donde:

$\pi$ : proporción poblacional

$p$ : proporción muestral

$Z_{\alpha/2}$ : valor asociado al nivel de confianza

$n$ : tamaño de la muestra

Las expresiones matemáticas referentes a los intervalos de confianza para la media y la proporción se muestran con la finalidad de explicar que existen métodos estadísticos ya elaborados para determinar los intervalos de confianza para diferentes parámetros, y para evidenciar el uso de tamaño de la muestra  $n$  dentro de estas fórmulas. El significado de cada uno de los elementos de estas expresiones matemáticas se abordará con detalles y ejemplos en el capítulo 5.

### 3.4.2. Prueba de hipótesis

Es una regla que permite aceptar o rechazar una hipótesis planteada con base en la observación de los resultados y el análisis estadístico. La prueba de hipótesis es un paso importante en el desarrollo de las investigaciones de enfoque cuantitativo.

**Hipótesis estadística:** Es una aseveración que se hace acerca de una o más características de una población. Moya y Saravia (2004) plantean que la hipótesis estadística es una afirmación que se hace acerca de la distribución de una o más variables aleatorias. También se puede decir que son aseveraciones con respecto a un parámetro o parámetros.

- El 1.07% de los productos elaborados son defectuosos.
- La edad promedio de los trabajadores de la empresa es 38 años.
- El IMC promedio del grupo “A” es mayor al IMC promedio del grupo “B”.
- El nivel de satisfacción en los clientes es mayor luego de la capacitación del personal.

**Hipótesis nula ( $H_0$ ):** Es la hipótesis que se quiere probar.

**Hipótesis alternativa ( $H_1$ ):** Es la hipótesis que se acepta en caso de que la  $H_0$  sea rechazada.

**Error tipo I ( $\alpha$ ):** También se le denomina significancia. Es la probabilidad de rechazar  $H_0$  cuando  $H_0$  es verdadera.

**Error tipo II ( $\beta$ ):** Probabilidad de aceptar  $H_0$  cuando  $H_0$  es falsa.

**Nivel de confianza ( $1-\alpha$ ):** Probabilidad de que el parámetro a estimar se encuentre en el intervalo de confianza.

**Poder estadístico ( $1-\beta$ ):** Probabilidad de que la hipótesis alternativa sea aceptada cuando la hipótesis alternativa es verdadera.

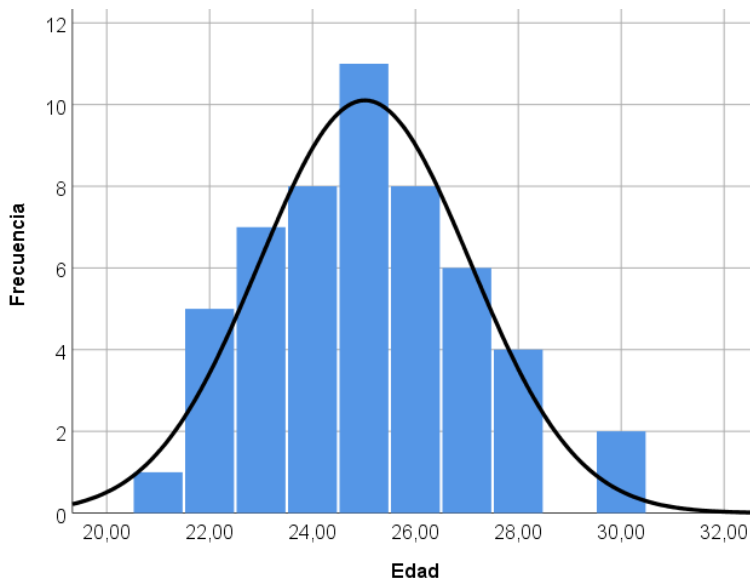
Martínez et al. (2006) señalan que el uso de los valores 5% y 20% para  $\alpha$  y  $\beta$  respectivamente, se ha hecho costumbre en investigación y que no está mal regirse por lo menos por estos requisitos puesto que se pueden utilizar valores aún más pequeños. Hay que tener en cuenta que usar valores de  $\alpha$  y  $\beta$  más pequeños implica recurrir a un tamaño de muestra más grande.

**Tamaño del efecto:** Pardo y San Martín (2010) definen al tamaño del efecto como la magnitud del efecto estudiado. Por ejemplo, si se realiza una comparación, el tamaño del efecto indica la magnitud de la diferencia. Si se relacionan variables, el tamaño del efecto indica la intensidad de la relación.

### 3.5. La distribución normal

Para los fines de este texto basta con indicar que la distribución normal es una distribución teórica a la cual se ajustan muchas distribuciones de variables del contexto cotidiano. En muchos de los ejemplos que se desarrollan en los capítulos posteriores se asume que los datos recolectados provienen de una población con distribución aproximadamente normal.

Para entender mejor esta distribución normal recurriremos a las características de su gráfica. Por ejemplo, si se realiza una reunión de estudiantes universitarios con diferentes edades en el intervalo de 21 a 30 años, es más probable encontrar estudiantes con edades intermedias que estudiantes con edades cercanas a los extremos del intervalo de edades.



**Figura 3.3** La curva representa la distribución normal.

En la figura 3.3 se puede observar que la distribución de las edades se aproxima mucho a la distribución normal. Observe que las mayores frecuencias se encuentran en la parte central de la distribución, en cambio las menores frecuencias se encuentran en los extremos.

No solo la distribución de frecuencias de las variables puede aproximarse a una distribución normal, la distribución de probabilidad de estadígrafos como la media también pueden presentar una distribución aproximadamente normal.

Para entender la distribución de probabilidad de la media podemos utilizar la población compuesta por las edades de cinco personas:

10; 12; 12; 14 y 16

Luego extraemos muestras de tamaño 2:

$$M_1 = \{10; 12\}$$

$$M_2 = \{10; 12\}$$

$$M_3 = \{10; 14\}$$

$$M_4 = \{10; 16\}$$

$$M_5 = \{12; 12\}$$

$$M_6 = \{12; 14\}$$

$$M_7 = \{12; 16\}$$

$$M_8 = \{12; 14\}$$

$$M_9 = \{12; 16\}$$

$$M_{10} = \{14; 16\}$$

Después calculamos los promedios para cada grupo:

$$\bar{X}_1 = \frac{10 + 12}{2} = 11$$

$$\bar{X}_2 = \frac{10 + 12}{2} = 11$$

$$\bar{X}_3 = \frac{10 + 14}{2} = 12$$

$$\bar{X}_4 = \frac{10 + 16}{2} = 13$$

$$\bar{X}_5 = \frac{12 + 12}{2} = 12$$

$$\bar{X}_6 = \frac{12 + 14}{2} = 13$$

$$\bar{X}_7 = \frac{12 + 16}{2} = 14$$

$$\bar{X}_8 = \frac{12 + 14}{2} = 13$$

$$\bar{X}_9 = \frac{12 + 16}{2} = 14$$

$$\bar{X}_{10} = \frac{14 + 16}{2} = 15$$

Con los promedios calculados podemos establecer la función de probabilidad para la variable aleatoria X: promedio de las edades de las personas elegidas.

$$P[X = 11] = \frac{2}{10} = 0.2$$

$$P[X = 12] = \frac{2}{10} = 0.2$$

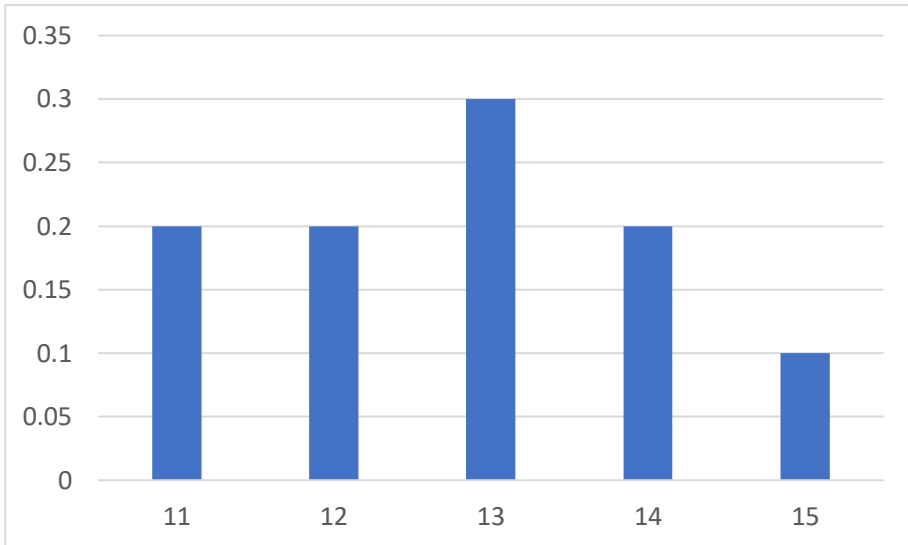
$$P[X = 13] = \frac{3}{10} = 0.3$$

$$P[X = 14] = \frac{2}{10} = 0.2$$

$$P[X = 15] = \frac{1}{10} = 0.1$$

Usando el programa Excel podemos elaborar la gráfica de la distribución de probabilidad.





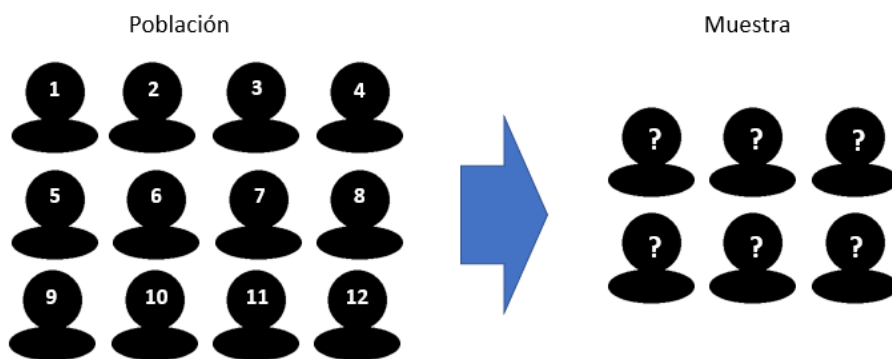
**Figura 3.4** Distribución de probabilidad de la media.

En la figura 3.4 se observa que la distribución de probabilidad para la media en este ejemplo no tiene una buena aproximación a la distribución normal. Sin embargo, el ejemplo es útil para comprender definiciones posteriores. Cabe resaltar que la distribución de probabilidad de la media se puede ajustar más a la distribución normal cuando aumenta el tamaño de la muestra.

La distribución normal no es la única distribución teórica que se utiliza en los trabajos de investigación. Además de la distribución normal existen otras distribuciones de probabilidad como la distribución t de Student, la distribución ji-cuadrada y la distribución F.

## Capítulo 4: Selección de los elementos para la muestra

En este apartado se establecen algunos criterios que se usan con frecuencia para elegir los elementos de la muestra. Es decir, se trata de responder a la interrogante que surge dada una población de tamaño  $N$  y una muestra calculada de tamaño  $n$  ¿Qué elementos de la población deben ser elegidos para ser parte de la muestra?



¿Cómo elegir los 6 elementos necesarios para la muestra?

**Figura 4.1** Elección de los elementos muestrales.

### 4.1. Muestreo

En el desarrollo de este texto se considerará al muestreo como el estudio de la relación que existe entre una población y las muestras que se pueden obtener de esta población. El muestreo es importante porque permite realizar inferencias partiendo de una muestra hacia la población

por medio de una prueba de hipótesis o estimación de parámetros (Spiegel y Stephens, 2009).

De acuerdo con Lind, et al. (2016) podemos considerar una definición más general indicando que el muestreo es un instrumento para establecer inferencias sobre la población. Sin embargo, esta última definición solo se brinda para entender la complejidad del muestreo, pues en esta parte del texto nos centraremos en los criterios de uso común para la selección de los elementos de la muestra.

Cochran (1980) resalta que el muestreo es una herramienta que se utiliza en diferentes entidades como en la administración, en la investigación de mercados, en las empresas, en la ingeniería, en las ciencias de la salud, la industria y en las ciencias de la educación. También, señala que la teoría del muestreo aumenta la eficiencia del proceso mismo de muestreo.

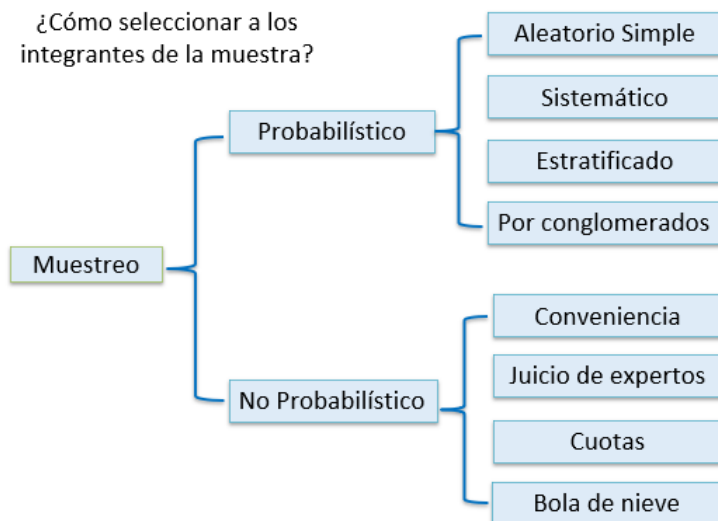
Por ello, el procedimiento para la selección de la muestra es una parte importante de la investigación que aumenta la probabilidad de obtener conclusiones válidas para la población. El proceso de muestreo inicia luego de plantear los objetivos, de realizar la elección del diseño de investigación y los métodos estadísticos para el análisis inferencial. Posteriormente se elige el método de muestreo más pertinente según las características de la población, de este modo se puede elegir una muestra representativa a partir de la población bien definida.

## 4.2. Tipos de muestreo

Existen diferentes criterios de clasificación para los tipos de muestreo, aunque en general pueden dividirse en dos grandes grupos: métodos de muestreo probabilísticos y métodos de muestreo no probabilísticos.

Webster (2000) señala que es adecuado utilizar métodos de muestreo que permitan minimizar el error y entre estos métodos aborda: el muestreo aleatorio simple, el muestreo sistemático, el muestreo

estratificado y el muestreo por conglomerados. Sin embargo, existen otros tipos de muestreo como el muestreo por conveniencia, por cuotas, por juicio de expertos y el muestreo de bola de nieve.



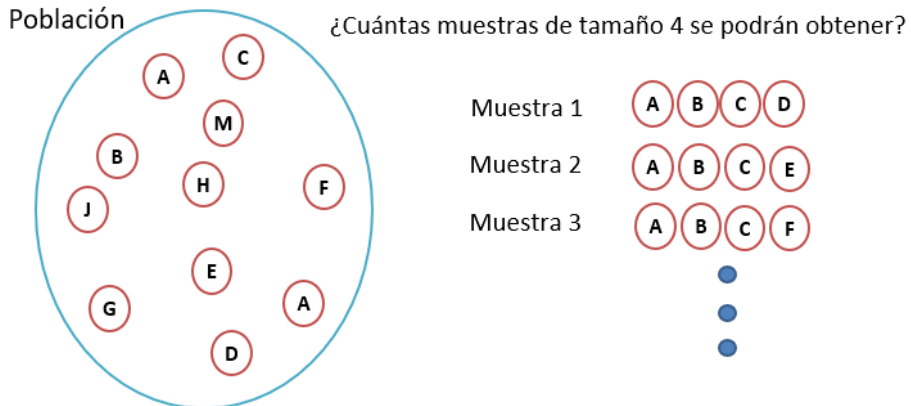
**Figura 4.2** Diferentes medios para seleccionar los elementos de la muestra.

Como señala Supo (2014) la elección de un tipo de muestreo probabilístico o uno no probabilístico depende de las circunstancias en las que se desarrolla la investigación. Por ello, en los siguientes apartados se describen las características de estos tipos de muestreo de modo que el investigador pueda reconocer el tipo de muestreo más adecuado a los objetivos de su trabajo y a sus recursos disponibles.

#### 4.2.1. Muestreo probabilístico

Se basa en que todos los individuos de la población tienen la misma probabilidad de ser elegidos para formar parte de una muestra de tamaño  $n$  y, consiguientemente, todas las posibles muestras de tamaño  $n$  tienen la misma probabilidad de ser seleccionadas (Anderson et al., 2012).

Navarro (2011) define como muestreo irrestricto aleatorio a aquel tipo de muestreo donde cualquier muestra de tamaño  $n$  tomada de una población de tamaño  $N$  tiene la misma probabilidad de ser seleccionada.



**Figura 4.3** Diferentes muestras que se pueden obtener de una misma población.

#### 4.2.1.1. Muestreo aleatorio simple

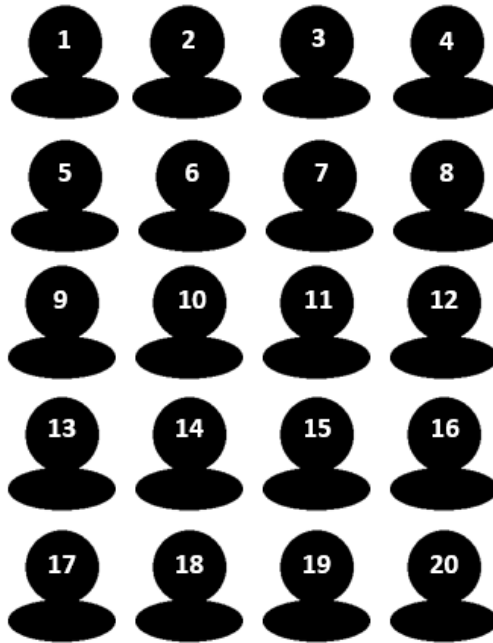
Este tipo de muestreo se elabora con base en el azar. Es decir, cualquier elemento de la población tiene posibilidad de ser incluido en la muestra (Lind et al., 2016).

El procedimiento empleado es el siguiente:

- Se asigna un número a cada individuo de la población. En este tipo de muestreo se debe tener acceso al marco muestral.
- Se elige algún medio mecánico como un sorteo, bolas dentro de una urna, tablas de números aleatorios, números aleatorios generados con algún programa, etc.
- Se extrae la cantidad de elemento necesarios para completar el tamaño de muestra requerido.

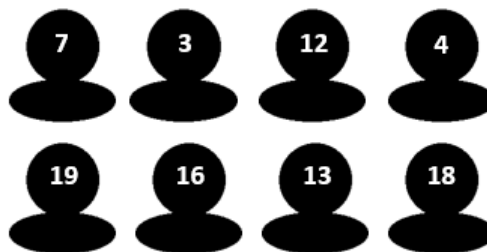
**Ejemplo 1:** Si tenemos una población constituida de 20 individuos y queremos extraer una muestra de 8 individuos, debemos seguir el siguiente procedimiento.

Como primer paso, enumeramos a todos los elementos de la población:



**Figura 4.4** Todos los elementos de la población se enumeran.

Como segundo paso, podemos realizar un sorteo luego del cual se obtienen los siguientes elementos para la muestra:



**Figura 4.5** Elementos de la muestra final.

Otra forma de seleccionar los elementos de forma aleatoria es utilizando Excel. Para ello, se debe escribir: =ALEATORIO.ENTRE(1; 20) en la casilla de fórmulas.

	C	D	E	F
	<b>Población</b>	<b>Muestra</b>		
	1	2		
	2	11		
	3	5	Elementos de la muestra	
	4	13		
	5	15		
	6	3		
	7	4		
	8	7		
	9			
	10			
	11			
	12			
	13			
	14			
	15			
	16			
	17			
	18			
	19			
	20			

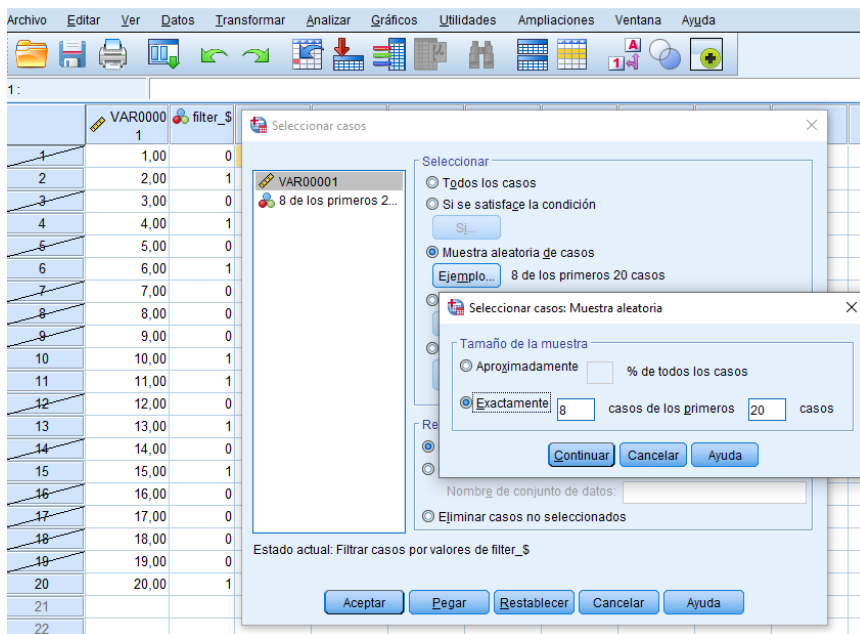
**Figura 4.6** Función para el muestreo aleatorio en Excel.

De acuerdo con los resultados del programa EXCEL, los elementos seleccionados para la muestra son:



**Figura 4.7** Elementos de la muestra final usando Excel.

Si utilizamos el SPSS se debe seguir la siguiente ruta: DATOS-SELECCIONAR CASOS-MUESTRA ALEATORIA DE CASOS-EJEMPLO. Este procedimiento nos genera los elementos para la muestra.



**Figura 4.8** Muestreo aleatorio en SPSS.

De acuerdo con este procedimiento aleatorio los elementos seleccionados para la muestra son aquellos que no han sido tachados por el programa:





**Figura 4.9** Elementos de la muestra final usando SPSS.

**Ejemplo 2:** Dada una población constituida de 45 elementos se desea extraer una muestra de 14 de ellos. Realiza la selección de los elementos de la muestra usando el muestreo aleatorio simple.

Como primer paso, enumeramos a todos los elementos de la población. Luego se escribe: `=ALEATORIO.ENTRE(1; 45)` en la primera casilla de la columna donde se ubicarán los elementos de la muestra. Finalmente arrastramos la fórmula a las casillas de la parte inferior hasta completar los 14 elementos necesarios.

B	C	D	E
<b>Población</b>	<b>Muestra</b>		
1	26	n=14	
2	23		
3	10		
4	7		
5	4		
6	15		
7	5		
8	35		
9	32		
10	6		
11	43		
12	2		
13	45		
14	37		
15			
16			
17			
18			
19			
20			

**Figura 4.10** Función para el muestreo aleatorio en Excel.

El muestreo aleatorio simple aparenta cierta facilidad para su uso, sobre todo cuando la población es pequeña. Sin embargo, pierde utilidad práctica cuando la población es muy grande.

### **Tabla de números aleatorios**

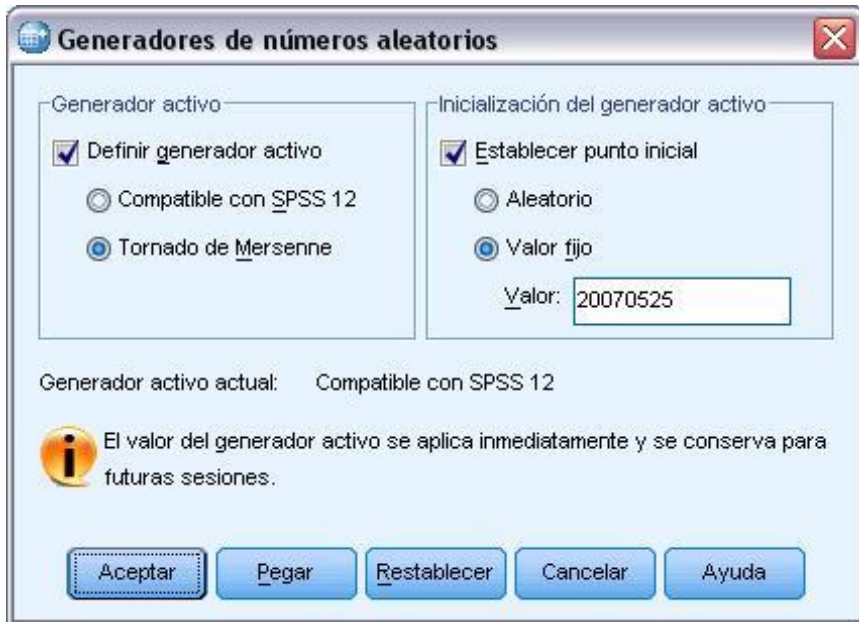
Una tabla de números aleatorios contiene números que han sido elegidos al azar. Por ejemplo, números elegidos por medio de un sorteo. Los libros de estadística inferencial, generalmente, presentan tablas de

números aleatorios en sus anexos. Sin embargo, actualmente se pueden encontrar diferentes tablas de números aleatorios en la web.

73735	45963	78134	63873
02965	58303	90708	20025
98859	23851	27965	62394
33666	62570	64775	78428
81666	26440	20422	05720
15838	47174	76866	14330
89793	34378	08730	56522
78155	22466	81978	57323
16381	66207	11698	99314
75002	80827	53867	37797
99982	27601	62686	44711
84543	87442	50033	14021
77757	54043	46176	42391
80871	32792	87989	72248
30500	28220	12444	71840

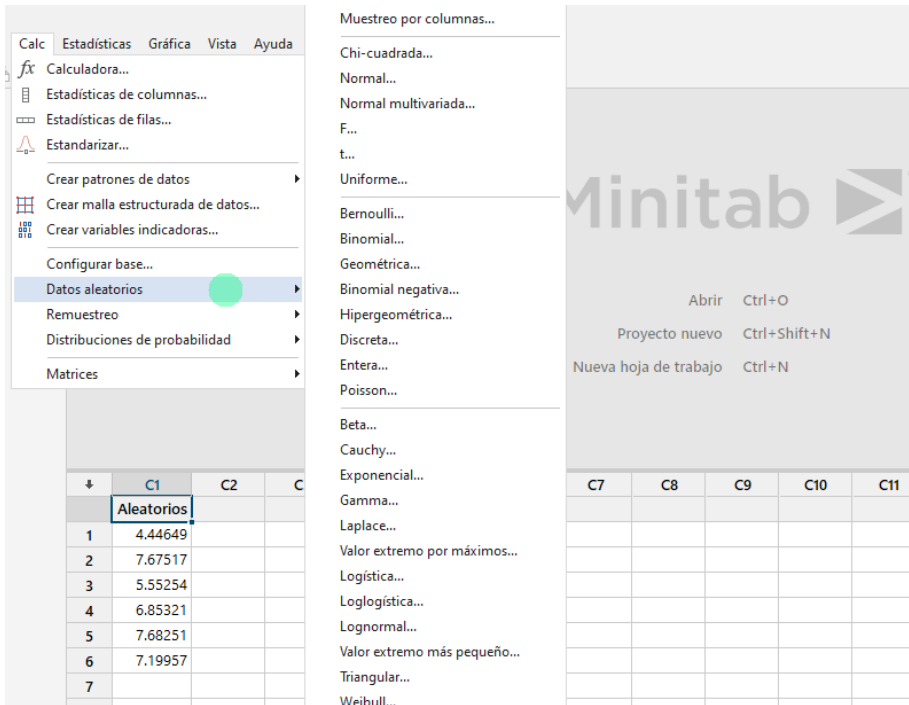
**Figura 4.11** Esta es una parte del libro Million Random Digits  
Fuente: Rand Corporation

Existen muchos programas estadísticos que brindan opciones para generar números aleatorios, por ejemplo, el SPSS versión 25 brinda la siguiente opción:



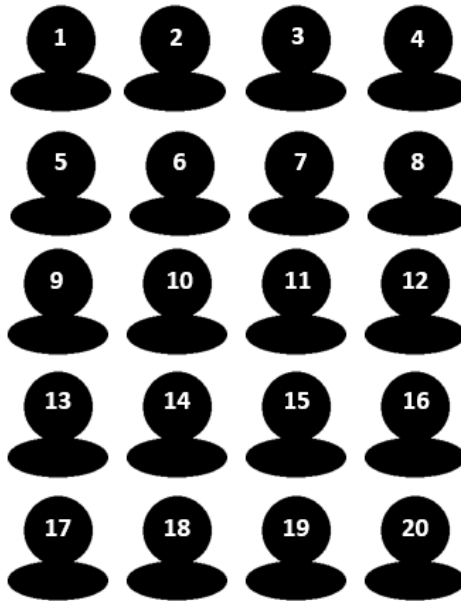
**Figura 4.12** Generación de números aleatorios en SPSS.

El programa estadístico MINITAB también nos ofrece la opción para calcular números aleatorios.



**Figura 4.13** Generación de números aleatorios en Minitab.

**Ejemplo 1:** Dado el siguiente grupo de 20 individuos. Seleccione una muestra de 8 individuos usando la tabla de números aleatorios.



**Figura 4.14** Población inicial.

Usaremos la tabla de números aleatorios mostrada en la sección anterior. Cochran (1980) indica que se puede trabajar por columnas, en nuestro caso usaremos las dos primeras columnas de cada bloque porque tenemos que seleccionar números de 1 al 20 (dos cifras como máximo). Si tuviésemos que elegir números hasta el 120, usaríamos las tres primeras columnas de cada bloque (tres cifras como máximo).

73735	45963	78134	63873
02965	58303	90708	20025
98859	23851	27965	62394
33666	62570	64775	78428
81666	26440	20422	05720
15838	47174	76866	14330
89793	34378	08730	56522
78155	22466	81978	57323
16381	66207	11698	99314
75002	80827	53867	37797
99982	27601	62686	44711
84543	87442	50033	14021
77757	54043	46176	42391
80871	32792	87989	72248
30500	28220	12444	71840

**Figura 4.15** Primeras columnas de la tabla de números aleatorios.

Podemos iniciar la selección de números desde cualquier bloque. En nuestro caso, iniciaremos desde el primer bloque y seleccionaremos, hacia abajo, los 8 primeros números que se encuentran en el intervalo del 1 al 20.

73735	45963	78134	63873
02965	58303	90708	20025
98859	23851	27965	62394
33666	62570	64775	78428
81666	26440	20422	05720
15838	47174	76866	14330
89793	34378	08730	56522
78155	22466	81978	57323
16381	66207	11698	99314
75002	80827	53867	37797
99982	27601	62686	44711
84543	87442	50033	14021
77757	54043	46176	42391
80871	32792	87989	72248
30500	28220	12444	71840

**Figura 4.16** Elección de los primeros números de 1 a 20.

Luego de seleccionar de arriba hacia abajo en la primera columna, se sigue pasando a la siguiente columna de la derecha hasta completar la cantidad de números que componen la muestra.

Finalmente, los elementos seleccionados para la muestra son:



**Figura 4.17** Muestra final usando la tabla de números aleatorios por columnas.

Anderson et al. (2012) indican que la selección de la muestra usando la tabla de números aleatorios también se puede hacer por filas.



**Ejemplo 2:** Dado un grupo de 20 individuos previamente enumerados. Seleccione una muestra de 8 individuos usando la tabla de números aleatorios.

73735	45963	78134	63873
02965	58303	90708	20025
98859	23851	27965	62394
33666	62570	64775	78428
81666	26440	20422	05720
15838	47174	76866	14330
89793	34378	08730	56522
78155	22466	81978	57323
16381	66207	11698	99314
75002	80827	53867	37797
99982	27601	62686	44711
84543	87442	50033	14021
77757	54043	46176	42391
80871	32792	87989	72248
30500	28220	12444	71840

**Figura 4.18** Primeras filas de la tabla de números aleatorios.

Como el mayor número 20 tiene dos cifras, formamos parejas de números con los dígitos de cada fila:

73	73	54	59	63	78	13	46	38	73
02	96	55	83	03	90	70	82	00	25
98	85	92	38	51	27	96	56	23	94
33	66	66	25	70	64	77	57	84	28
81	66	62	64	40	20	42	20	67	20
15	83	84	71	74	76	86	61	43	30
89	79	33	43	78	08	73	05	65	22
78	15	52	24	66	81	97	85	73	23
16	38	16	62	07	11	69	89	93	14

**Figura 4.19** Se unen todos los números de las primeras filas en parejas.

Luego se selecciona los 8 primeros números en el intervalo del 1 al 20 siguiendo la fila de izquierda a derecha y pasando luego a la siguiente fila. Cabe resaltar que el proceso no necesariamente se inicia en la primera fila.

73	73	54	59	63	78	<b>13</b>	46	38	73
<b>02</b>	96	55	83	<b>03</b>	90	70	82	00	25
98	85	92	38	51	27	96	56	23	94
33	66	66	25	70	64	77	57	84	28
81	66	62	64	40	<b>20</b>	42	20	67	20
<b>15</b>	83	84	71	74	76	86	61	43	30
89	79	33	43	78	<b>08</b>	73	<b>05</b>	65	22
78	15	52	24	66	81	97	85	73	23
<b>16</b>	38	16	62	07	11	69	89	93	14

**Figura 4.20** Se eligen los primeros números de 1 a 20.

Observe que no podemos elegir el mismo número más de una vez y que dejamos de lado el 00 dado que el menor número utilizado para la codificación de la población es 1.

Por lo tanto, la muestra seleccionada usando la tabla de números aleatorios por filas es:

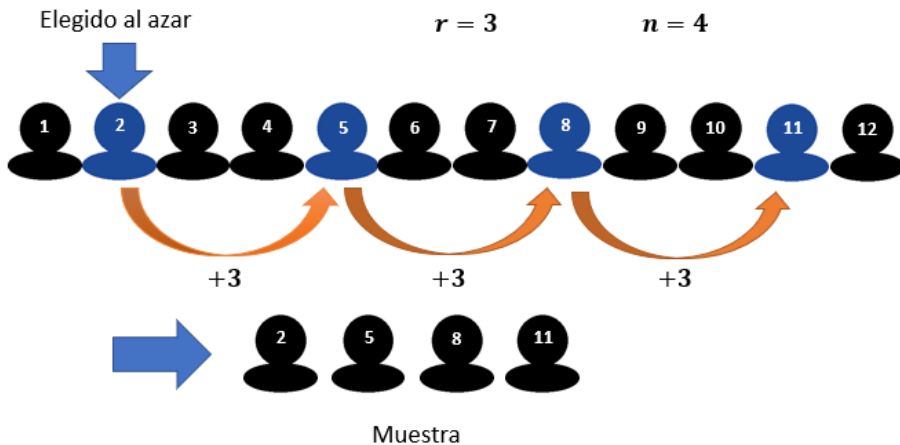


**Figura 4.21** Muestra final usando la tabla de números aleatorios por filas.

#### 4.2.1.2. Muestreo sistemático

En este tipo de muestreo se enumeran todos los elementos de la población, pero en lugar de extraer los  $n$  números de forma aleatoria

primero se extrae uno al azar. Luego se procede a seleccionar los siguientes elementos de acuerdo a un patrón recurrente.



**Figura 4.22** Muestreo sistemático para una muestra de tamaño 4.

### Procedimiento

- En primer lugar, se debe enumerar a todos los elementos de la población.
- Después, se calcula la constante  $r$  dividiendo el tamaño de la población entre el tamaño de la muestra (Blanco, 2011):

$$r = \frac{N}{n}$$

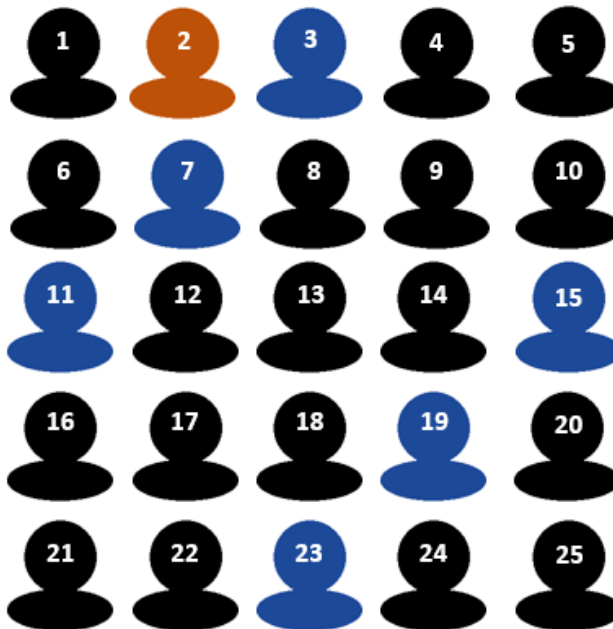
- Luego se elige el primer elemento de la posición  $a$ . Este número  $a$  puede tomar cualquier valor desde 1 hasta  $r$ . El número  $a$  es elegido de forma aleatoria (lanzamiento de un dado, sorteo, etc.).
- Los siguientes elementos que integran la muestra son los que ocupan la posición  $a + r$ ,  $a + 2r$ ,  $a + 3r$ , ...,  $a + (n - 1)r$ . Hay que resaltar que  $n$  es un entero mayor que cero.

**Ejemplo 1:** De una población de 25 individuos se debe tomar una muestra de tamaño 7. Seleccione la muestra utilizando el muestreo sistemático.

Luego de enumerar los 25 elementos de la población, calculamos la constante  $r$ :

$$r = \frac{N}{n} = \frac{25}{7} = 3.571 \approx 4$$

Para elegir el primer elemento realizamos un sorteo con los números del 1 al 4 y obtenemos, por ejemplo, el número 3. El elemento que se encuentra en la posición 3 es nuestro primer elemento seleccionado. Los demás elementos se eligen recorriendo posiciones de 4 en 4.



**Figura 4.23** Muestreo sistemático con  $r=4$  y  $n=7$ .

Luego de seleccionar el elemento de la posición 3 recorreremos 4 posiciones retornando a las primeras posiciones, de ser necesario. Por

lo tanto, la muestra seleccionada se compone de los siguientes elementos:



**Figura 4.24** Muestra final usando el muestro sistemático.

**Ejemplo 2:** De una población de 58 individuos se debe tomar una muestra de tamaño 12. Seleccione la muestra utilizando el muestreo sistemático.

Luego de enumerar los 58 elementos de la población, calculamos la constante  $r$ :

$$r = \frac{N}{n} = \frac{58}{12} = 4.83 \approx 5$$

Para elegir el primer elemento realizamos un sorteo con los números del 1 al 5 y obtenemos, por ejemplo, el número 2. El elemento que se encuentra en la posición 2 es nuestro primer elemento seleccionado. Los demás elementos se eligen recorriendo posiciones de 5 en 5. Pero esta vez utilizaremos directamente el programa Excel.

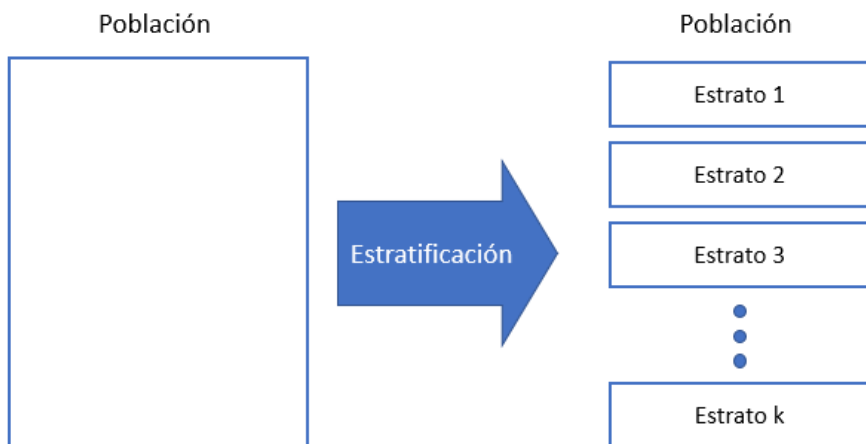
B	C	D
<b>Población</b>		<b>Muestra</b>
1		2
2	r=5	7
3		12
4		17
5		22
6	Se elige, al azar, al número 2 como primer elemento.	27
7		32
8		37
9		42
10		47
11		52
12		57
13		
14		
15		
16		
17		

**Figura 4.25** Muestreo sistemático usando Excel.

En la figura 4.25 se puede observar que para obtener los elementos luego del primer elemento de la posición 2 se le fue sumando 5 progresivamente a cada elemento siguiente.

#### 4.2.1.3. Muestreo estratificado

Es aquel muestreo en el que se divide una población de  $N$  individuos, en  $k$  subpoblaciones o estratos, atendiendo a criterios de división importantes para el estudio.



**Figura 4.26** Estratificación de la población.

Los criterios para establecer la estratificación son diversos y dependen del conocimiento que tenga el investigador sobre la población en estudio. Así, si se trabaja con grupos humanos, se puede realizar una estratificación por sexo, grupo etario, estado civil, nivel de instrucción, etc. De los criterios que se puedan establecer para crear los estratos el investigador decide cuales son los más pertinentes y los que generan un mayor aporte al momento de realizar el análisis de resultados.

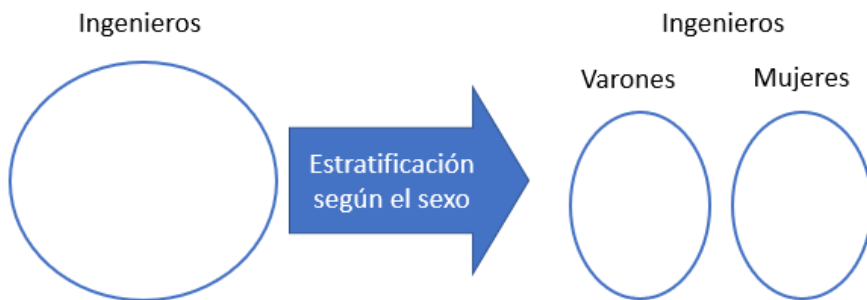


**Figura 4.27** Ejemplo de estratificación de las empresas.

Fuente: INEI

En la figura 4.27 el conjunto de las empresas nuevas y reactivadas ha sido estratificado según el tipo de actividad de la empresa.

Los estratos se forman de acuerdo a características mutuamente excluyentes. Es decir, de acuerdo con algún criterio de estratificación, los elementos de un estrato no tienen posibilidad de ser incluidos en otro estrato. Navarro (2011) define a los estratos como grupos de la población que no presentan traslapes. El término traslape puede entenderse mejor teniendo en cuenta los diagramas de Venn, cuando dos conjuntos tienen elementos en común, gráficamente se representa la intersección como un traslape entre los conjuntos. Cuando los conjuntos no tienen elementos en común no se observa ningún traslape y los conjuntos pueden representarse totalmente separados.



**Figura 4.28** Población dividida en dos estratos.

En la figura 4.28 se observa que los estratos varones y mujeres son conjuntos disjuntos.

**Ejemplo 1:** Un investigador planifica un estudio para conocer las características óptimas para la conservación de la carne roja para consumo humano. Por ello realiza la siguiente estratificación:



**Tabla 4.1** Estratificación de carnes rojas.

Criterio de estratificación	Estrato
Tipo de ganado	Vacuno
	Ovino
	Porcino
	Caprino

Los tamaños de cada estrato se pueden simbolizar de la siguiente manera:  $N_i$ , donde  $i$  es un número entero  $i = 1, 2, 3, \dots; k$  y  $k$  es la cantidad de estratos:

$$N_1, N_2, N_3, \dots, N_k$$

Los estratos son particiones de la población. Es decir, los estratos son conjuntos disjuntos, no comparten elementos en común. Por lo tanto, se cumple que la suma de los tamaños de los estratos es igual al tamaño de la población.

$$N = N_1 + N_2 + N_3 + \dots + N_k$$

**Ejemplo 2:** Si se realiza un estudio sobre las perspectivas que tienen los trabajadores de la empresa “X” con respecto a su jubilación, podemos realizar una estratificación de acuerdo al sistema de pensiones al que pertenecen. Además, se sabe que la empresa cuenta con 137 empleados registrados en planilla.

**Tabla 4.2** Estratificación según el sistema de pensiones.

Criterio	Estrato	Tamaño del estrato
Sistema de pensiones	Privado	$N_1 = 84$
	Nacional	$N_2 = 53$

En la tabla 4.2 se puede verificar que la suma de los tamaños de los estratos es igual al tamaño de la población:

$$84 + 53 = 137$$

Con este mismo criterio se puede establecer que la suma de los tamaños de las muestras para cada estrato de la población debe ser igual al tamaño de la muestra total. Es decir, si realizamos muestreos aleatorios simples de tamaño  $n_i$  en cada uno de los estratos, donde  $i$  es un número entero  $i = 1, 2, 3, \dots; k$ , se obtiene:

$$n = n_1 + n_2 + n_3 + \dots + n_k$$

Recordemos que  $n$  es el tamaño de la muestra. Más adelante se abordarán los procedimientos para el cálculo del tamaño de muestra  $n$ .

**Tabla 4.3** Simbología para el tamaño del estrato y del tamaño de muestra para cada estrato.

Estratos	Tamaño de los estratos	Tamaño de la muestra de cada estrato
$E_1$	$N_1$	$n_1$
$E_2$	$N_2$	$n_2$
$E_3$	$N_3$	$n_3$
$E_k$	$N_k$	$n_k$
Total	$N$	$n$

Anderson et al. (2012) señala que se obtienen mejores resultados cuando los elementos que conforman cada estrato son lo más parecidos en lo posible. Es decir, cuando los elementos del estrato 1 son muy similares entre sí, los elementos del estrato 2 son similares entre sí, etc.

### Afijación

En este apartado se muestra un acercamiento al proceso para establecer la cantidad de elementos muestrales para cada estrato de la población. Cochran (1980) señala que los tamaños de muestra para cada estrato son definidos por el encargado de hacer el muestreo y el proceso responde a criterios de minimización del costo de la varianza de la estimación de la media de la población.

La distribución de los tamaños de muestra para cada estrato se denomina afijación (Flores et al., 2008), y puede ser de diferentes tipos:

**Afijación Simple:** En este caso a cada estrato le corresponde igual número de elementos muestrales. También se le denomina afijación uniforme.

**Ejemplo 1:** En un congreso estudiantil universitario se realiza una encuesta a los estudiantes. Se sabe que estos provienen de diferentes distritos como se muestra a continuación. Realiza un muestreo estratificado con afijación simple si se quiere encuestar a 400 estudiantes en total.

**Tabla 4.4** Tamaño de los estratos

Distrito	Cantidad de estudiantes
San Luis	250
San Borja	200
Jesús maría	150
Lince	140
San Miguel	90
Chorrillos	100
Miraflores	90
Magdalena	80

Dado que a cada uno de los 8 estratos se le debe asignar la misma cantidad de elementos muestrales, realizamos la división:

$$\frac{400}{8} = 50$$

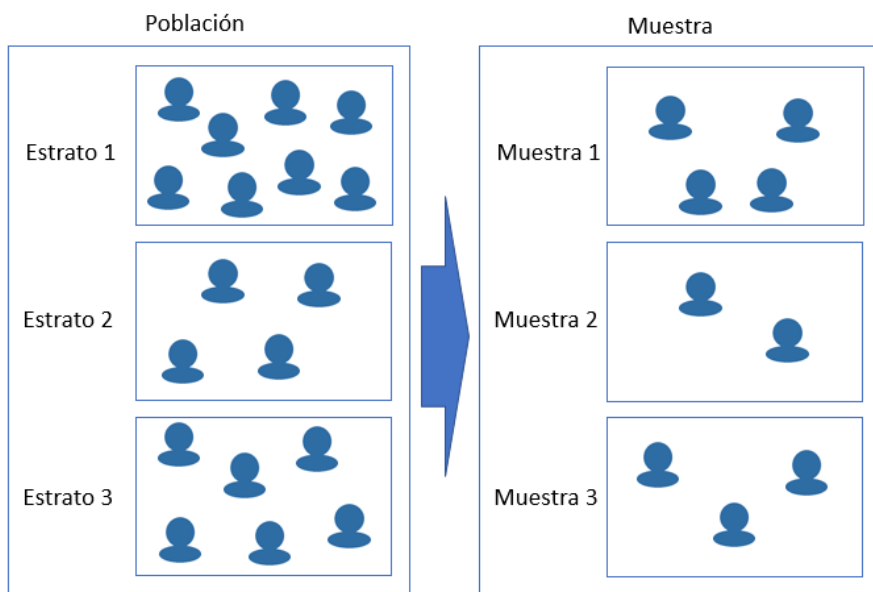
**Tabla 4.5** Tamaño de la muestra para cada estrato

Distrito	Cantidad de estudiantes
San Luis	50
San Borja	50
Jesús maría	50
Lince	50

San Miguel	50
Chorrillos	50
Miraflores	50
Magdalena	50

---

**Afijación Proporcional:** La distribución los tamaños de muestra se hace de acuerdo con el tamaño de la población en cada estrato.



**Figura 4.29** El tamaño de cada muestra es la mitad del tamaño de su respectiva población.

En la figura 4.29 se observa que el tamaño del estrato 1 es mayor al tamaño del estrato 2 y al tamaño del estrato 3, por lo tanto, le corresponde un mayor tamaño de muestra. Asimismo, el tamaño del estrato 2 es menor al tamaño de los otros estratos, por lo tanto, le corresponde el menor tamaño de muestra. También se puede verificar que el tamaño de la muestra es la mitad del tamaño de la población ( $n=9$  y  $N=18$ ), por ello, el tamaño de cada sub muestra es la mitad del tamaño de cada estrato.

Sea  $n$  el número de individuos de la población total que forman parte de alguna muestra, entonces se cumple:

$$n = n_1 + n_2 + n_3 + \dots + n_k$$

Cuando la asignación es proporcional el tamaño de la muestra de cada estrato es proporcional al tamaño del estrato correspondiente con respecto a la población total:

$$n_i = n \times \frac{N_i}{N}$$

También se puede utilizar la constante de proporcionalidad:

$$c = \frac{n}{N}$$

Esta constante  $c$  se multiplica por el tamaño de la población de cada estrato y se obtiene el tamaño de muestra para cada estrato.

$$n_i = c \times N_i$$

**Ejemplo 1:** Una empresa realiza un estudio sobre la preferencia de cierto tipo de bebida en los colegios de Lima norte. Para ello trabajará con los colegios privados de los distritos de Puente Piedra, Ancón, Carabayllo y Santa Rosa. Además, se piensa utilizar una muestra de 100 colegios. Realiza el muestreo estratificado proporcional.

**Tabla 4.6** Tamaño de la población de cada colegio.

Distrito	Cantidad de colegios
Puente Piedra	515
Ancón	75
Carabayllo	499
Santa Rosa	31

Fuente: Google Data Studio 2021

Primero calculamos el tamaño de la población  $N$ :

$$N = 515 + 75 + 499 + 31 = 1120$$

Luego calculamos la constante de proporcionalidad  $c$ :

$$c = \frac{n}{N} = \frac{100}{1120} = 0.089$$

Después multiplicamos el tamaño de cada estrato por la constante  $c$ :

$$515 \times 0.089 = 45.835$$

$$75 \times 0.089 = 6.675$$

$$499 \times 0.089 = 44.411$$

$$31 \times 0.089 = 2.759$$

Luego aproximamos los resultados a cantidades enteras.

**Tabla 4.7** Tamaño de la muestra para cada estrato.

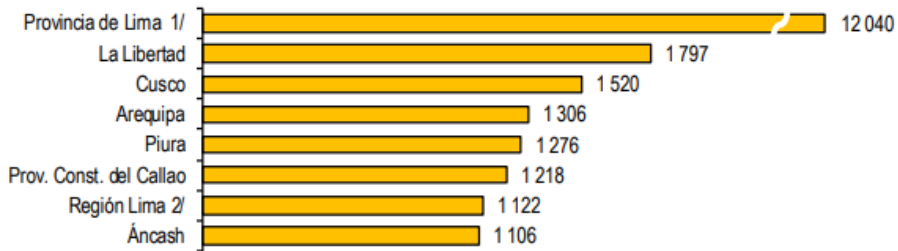
Distrito	Cantidad de colegios
Puente Piedra	46
Ancón	7
Carabayllo	44
Santa Rosa	3

Se puede verificar que la suma de los tamaños de muestra para cada estrato es igual al tamaño de muestra total.

$$46 + 7 + 44 + 3 = 100$$

**Ejemplo 2:** El ente fiscalizador de las municipalidades del Perú planea analizar los conocimientos sobre los protocolos de trabajo y el desempeño laboral en los serenos de las municipalidades del Perú que tienen un mayor número de personal. Para este trabajo de investigación se calculó una muestra de 584 serenos. Realiza el muestreo estratificado proporcional.

Para realizar el muestreo primero seleccionamos los municipios que tienen más de 1 000 serenos con participación activa con base en la información presentada por el INEI en el 2019.



**Figura 4.30** Personal de serenazgo al 31 de marzo de 2019.  
Fuente: INEI - Registro Nacional de Municipalidades 2019.

Primero calculamos el tamaño de la población  $N$ :

$$N = 12040 + 1797 + 1520 + \dots + 1106 = 21\,385$$

Luego calculamos la constante de proporcionalidad  $c$ :

$$c = \frac{n}{N} = \frac{584}{21\,385} = 0.0273$$

Después multiplicamos el tamaño de cada estrato por la constante  $c$ :

$$\begin{aligned} 12040 \times 0.0273 &= 328.692 \\ 1797 \times 0.273 &= 49.0581 \\ 1520 \times 0.0273 &= 41.496 \\ 1306 \times 0.0273 &= 35.6538 \\ 1276 \times 0.0273 &= 34.8348 \\ 1218 \times 0.0273 &= 33.2514 \\ 1122 \times 0.0273 &= 30.6306 \\ 1106 \times 0.0273 &= 30.1938 \end{aligned}$$

Luego aproximamos los resultados a cantidades enteras.

**Tabla 4.8** Tamaño de la muestra para cada estrato.

Municipalidad	Cantidad de serenos
Provincia de Lima	329
La Libertad	49
Cusco	41
Arequipa	36
Piura	35
Prov. Const. del Callao	33
Región Lima	31
Ancash	30

**Afijación Óptima:** Para este tipo de afijación se tiene en consideración el grado de dispersión de los resultados. Por ello, para el cálculo de los tamaños muestrales de cada estrato se considera la proporción y la desviación típica. No tiene mucha aplicación ya que no se suele conocer la desviación.

El tamaño de muestra para cada estrato  $k$  se calcula mediante la expresión (Cochran, 1980):

$$n_k = \frac{n(W_k S_k)}{\sum W_k S_k}$$

Donde  $W_k$  y  $S_k$  corresponden a la ponderación y la desviación estándar en el estrato  $k$  respectivamente. Cochran también resalta que la fórmula mostrada permite deducir que la extracción de una muestra grande de un determinado estrato se justifica por el mayor tamaño del estrato, por una mayor variabilidad dentro del estrato o porque es más económico extraer la muestra de dicho estrato.

### **Cuando usar el muestreo estratificado.**

Cochran (1980) señala que existen ciertas situaciones en las que es recomendable utilizar el muestreo estratificado.

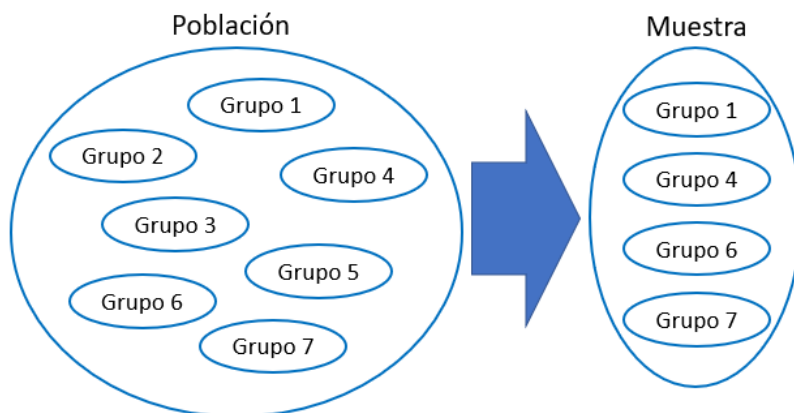


- La población consta de instituciones que varían mucho en tamaño.
- Las principales variables a medir están íntimamente relacionadas con los tamaños de las instituciones.
- Se cuenta con una buena medida de tamaño para establecer los estratos.

Supo (2014) señala que en el muestro estratificado los grupos o estratos son heterogéneos entre sí, pero homogéneos dentro de cada grupo. Esto resume la explicación desarrollada anteriormente donde se indica que los estratos son conjuntos disjuntos y los elementos que conforman cada estrato responden estrictamente a ciertos criterios de selección.

#### 4.2.1.4. Muestreo por conglomerados

En este tipo de muestreo una población grande se subdivide en grupos y se eligen algunos de estos grupos. Luego, se toman elementos de cada grupo elegido para conformar la muestra. Mason y Lind (1992) señalan que este tipo de muestreo se utiliza con la finalidad de reducir el costo de muestrear una población distribuida en un área geográfica grande.



**Figura 4.31** Población dividida en conglomerados.

En este tipo de muestreo, además de reducir el costo, se puede reducir el tiempo en la recolección de datos, teniendo en consideración que en

muchas ocasiones las unidades de análisis que nos interesan se encuentran encapsuladas o distribuidas en determinados lugares físicos o geográficos a los que se les llama racimos o conglomerados (Hernández et al., 2014).

Las unidades de análisis se encuentran en diferentes lugares que, luego de identificarlos, nos permiten realizar un muestreo más eficiente. Por ejemplo, cuando queremos analizar los hábitos de las amas de casa, los posibles racimos o conglomerados son los mercados, tiendas, supermercados, etc.



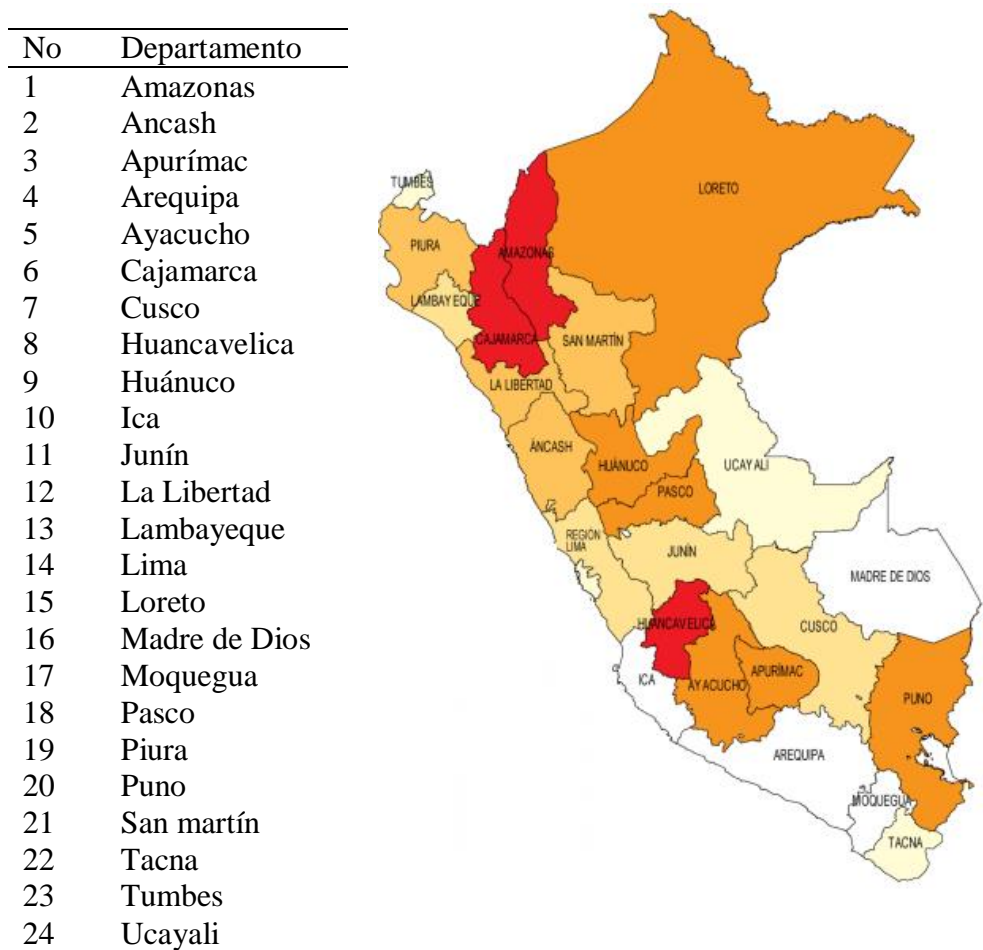
**Figura 4.32** Población de deportistas dividida en conglomerados.

### **Procedimiento:**

- Se divide a la población en grupos o conglomerados bajo ciertos criterios. Anderson, et al. (2012) indican que lo ideal es que cada grupo o conglomerado sea una representación, a pequeña escala, de la población completa.
- Se toma una muestra aleatoria simple de los conglomerados. Es decir, elegimos un grupo de conglomerados al azar y estos grupos en conjunto conforman la muestra.

Este muestreo tiende a proporcionar mejores resultados cuando los elementos dentro de los conglomerados no son semejantes (Anderson, et al., 2012).

**Ejemplo:** Si queremos analizar la satisfacción de las personas con respecto a la galleta “Doraditas” que se distribuye en todo el Perú, debemos reconocer los distintos departamentos que componen el territorio peruano. Cada departamento representa un conglomerado.



**Figura 4.33** Población peruana dividida por departamentos.

Según el muestreo por conglomerados se debe seleccionar solo algunos departamentos del país, esta selección se debe realizar de forma aleatoria.

**Tabla 4.9** Selección de conglomerados.

No	Departamento
1	Arequipa
2	Ayacucho
3	Cajamarca
4	Cusco
5	Huánuco
6	Ica
7	La Libertad
8	Lima
9	Puno
10	Tacna
11	Tumbes

Mayor, J. (2007) señala que después de elegir los conglomerados de forma aleatoria. Se deben analizar todos los elementos de cada conglomerado seleccionado, es decir se realiza un censo de cada conglomerado. Sin embargo, esta opción es algo compleja cuando la muestra es grande. La opción alterna consiste en realizar una segunda etapa de muestreo dentro de cada conglomerado.



**Figura 4.34** Muestreo por conglomerados en dos etapas.

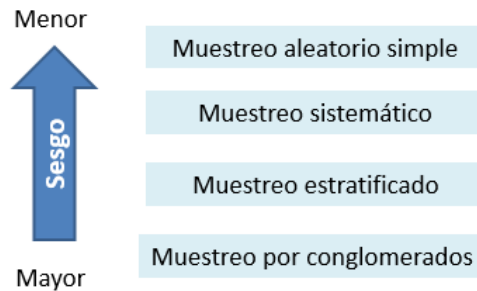
En la figura 4.34 se observa un muestreo por conglomerados con dos etapas. Sin embargo, se pueden agregar más etapas al muestreo y obtener un muestreo denominado de múltiples etapas.

#### 4.2.1.5. ¿Qué tipo de muestreo probabilístico debo utilizar?

En la exposición de cada uno de los tipos de muestro probabilístico se han establecido las características principales. Por ello, la elección del muestreo adecuado depende de las características de la investigación que se está realizando y de los recursos con que se cuenta para realizar dicho trabajo. Debemos agregar que los objetivos y el alcance de investigación también son relevantes en la elección de método de muestreo. Esto último es evidente, por ejemplo, cuando elegimos una población a nivel nacional para que nuestros resultados tengan mayor alcance o cuando nos decidimos realizar nuestra investigación solo en una determinada región del país. Se suele pensar que en el primer caso podemos optar por un muestro estratificado o por conglomerados y en el segundo caso se puede optar por un muestreo aleatorio o sistemático, sin embargo, en ambos casos se puede elegir cualquiera de los tipos de muestreo.

A continuación, se muestra un ejemplo donde se aprecia como se pueden aplicar los diferentes tipos de muestreo a una misma situación. Esto quiere decir que el investigador debe analizar diferentes escenarios e identificar las características de su investigación y como estas se adecúan a un determinado tipo de muestreo para hacer la elección del proceso adecuado.

Supo (2014) resalta que el muestreo probabilístico presenta menor sesgo en comparación con las técnicas de muestreo no probabilístico. Sin embargo, dentro de las técnicas de muestro probabilístico también existe sesgo. Supo nos orienta nuevamente indicando que se puede ordenar estas técnicas de acuerdo a la magnitud del sesgo que puedan generar.



**Figura 4.35** Ordenamiento de los tipos de muestreo según el sesgo.

**Ejemplo:** En una investigación, por medio de una encuesta, se desea analizar las opiniones de los estudiantes de la universidad “X” con respecto al retorno a las clases presenciales en el año 2022. Explique la aplicación del muestreo aleatorio simple, el muestreo sistemático, el muestreo estratificado y el muestreo por conglomerados.

**Tabla 4.10** Aplicación de los tipos de muestreo probabilístico.

Población	Muestreo	Procedimiento
Estudiantes de la universidad “X”.	Aleatorio simple	Se enumera a todos los estudiantes de la universidad. Mediante un sorteo entre todos ellos se elige un grupo de tamaño $n$ y se les toma la encuesta.
	Sistemático	Se enumera a todos los estudiantes de la universidad. Se aproxima $r = n/N$ al entero más cercano. Se elige al azar un entero entre 1 y $r$ , este número indica la posición del primer estudiante que será encuestado. Los demás estudiantes a encuestar serán elegidos de $r$ en $r$ posiciones

---

	a partir del primer estudiante encuestado.
Estratificado	La población se separa por ciclos y se selecciona una muestra aleatoria simple de estudiantes de cada ciclo, finalmente se les toma la encuesta a los estudiantes seleccionados.
Por conglomerados	Se elige 6 de los 10 ciclos de forma aleatoria y se encuesta a cada uno de los estudiantes de los ciclos elegidos.

---

#### 4.2.2. Muestreo no probabilístico

En este tipo de muestreo la elección de los elementos de la muestra no depende de la probabilidad. La elección de estos elementos es decisión del investigador. En este tipo de muestro se seleccionan solo elementos que cumplan ciertas características para beneficiar a la investigación con mejores resultados.

El muestreo no probabilístico es utilizado cuando los investigadores realizan investigaciones cualitativas, estudios piloto o investigación exploratoria.

De acuerdo con Rubio (citado por Kleeberg y Ramos, 2009) los resultados obtenidos en una investigación basada en el muestreo no probabilístico no son generalizables, es decir no se puede hacer inferencia estadística.

#### 4.2.2.1. Por conveniencia o accidental

Como todo muestreo no probabilístico el investigador deliberadamente trata de obtener muestras representativas. Por ello, elige elementos que presenten las mismas condiciones y tengan las mismas características para que conformen la muestra.

En este caso la selección de la muestra depende de la facilidad en el acceso del investigador a los posibles elementos muestrales. Es decir, el investigador puede elegir una muestra porque el acceso a ella es más sencillo.

Por ejemplo, un docente que desea investigar sobre los problemas de aprendizaje en los estudiantes podría utilizar a sus propios estudiantes como muestra por la facilidad que esto implica en el proceso de recolección de datos.

Kleeberg y Ramos (2009) afirman que las muestras por conveniencia o accidentales están conformadas por elementos que voluntariamente acceden a ser observados. En esta descripción también se pone en evidencia la facilidad que encuentra el investigador en trabajar con aquellos elementos que son más accesibles durante el proceso de muestreo.

Otro ejemplo muy recurrente para este tipo de muestreo es cuando se realiza una encuesta en alguna institución y se toma como muestra a las 10 o 15 primeras personas (según lo que encuentre disponible) que voluntariamente respondieron a la encuesta.

#### 4.2.2.2. Juicio de expertos o de criterio

Supo (2014) señala que en este tipo de muestreo el criterio para la selección de la muestra puede ser del investigador o de un grupo de expertos. En el caso del criterio del investigador, se recurre a su



conocimiento y su experiencia de modo que pueda seleccionar a los elementos que reúnen las condiciones para ser parte de la muestra.

Hernández et al. (2014) indican que en este tipo de muestreo solo se incluye dentro de los elementos de la muestra a los expertos necesarios de acuerdo con el tema de investigación. Es decir que, si se ha de realizar una investigación cualitativa sobre el impacto de las condiciones de pandemia en el sistema educativo peruano, la muestra para el estudio debe estar compuesta de expertos en educación.

Como se aprecia Supo y Hernández et al. presentan versiones distintas de este tipo de muestreo, pero ambas versiones son complementarias. Por ejemplo, cuando se quiere realizar una investigación sobre la formación musical de los diferentes artistas de música peruana podemos confiar por nuestro conocimiento sobre estos cantantes, siempre y cuando se tenga un conocimiento sólido incluyendo a cantantes que cultivan los géneros de música peruana que no son muy comerciales. La otra opción sería recurrir a los expertos que están representados por los directores de las escuelas musicales donde los artistas inician o complementan su formación musical. Así, estos directores nos pueden brindar información sobre la trayectoria del artista o nos puede referenciar artistas cuyas características son las adecuadas para nuestra investigación.

#### 4.2.2.3. Por cuotas

Cochran (1980) señala que el muestro por cuotas puede definirse como un muestreo estratificado donde la selección de elementos para cada estrato no es estrictamente aleatoria. También señala que este tipo de muestreo es muy utilizado en los sondeos de opinión y la investigación de mercados.

Para este tipo de muestro se debe contar con un buen conocimiento de la población para poder establecer los estratos. Las cuotas se completan de acuerdo a las variables demográficas de la población que se han definido previamente.

Entre las variables sociodemográficas que se utilizan para la estratificación se pueden mencionar:

- Edad
- Sexo
- Estado civil
- Ocupación
- Área geográfica de residencia
- Ingreso económico
- Nivel educativo
- Etnia

Por ejemplo, en una encuesta de opinión podemos considerar que nuestra muestra este compuesta por un 50% de personas con nivel educativo básico, 30% de personas con nivel educativo técnico y 20% de personas con estudios universitarios. También se puede establecer que el 60% de los encuestados sean mujeres y el 40% varones. Estas condiciones representan las cuotas.

#### 4.2.2.4. Bola de nieve

En algunas investigaciones de tipo exploratorio o cualitativo los elementos que componen la muestra no se pueden localizar con facilidad. Para acceder a estos elementos se puede usar un sistema de información de modo tal que a través de algunos elementos de la muestra ya localizada nos permitan contactar con otros sujetos de características similares y adecuadas para completar la muestra.

Atkinson y Flint (citados por Baltar y Gorjup, 2012) definen este tipo de muestreo como una técnica que para encontrar a los sujetos que componen la muestra a través de una serie de referencias. Es decir, un sujeto que compone la muestra inicial da al investigador el nombre de otro posible sujeto adecuado para la muestra, y este proporciona el nombre de un sujeto más, y así sucesivamente.

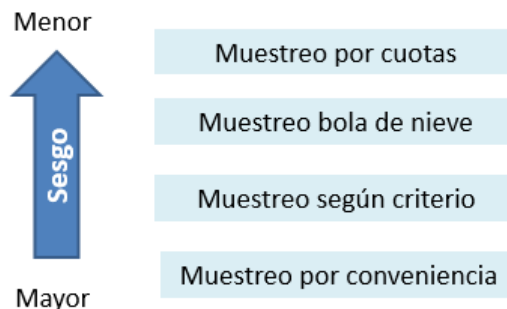
Este tipo de muestreo se utiliza en investigaciones donde los elementos que se desea analizar se presentan en número escaso y es difícil acceder a ellos.

Baltar y Gorjup (2012) indican que los primeros elementos del muestreo bola de nieve, por cuestiones prácticas, son elegidos a través de un muestreo de conveniencia.

Por ejemplo, en una investigación donde se trata de conocer las características psicológicas y los hábitos de las personas coleccionistas de discos de vinilo no es fácil acceder a los elementos de la muestra. No existe un registro de coleccionistas de discos de vinilo, sin embargo, en redes sociales podemos encontrar grupos de coleccionistas, de esta forma podemos iniciar el muestreo. Luego, nuestros encuestados nos pueden referenciar otros grupos de coleccionistas donde podemos encontrar más elementos muestrales y de esa se forma proseguimos el muestreo hasta completar el tamaño muestral que se considera necesario.

#### 4.2.2.5. ¿Qué tipo de muestreo no probabilístico debo utilizar?

De la misma forma que en el muestreo probabilístico, la elección del tipo de muestreo depende de las condiciones y recursos con que se cuenta para realizar la investigación. Sin embargo, podemos recurrir a lo señalado por Supo (2014) con respecto al orden de las técnicas de muestro según el sesgo que puedan generar.



**Figura 4.36** Ordenamiento de los tipos de muestreo según el sesgo.

## Capítulo 5. Cálculo del tamaño de la muestra

No existe una fórmula general que nos permita determinar el tamaño de la muestra para todo tipo de investigación. El tamaño de la muestra depende, entre otros factores, de la variabilidad de la población en estudio. En este sentido, se puede concluir que mientras más homogénea sea la población, menor será el tamaño de muestra requerido para establecer las conclusiones.

Por ejemplo, si se quiere determinar el efecto de la temperatura sobre la durabilidad de cierto tipo de embutido, no se necesita una muestra muy grande, debido a que los embutidos han sido parte de un mismo proceso de elaboración, poseen casi la misma consistencia, por tanto, se puede decir que provienen de una población homogénea. En cambio, cuando se quiere determinar si las condiciones laborales afectan la productividad de los empleados de una empresa, no se puede trabajar con una muestra pequeña debido a que las personas son seres complejos y esto acentúa el grado de variabilidad de la población de empleados.

Esto último, no quiere decir que siempre que se trate de poblaciones compuestas de personas se requiera una muestra grande. Si se aplican criterios adecuados de inclusión y exclusión se puede establecer una población con características similares de modo que la variabilidad se puede reducir notablemente.

Para comprender de donde se obtienen las fórmulas para establecer el tamaño de la muestra en los siguientes apartados, es necesario recurrir a algunas nociones previas como los intervalos de confianza.

Por ejemplo, se puede determinar el tamaño de la muestra a partir del intervalo de confianza para la media:

$$\mu \in \left[ \bar{X} - \frac{Z_{\alpha/2} \sigma}{\sqrt{n}}; \bar{X} + \frac{Z_{\alpha/2} \sigma}{\sqrt{n}} \right]$$

En términos más simples, si se tiene que la media muestral de un conjunto de masas corporales es igual a 35 kg y existe un error de estimación de 4.8 kg para la media poblacional, se puede decir que la media poblacional es  $(35 \pm 4.8)$  kg. Entonces, el intervalo de confianza para la media poblacional se puede expresar de la siguiente manera:

$$\mu \in [35 - 4.8; 35 + 4.8]$$

Restando y sumando el error de estimación a la media muestral tenemos.

$$\mu \in [30.2; 39.8]$$

Es decir, la media poblacional se encuentra en el intervalo de 30.2 kg a 39.8 kg.

Si consideramos una forma más sencilla de expresar el intervalo de confianza podemos utilizar:

$$\mu \in [\bar{X} - e; \bar{X} + e]$$

Donde  $e$  es el error de estimación para la media poblacional. Note que, si el error es grande, al intervalo de confianza tiene mayor amplitud. En otras palabras, para hacer una estimación más precisa, el error debe ser numéricamente más pequeño.

Para hacer la estimación de parámetros, el investigador puede establecer un máximo valor para el error de estimación, teniendo en cuenta que este error debe ser pequeño. Si consideramos que  $e$  es el máximo error de estimación, podemos establecer la siguiente desigualdad para deducir el tamaño de la muestra.

$$\frac{Z_{\alpha/2}\sigma}{\sqrt{n}} \leq e$$

Usamos la expresión menor o igual ya que el valor del primer miembro no excede al valor de  $e$ . Luego despejamos el valor de  $n$ , considerando que todos los términos de la expresión son positivos:

$$\frac{Z_{\alpha/2}\sigma}{e} \leq \sqrt{n}$$

$$\frac{Z_{\alpha/2}^2\sigma^2}{e^2} \leq n$$

Como se puede apreciar, se ha determinado el valor mínimo de  $n$  para que el error de estimación alcance el valor de estimación máximo que se había definido.

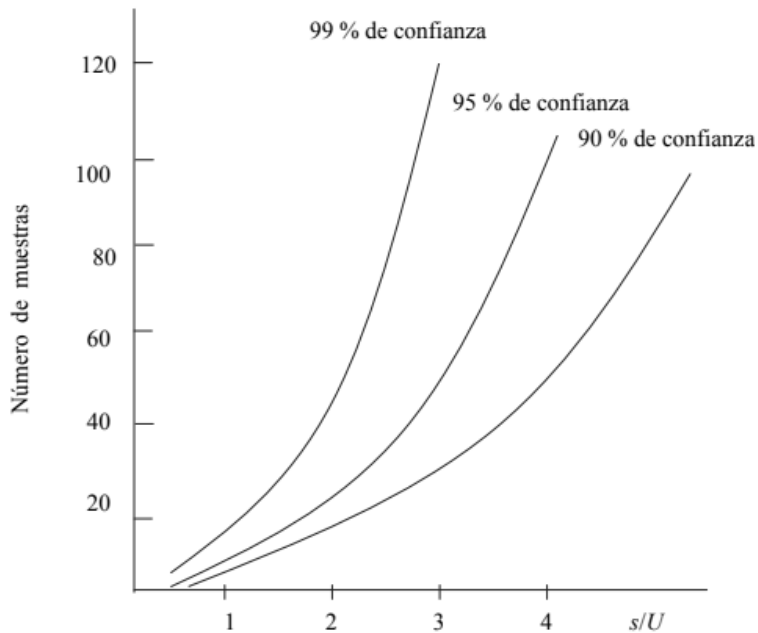
El proceso mostrado permite el cálculo del tamaño de muestra mínimo para determinar la media poblacional. Pero, en las investigaciones no siempre se quiere estimar el valor de la media. Existen otros parámetros y para cada uno de ellos existe un proceso particular para calcular el tamaño de muestra necesario a utilizar en el análisis estadístico de la variable en estudio.

Mason y Lind (1992) indican que existen tres factores que determinan el tamaño de la muestra sin tener en consideración el tamaño de la población. Estos factores son: el nivel de confianza, el máximo error permisible y la variabilidad de la población. Este último elemento, la variabilidad, se representa con la varianza o la desviación estándar. No obstante, los factores que afectan el tamaño de la muestra son distintos según el análisis que se desea realizar. Es decir, por ejemplo, en el cálculo del tamaño de la muestra para analizar la media de una población intervienen elementos que no están presentes en el cálculo del tamaño de muestra para analizar la proporción de cierta característica en una población.

García et al. (2013) inciden en que los factores estadísticos que intervienen en el tamaño de la muestra son:

- La hipótesis.
- El error tipo I o error  $\alpha$ .
- El error tipo II o error  $\beta$ .
- El poder estadístico.
- La variabilidad.
- Las pérdidas en el seguimiento del estudio.
- La relevancia del tamaño del efecto y significancia estadística.

De igual manera, Luna et al. (2011) plantean que el tamaño de la muestra depende del nivel de confianza seleccionado, la desviación estándar y el error aceptable de estimación. La desviación estándar hace referencia a la variabilidad de la población y el error de estimación nos indica la precisión con la que se desea estimar el parámetro.



**Figura 5.1** Tamaño de la muestra según el nivel de confianza.

Fuente: Luna, V. M., Alva, A., Bernal, A. y Arteaga, M. (2011). “Cuerpos de agua superficiales”. En F. Bautista (Ed.), “Técnicas de muestreo para manejadores de recursos naturales”.

### 5.1. Variable cualitativa y variable cuantitativa

Para seleccionar adecuadamente la fórmula para el cálculo del tamaño de la muestra también es necesario conocer el nivel de medición de la variable o el tipo de variable que se analizará. La correcta identificación del tipo de variable orienta la elección de la prueba estadística a utilizar.

Para establecer el tipo de variable usaremos una definición sencilla, si los datos que se obtienen al medir la variable son numéricos, la variable es cuantitativa. Cuando los datos que se obtienen al medir la variable no son numéricos sino categorías, la variable es cualitativa.



**Tabla 5.1** Posibles objetivos estadísticos según el tipo de variable.

Variable	Población	Objetivo
Cualitativa	1	Estimar la proporción.
	Más de una	Estimar la diferencia de proporciones. Establecer la asociación.
Cuantitativa	1	Estimar una media.
	Más de una	Estimar la diferencia de medias. Establecer la correlación.
		Establecer la regresión.

En la tabla 5.1 se observa que cuando la variable es cuantitativa se puede estimar el parámetro media, pero si la variable es cualitativa no se puede calcular dicho parámetro puesto que no hay números que promediar.

También se debe tener en consideración que todas las fórmulas mostradas a continuación se basan en un muestreo probabilístico y se asume que los datos recolectados se comportan aproximadamente de acuerdo con la distribución normal.

## 5.2. Población finita y población infinita

### 5.2.1. Población finita

Zappino (2020) define a la población finita como aquella población cuyo número de elementos es numerable. También se puede decir que en una población finita se puede conocer la cantidad de elementos que componen dicha población.

Por ejemplo:

- Escuelas en el distrito de Los Olivos.

- Comerciantes formales del Mercado Central de Lima.
- Camas operativas del Hospital Regional de Cusco.
- Empleados del Ministerio de Educación del Perú.
- Hospitales de Ecuador.
- Entidades financieras en la ciudad de Arequipa.

### 5.2.2. Población infinita

Cuando la población es grande y se desconoce su tamaño se asume que la población es infinita. Gutiérrez y De La Vara (2008) también señalan que, en muchos casos, en los que se desconoce el tamaño de la población, las poblaciones se pueden suponer infinitas o grandes. Por ejemplo, una población que se puede considerar infinita es la compuesta por los pescados que llegan a un terminal pesquero de la ciudad de Lima.

Navarro (2011) señala que algunas poblaciones son lo suficientemente grandes para considerarlas infinitas y en otros casos las poblaciones son conceptualmente infinitas. Para aclarar la noción de conceptualmente infinito se toma como ejemplo los estudios experimentales donde la población se compone del conjunto de resultados del experimento. Luego, se sabe que ciertos experimentos pueden efectuarse una infinidad de veces, con esto se llega a la noción de una población conceptualmente infinita.

Por ejemplo:

- Producción de palillos de fósforo de una empresa.
- Vendedores ambulantes en el Perú.
- Consumidores de bebidas gaseosas en Lima Metropolitana.
- Mascarillas vendidas en la ciudad de Puno.
- Inquilinos extranjeros en la ciudad de Lima.

### 5.3. Tamaño de la muestra para variable cualitativa

#### 5.3.1. Tamaño de la muestra para estimar la proporción

Cochran W. G. (1980) nos indica la siguiente fórmula para calcular el tamaño de la muestra:

$$n = \frac{\frac{t^2 PQ}{d^2}}{1 + \frac{1}{N} \left( \frac{t^2 PQ}{d^2} - 1 \right)}$$

Donde:

$N$ : tamaño de la población.

$d$ : margen de error para una proporción  $p$  de una categoría determinada.

$t$ : valor asociado a la significancia  $\alpha$ .

$p$ : proporción poblacional que presenta cierta característica.

$q$ : proporción poblacional que no presenta la característica.

Existe otra expresión equivalente para calcular el tamaño de la muestra cuando se quiere estimar la proporción y tiene algunas variaciones que dependen del tipo de población, es decir, si la población es considerada finita o infinita.

En la fórmula mostrada podemos observar que la distribución de probabilidad que se utiliza no es la distribución normal ( $Z$ ) sino la distribución  $t$  de Student. Como se mencionó en el capítulo 2 existen otras distribuciones de probabilidad aparte de la distribución normal, sin embargo, en casi todos los casos siguientes para realizar el cálculo del tamaño muestral usaremos la distribución normal.

### 5.3.1.1. Para una población infinita

$$n = \frac{Z^2 pq}{e^2}$$

Donde:

$p$ : Proporción poblacional que presenta cierta característica.

$q$ : Proporción poblacional que no presenta la característica.

$e$ : Máximo error de estimación de  $p$ .

$Z$ : Valor relacionado al nivel de confianza.

**Nota:** En adelante podemos utilizar  $Z$  en lugar de  $Z_{\alpha/2}$  para simplificar las expresiones matemáticas. En otros textos se utiliza la expresión  $Z_{1-\alpha/2}$  para referirse al valor asociado al nivel de confianza. Sin embargo, en este texto no existe ningún inconveniente en usar los símbolos mostrados de manera indistinta.

Más adelante se mostrará, usando la tabla de distribución de probabilidad normal, que para un nivel de confianza de 95% se verifica:

$$Z_{\alpha/2} = -1.96$$

$$Z_{1-\alpha/2} = 1.96$$

Como se aprecia los valores son distintos para  $Z_{\alpha/2}$  y  $Z_{1-\alpha/2}$  pero solo por el signo. En las fórmulas que veremos en adelante el valor  $Z$  siempre aparece elevado al cuadrado. Por lo tanto, se verifica la equivalencia:

$$Z_{\alpha/2}^2 = Z_{1-\alpha/2}^2 = Z^2$$

## Descripción de los elementos de la fórmula

### Proporción p y q

El elemento p representa el tanto por ciento de la población que presenta la característica que se desea analizar. También se puede entender que p es la probabilidad de que un elemento de la población presente la característica que se desea analizar. El elemento q representa el tanto por ciento de la población que no presenta la característica que se desea analizar.

Se debe tener en cuenta que p y q están expresados en tanto por ciento. Entonces, se cumple:

$$p + q = 100\%$$

Si utilizamos la expresión decimal en lugar de la porcentual, se tiene que:

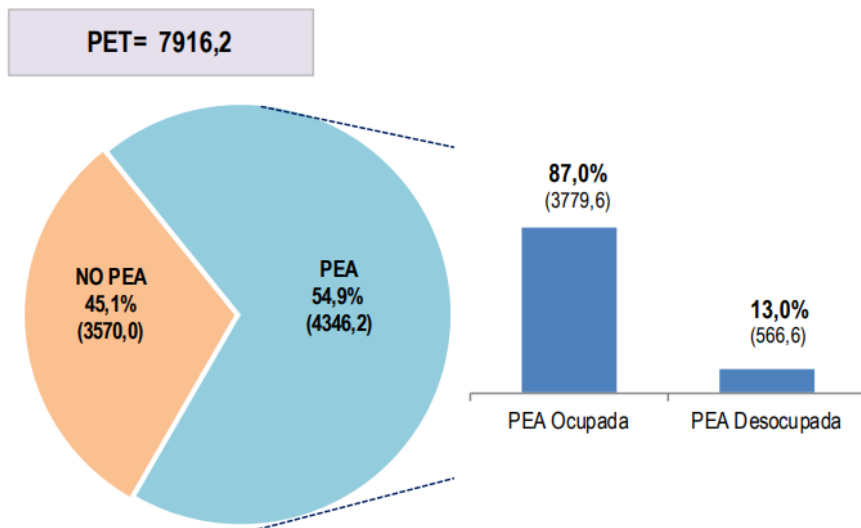
$$p + q = 1$$

Entonces q se puede expresar en función de p:

$$q = 1 - p$$

El valor de p se obtiene de investigaciones precedentes. Es decir, si se quiere analizar la prevalencia de una enfermedad en el año 2021, se pueden utilizar, como referencia, investigaciones anteriores sobre la prevalencia de dicha enfermedad en un contexto similar al contexto donde se realiza la investigación.

**Ejemplo:** Si queremos realizar una investigación sobre las características de los peruanos que trabajan actualmente, podemos utilizar la siguiente información:



PET= Población en Edad de Trabajar

PEA= Población Económicamente Activa

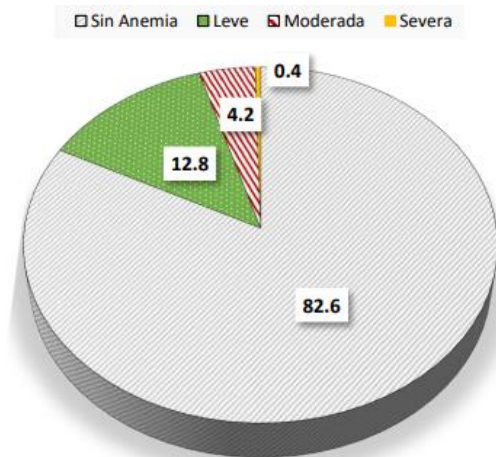
NO PEA= Población Económicamente No Activa

Fuente: Instituto Nacional de Estadística e Informática - Encuesta Permanente de Empleo.

**Figura 5.2** Lima Metropolitana: Población en edad de trabajar según condición de actividad, Anual: enero - diciembre 2020  
(Miles de personas y porcentaje)

De acuerdo con la figura 5.2,  $p=87\%$ , pues es el tanto por ciento de peruanos que trabaja según el informe del INEI y  $q=13\%$  es el tanto por ciento de peruanos que no trabaja.

**Ejemplo:** Si se desea realizar un estudio sobre la anemia en adultos en el Perú se puede utilizar la siguiente información brindada por el Instituto nacional de salud (INS):



**Figura 5.3** Nivel de anemia en adultos. Perú; 2017 – 2018.

Fuente: INS, (2021) “Estado nutricional en adultos de 18 a 59 años, Perú: 2017 – 2018”.

De acuerdo con la figura 5.3,  $p=17.4\%$ , pues es el tanto por ciento de peruanos que presentó anemia según el informe del INS y  $q=82.6\%$  es el tanto por ciento de peruanos que no presentó anemia según el informe.

### ¿Qué hacer cuando no se conoce $p$ ?

Cuando no se cuenta con información sobre  $p$ , se debe realizar una prueba piloto para estimar el valor buscado. Es decir, se debe realizar una encuesta previa con un número significativo de individuos. Martínez et al. (2006) consideran que parece contradictorio que, para estimar una proporción, se deba conocer dicha proporción previamente. Sin embargo, este paso es necesario para calcular el tamaño muestral adecuado. Por ello, señalan que se debe optar por apoyarse en investigaciones similares o realizar la prueba piloto con algunos integrantes de la población y obtener el valor de la proporción.

**Ejemplo:** Se quiere conocer las características el nivel de triglicéridos de los empleados de la empresa X. Dado que no se cuenta con información sobre el nivel de triglicéridos de los empleados en el año

pasado o en una investigación con una población similar, se opta por realizar una prueba piloto. Por ello, se realiza una prueba a un grupo de empleados de la empresa y se obtiene los siguientes resultados:

**Tabla 5.2** Nivel de triglicéridos en los empleados de la empresa, 2021.

Nivel de triglicéridos	Cantidad de empleados	%
Alto	38	34.5
Normal	72	65.5

De la tabla podemos indicar que:

$$p = 34.5\%$$

$$q = 65.5\%$$

Existe la opción de utilizar el valor 0.5 cuando no se conoce el valor de  $p$ . La justificación en el uso del valor  $p = 0.5$  se sustenta en que este número permite obtener el máximo valor de  $n$  (Martínez, 2012).

$$p = q = 0.5$$

Lwanga y Lemeshov (1991) también recomiendan el uso de  $p = 0.5$  y consideran que es la opción más adecuada para estimar la proporción dado que el tamaño muestral asumiendo ese valor es el más óptimo.

Al analizar la fórmula para el cálculo del tamaño de la muestra se puede apreciar que el tamaño de la muestra  $n$  es directamente proporcional al valor del producto  $pq$ . Es decir, el tamaño de la muestra será mayor cuando mayor sea el valor de  $pq$ .

$$n = \frac{Z^2 pq}{e^2}$$

Veamos que sucede cuando asignamos ciertos valores a  $p$  y a  $q$ .



**Tabla 5.3** Valor máximo del producto  $pq$ .

$p$	$q = 1 - p$	$pq$
0.1	0.9	0.09
0.2	0.8	0.16
0.3	0.7	0.21
0.4	0.6	0.24
0.5	0.5	0.25
0.6	0.4	0.24
0.7	0.3	0.21
0.8	0.2	0.16
0.9	0.1	0.09

Con los resultados de la tabla 5.3 se verifica que el mayor valor para  $pq$  se obtiene cuando  $p = q = 0.5$ . Podemos reemplazar estos valores en la expresión que define el tamaño de la muestra:

$$n = \frac{Z^2 \times 0.5 \times 0.5}{e^2}$$

Entonces, el tamaño de muestra se puede calcular utilizando la expresión:

$$n = \frac{0.25Z^2}{e^2}$$

También se puede expresar utilizando la siguiente expresión equivalente:

$$n = \frac{Z^2}{4e^2}$$

### Error de estimación de $p$

Si tenemos la proporción poblacional  $\pi$  y una proporción muestral  $p$ . Decimos que  $p$  es un estimador puntual de  $\pi$ , cuando  $p$  nos brinda un valor aproximado a  $\pi$ . Entonces, entre ambas proporciones existe una diferencia y a esta diferencia se le denomina error de estimación.

$$e = |p - \pi|$$

En los trabajos de investigación siempre se utilizan valores de error pequeños para que los resultados del estudio sean más significativos. Sin embargo, la reducción del error de estimación implica otros ajustes en el desarrollo de la investigación.

Cochran W. G. (1980) señala que la elección del error de estimación tolerable es algo arbitrario y depende del investigador y de la finalidad de los resultados.

Usualmente se trabaja con errores de 1% a 10%. En muchas investigaciones 5% representa un error tolerable bastante aceptable. Sin embargo, en otras investigaciones que requieren una mayor precisión, el error de 5% es demasiado alto.

Entre el error de estimación y el tamaño de la muestra existe una relación inversa. Es decir, cuanto menor sea el valor del error de estimación, mayor será el tamaño de la muestra.

¿Qué sucede con el tamaño de muestra cuándo  $e$  toma un valor pequeño?

Recordemos la fórmula del tamaño de muestra para una población infinita.

$$n = \frac{Z^2 pq}{e^2}$$

No podemos reemplazar directamente el valor  $e=0$  en la expresión porque no se puede efectuar dicha división. Sin embargo, podemos utilizar la noción de límite:

$$n = \lim_{e \rightarrow 0} \frac{Z^2 pq}{e^2}$$

El límite de la expresión cuando  $e$  toma un valor muy cercano a cero será:

$$n = \infty$$

Se verifica que, para eliminar el error de estimación, el tamaño de la muestra tiene que ser igual al tamaño de la población (infinita).

Para hacer más evidente este resultado podemos utilizar  $z=1.96$  y  $p = q = 0.5$ . Luego calculamos el tamaño de muestra para valores de  $e$  muy pequeños.

**Tabla 5.4** Relación entre los valores de  $e$  y  $n$ .

$e$	$n$
5%	385
4%	601
3%	1 068
2%	2 401
1%	9 604
0.5%	38 416
0.4%	60 025
0.3%	106 712
0.2%	240 100
0.1%	960 400

En la tabla 5.4 se observa como el tamaño de muestra crece a medida que se reduce el valor de  $e$ .

### El valor Z

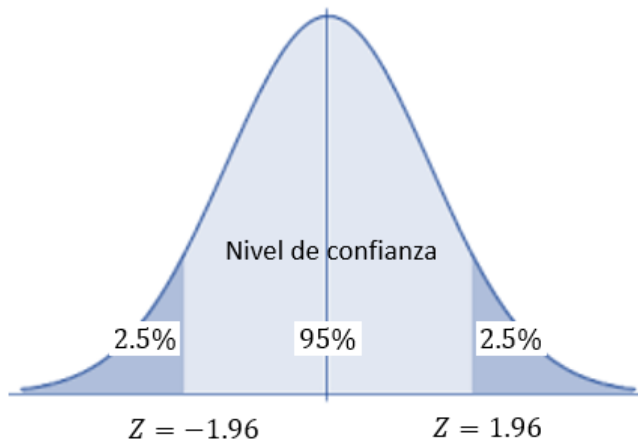
Para cada nivel de confianza existe un valor Z asociado y la relación entre estos valores se observa en la curva de distribución normal.

Estos valores  $Z$  se obtienen de la tabla de valores de probabilidad de distribución normal. Algunos de los niveles de confianza más usados se muestran en la siguiente tabla:

**Tabla 5.5** Valor  $Z$  según el nivel de confianza.

Nivel de confianza	Valor $Z$
90%	1.645
95%	1.96
99%	2.58

El nivel de confianza se complementa con la significancia  $\alpha$ . Por ejemplo, si el nivel de confianza es 95%, entonces la significancia es de 5%.



**Figura 5.4** Representación del nivel de confianza en la distribución normal.

En la figura 5.4 se puede ver que  $\alpha = 5\%$  se ha dividido en dos partes iguales a 2.5%. Esto se debe a que la prueba es bilateral, es decir, la gráfica presenta dos colas. Este tipo de pruebas será el que utilizaremos en todo el contenido del texto.

Por lo tanto,  $z = -1.96$  deja tras de sí una probabilidad de 0.025 y  $z = 1.96$  deja tras de sí una probabilidad de 0.975. Pero, delante de  $z = 1.96$  se observa la probabilidad 0.025.

En términos de probabilidad, la relación entre  $z$  y el nivel de confianza se puede expresar de la siguiente manera:

$$P[-1.96 < z < 1.96] = P[z < 1.96] - P[z < -1.96]$$

$$P[-1.96 < z < 1.96] = 0.975 - 0.025$$

$$P[-1.96 < z < 1.96] = 0.95$$

En términos sencillos podemos decir que existe un 95% de probabilidad de que el valor  $Z$  se encuentre en el intervalo de confianza  $[-1.96; 1.96]$ . Esta no es la definición formal, pero ayuda a comprender la relación entre intervalo de confianza, valor  $Z$  y probabilidad.

Los valores  $z$  asociados a cierto nivel de confianza  $z$  se pueden extraer de las tablas que aparecen en los anexos de los libros de estadística inferencial. Estos valores de probabilidad se han calculado usando la siguiente expresión que corresponde al área bajo una parte de la curva normal.

$$P[Z \leq z] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{x^2}{2}} dx$$

Pero para nuestros fines nos basta con entender el proceso para determinar los valores  $z$  a partir de la tabla. Debemos resaltar que todos los valores dentro de las filas y columnas que no están sombreadas representan probabilidades.

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767

**Tabla 5.5** Valores de probabilidad acumulativa para la distribución normal.

La tabla de la distribución normal de la figura nos permite ubicar los valores  $z$  para cada probabilidad. El 0.9750 de probabilidad se puede ubicar en la fila 1.9 y la columna 0.06 que finalmente se puede escribir como  $z = 1.9 + 0.06 = 1.96$ .

**Ejemplo:** Calcula el valor  $z$  para un nivel de confianza de 92%.

Solución:

$$92\% = 0.92$$

Para una prueba bilateral dividimos la significancia 8% entre 2 y resulta:

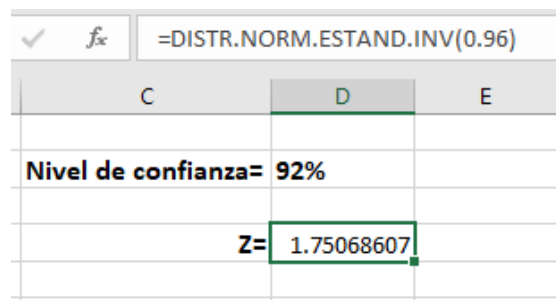
$$4\% = 0.04$$

Luego sumamos y buscamos el valor más cercano a dicha suma dentro de la tabla:

$$0.92 + 0.04 = 0.96$$

El valor más cercano es 0.9599 y se ubica en la fila 1.7 y la columna 0.05. Por lo tanto, el valor  $z$  que corresponde a un nivel de confianza de 92% es  $z = 1.75$ .

El cálculo del valor  $z$  también se puede realizar usando el programa Excel a través de la fórmula `DISTR.NORM.ESTAND.INV` (probabilidad).



	C	D	E
	Nivel de confianza= 92%		
		Z= 1.75068607	

**Figura 5.6** Cálculo de  $z$  usando Excel.

En la figura se puede observar que en el argumento de probabilidad se usa el valor 0.96, no el nivel de confianza 0.92.

### Ejemplos sobre cálculo del tamaño de muestra

**Ejemplo1:** Un investigador planifica realizar un estudio sobre el consumo de desayunos que se expenden en la vía pública en la ciudad de Lima. Un estudio anterior revela que el 27% de los limeños consume desayunos que se expenden en la vía pública debido a diferentes motivos. Estima el tamaño de la muestra que necesita el investigador. Utiliza un nivel de confianza de 95% y un error de 4%.

Solución:

$$p = 27\% = 0.27$$

$$q = 73\% = 0.73$$

$$z = 1.96$$

$$e = 4\% = 0.04$$

Reemplazamos los datos en la fórmula:

$$n = \frac{Z^2 pq}{e^2}$$
$$n = \frac{1.96^2 \times 0.27 \times 0.73}{0.04^2}$$
$$n = 473.237 \approx 474$$

Note que se ha redondeado al entero siguiente. Se opta por este criterio ya que siempre es mejor tener el mayor tamaño de muestra. Martínez et al. (2006) señalan que siempre se debe redondear hacia el entero superior para contar con el número suficiente de observaciones.

**Ejemplo 2:** Un estudiante de turismo decide realizar una investigación sobre turismo vivencial en diferentes distritos de Cusco, sin embargo, actualmente no cuenta con información referencial al respecto. Estima el tamaño de muestra adecuado para que el estudiante pueda realizar su investigación. Utiliza un nivel de confianza de 99% y un error de 5%

Solución:

$$p = 0.5 \text{ (usamos este valor porque se desconoce } p)$$

$$q = 0.5$$

$$z = 2.58$$

$$e = 5\% = 0.05$$

Reemplazamos los datos en la fórmula:

$$n = \frac{Z^2 pq}{e^2}$$
$$n = \frac{2.58^2 \times 0.5 \times 0.5}{0.05^2}$$



$$n = 665.64 \approx 666$$

El estudiante requiere una muestra de 666 personas para realizar su estudio.

**Ejemplo 3:** Un estudiante quiere realizar una encuesta para conocer el nivel de satisfacción de los estudiantes del nivel secundario de la ciudad de Lima con respecto al servicio de transporte público. Según los antecedentes de su investigación, el nivel de satisfacción es bajo, pero no hay un consenso con respecto a la proporción. Puesto que no conoce la proporción, realiza una prueba piloto y obtiene que el 60% de los encuestados tiene un nivel bajo de satisfacción con respecto al servicio de transporte público. Estima el tamaño de la muestra que se necesita para realizar la encuesta. Utiliza un nivel de confianza de 95% y un error de 3%.

Solución:

$$p = 60\% = 0.6$$

$$q = 40\% = 0.4$$

$$z = 1.96$$

$$e = 3\% = 0.03$$

Reemplazamos los datos en la fórmula:

$$n = \frac{Z^2 pq}{e^2}$$

$$n = \frac{1.96^2 \times 0.6 \times 0.4}{0.03^2}$$

$$n = 1024.43 \approx 1025$$

Entonces, la muestra para este caso se compone de 1025 estudiantes.

### 5.3.1.2. Para población finita

Para poder calcular el tamaño de la muestra para una población finita, es decir, cuando se conoce el tamaño de la población se puede utilizar la siguiente expresión ya conocida pero multiplicada por un factor de corrección.

$$n_0 = \frac{Z^2 pq}{e^2}$$

Multiplicamos  $n_0$  por el factor de corrección para población finita:  $\frac{N-n}{N-1}$

$$n = \frac{Z^2 pq}{e^2} \times \frac{N-n}{N-1}$$

$$n = \frac{NZ^2 pq - nZ^2 pq}{(N-1)e^2}$$

$$n(N-1)e^2 = NZ^2 pq - nZ^2 pq$$

$$n(N-1)e^2 + nZ^2 pq = NZ^2 pq$$

$$n[(N-1)e^2 + Z^2 pq] = NZ^2 pq$$

$$n = \frac{NZ^2 pq}{(N-1)e^2 + Z^2 pq}$$

Donde:

N: Tamaño de la población.

p: Proporción poblacional que presenta cierta característica.

q: Proporción poblacional que no presenta la característica.

e: Máximo error de estimación de la proporción poblacional.

Z: Valor relacionado al nivel de confianza.

## El tamaño de población $N$

La fórmula mostrada se utiliza cuando se puede precisar el tamaño de la población. Por ejemplo, los estudiantes de un centro educativo, los empleados de una empresa, las universidades de un país, etc.

**Ejemplo:** Según el blog Rankía al 2020, en el Perú existen 51 universidades públicas y 92 universidades privadas. Por lo tanto, si queremos analizar características de las universidades en el Perú, el tamaño de la población será:

$$N = 51 + 92$$

$$N = 143$$

**Ejemplo:** La Superintendencia Nacional de Administración Tributaria (SUNAT) informa que hasta enero del 2020 su personal se compone de 7172 personas. Por lo tanto, si queremos analizar características de la satisfacción laboral en el personal de la SUNAT, el tamaño de la población será:

$$N = 7172$$

**Ejemplo:** En el año 2020 el Colegio Nacional “X” cuenta con 80 estudiantes en el nivel inicial, 150 estudiantes en el nivel primaria y 164 estudiantes en el nivel secundaria. Por lo tanto, si queremos analizar el nivel de comprensión lectora entre los estudiantes de primaria y secundaria en dicho colegio en el año 2020, el tamaño de la población será:

$$N = 150 + 164$$

$$N = 314$$

### ¿Qué sucede cuando se trabaja con un error pequeño?

Si deseamos maximizar el tamaño de muestra, en la siguiente expresión podemos reducir el error:

$$n = \frac{NZ^2pq}{(N-1)e^2 + Z^2pq}$$

Eliminar el error implica que  $e = 0$ . Luego tenemos la siguiente expresión:

$$n = \frac{NZ^2pq}{(N-1)(0)^2 + Z^2pq}$$

$$n = \frac{NZ^2pq}{Z^2pq}$$

$$n = N$$

Este resultado es consecuente con la idea de obtener muestras numéricamente representativas. Si queremos eliminar el error, nuestro tamaño de muestra tiene que ser igual al tamaño de la población.

### Ejemplos sobre cálculo del tamaño de muestra

**Ejemplo 1:** En un centro poblado se planea realizar una encuesta para conocer qué tanto por ciento de pobladores está de acuerdo con la política de turismo implementada en la última década. Se sabe que el año pasado el 62% de los pobladores estaba de acuerdo con dicha política. Calcule el tamaño de muestra adecuado si la zona es habitada por 3450 personas. Utilice un nivel de confianza de 95% y la disposición de permitir un error de 3%.

Solución:

En primer lugar, identificamos los elementos:

$$N = 3450$$

$$p = 62\% = 0.62$$

$$q = 38\% = 0.38$$

$$e = 3\% = 0.03$$

$$Z = 1.96$$

Luego reemplazamos los datos en la fórmula:

$$n = \frac{3450 \times 1.96^2 \times 0.62 \times 0.38}{(3450 - 1)0.03^2 + 1.96^2 \times 0.62 \times 0.38}$$

$$n = 778.845$$

$$n = 779$$

Entonces, de los 3450 pobladores se puede seleccionar una muestra de 779 de ellos.

**Ejemplo 2:** El área de salud y prevención de una ciudad requiere conocer las lesiones oculares ocasionadas por el uso excesivo de la computadora en estudiantes de nivel secundario. Por ello, se debe seleccionar una muestra de estudiantes y realizar los exámenes respectivos. Calcule el tamaño de muestra adecuado si en dicha ciudad se ha registrado un total de 4930 estudiantes en el nivel secundario. Utilice un nivel de confianza de 99% y la disposición de permitir un error de 5%.

Solución:

Identificamos los elementos:

$$N = 4930$$

$$p = 0.5 \text{ (usamos este valor porque se desconoce } p)$$

$$q = 0.5$$

$$e = 5\% = 0.05$$

$$Z = 2.58$$

Luego reemplazamos los datos en la fórmula:

$$n = \frac{4930 \times 2.58^2 \times 0.5 \times 0.5}{(4930 - 1)0.05^2 + 2.58^2 \times 0.5 \times 0.5}$$

$$n = 586.562$$

$$n = 587$$

Entonces, de los 4930 estudiantes se puede seleccionar una muestra de 587 de ellos.

**Ejemplo 3:** Un estudiante de posgrado quiere realizar un estudio para conocer el nivel de estrés de los docentes de la universidad “X” durante el semestre 2021-1. El personal administrativo informa al estudiante que en el semestre 2021-1 laboraron 324 docentes, pero que no todos ellos trabajan actualmente en dicha universidad, por lo que será muy complicado encuestar a los 324 docentes. Calcule el tamaño de muestra adecuado usando un nivel de confianza de 95% y con disposición de permitir un error de 4%.

Solución:

$$N = 324$$

$$p = 0.5 \text{ (usamos este valor porque se desconoce } p)$$

$$q = 0.5$$

$$e = 4\% = 0.04$$

$$Z = 1.96$$

Luego reemplazamos los datos en la fórmula:

$$n = \frac{324 \times 1.96^2 \times 0.5 \times 0.5}{(324 - 1)0.04^2 + 1.96^2 \times 0.5 \times 0.5}$$

$$n = 210.65 \approx 211$$

Entonces, de los 324 docentes se puede seleccionar una muestra de 211 de ellos.

### 5.3.2. Tamaño de la muestra para la diferencia de proporciones

García et al. (2013) señalan que la siguiente fórmula se utiliza para el cálculo del tamaño de muestra cuando se cumplen los supuestos para la aplicación de la prueba Ji cuadrado con tablas de 2x2.

$$n = \left( \frac{Z_{\alpha/2} \sqrt{2P(1-P)} + Z_{\beta} \sqrt{P_1(1-P_1) + P_2(1-P_2)}}{P_1 - P_2} \right)^2$$

$P_1$ : Proporción en el primer grupo.

$P_2$ : Proporción en el segundo grupo.

$P$ : Proporción media.

$Z_{\alpha/2}$ : valor asociado al nivel de confianza.

$Z_{\beta}$ : Valor asociado al poder estadístico.

**Ejemplo 1:** En un estudio para determinar la asociación entre el uso de redes sociales y el nivel de ansiedad en adolescentes. Para determinar el tamaño de muestra se ha de utilizar una significancia de 5% y un poder estadístico de 80%. Estima el tamaño de muestra necesario para realizar dicha investigación. Además, según un estudio anterior, se estima que los jóvenes que usan con frecuencia las redes sociales presentan una probabilidad de 30% de presentar altos niveles de ansiedad en comparación con los jóvenes que acceden con menor frecuencia a las redes sociales quienes tiene una probabilidad de 10% de presentar altos niveles de ansiedad.

Solución:

$$P_1 = 30\% = 0.3$$

$$P_2 = 10\% = 0.1$$

$$P = \frac{0.3 + 0.1}{2} = 0.2$$

$$Z_{\alpha/2} = 1.96$$

$$Z_{\beta} = 0.842 \text{ (véase la tabla 5.10)}$$

Reemplazando los valores en la fórmula tenemos:

$$n = \left( \frac{1.96\sqrt{2 \times 0.2 \times 0.8} + 0.842\sqrt{0.3 \times 0.7 + 0.1 \times 0.9}}{0.3 - 0.1} \right)^2$$

$$n = 61.617$$

$$n \approx 62$$

Entonces la muestra se compone de un grupo de 62 adolescentes que usan con frecuencia las redes sociales y 62 adolescentes que no usan con frecuencia las redes sociales.

Martínez et al. (2006) sugiere la siguiente expresión para el cálculo del tamaño de la muestra en una comparación de proporciones.

$$n = 2P_M Q_M \left( \frac{Z_{\alpha/2} + Z_{\beta}}{P_A - P_B} \right)^2$$

Esta expresión se deriva de la fórmula para el tamaño de la muestra para la diferencia de medias.

Además:

$$P_M = \frac{P_A + P_B}{2}$$



- $P_A$ : Proporción en el primer grupo.  
 $P_B$ : Proporción en el segundo grupo.  
 $P_M$ : Proporción media.  
 $Z_{\alpha/2}$ : Valor asociado al nivel de confianza.  
 $Z_{\beta}$ : Valor asociado al poder estadístico.

### 5.3.3. Tamaño de la muestra para estudios de casos y controles

La siguiente fórmula es solo una adaptación de la fórmula para el cálculo del tamaño de muestra para la diferencia de proporciones. Los cambios realizados hacen referencia a la notación específica para el diseño de casos y controles.

$$n_c = n_e = \left( \frac{Z_{\alpha/2} \sqrt{2PQ} + Z_{\beta} \sqrt{P_c Q_c + P_e Q_e}}{P_e - P_c} \right)^2$$

- $n_c$ : cantidad de controles  
 $n_e$ : cantidad de casos  
 $P_c$ : Proporción favorable en el grupo de control.  
 $P_e$ : Proporción favorable en el grupo expuesto.  
 $P$ : Proporción media.  
 $Q_e$ :  $1 - P_e$ .  
 $Q_c$ :  $1 - P_c$ .  
 $Z_{\alpha/2}$ : valor asociado a la significancia estadística.  
 $Z_{\beta}$ : Valor asociado al poder estadístico.

Cuando se quiere utilizar una cantidad diferente de casos y controles. Es decir, si se decide utilizar  $n$  casos y  $m$  controles, el tamaño de muestra se calcula de la siguiente forma:

$$c = \frac{m}{n}$$

$$n = \frac{1}{c} \left( \frac{Z_{\alpha/2} \sqrt{(c+1)PQ} + Z_{\beta} \sqrt{cP_c Q_c + P_e Q_e}}{P_e - P_c} \right)^2$$

Soto y Cvetkovic (2020) en su artículo de investigación sobre estudios de casos y controles muestran una fórmula más sencilla para el cálculo del tamaño de muestra de casos y controles no pareados.

$$n_c = \frac{(P_1 Q_1 + P_2 Q_2)(z_{\alpha} + z_{\beta})^2}{(P_1 - P_2)^2}$$

Donde

$n_c$ : número de controles

$P_1$ : Proporción en el primer grupo.

$P_2$ : Proporción en el segundo grupo.

$Z_{\alpha/2}$ : Valor asociado al nivel de confianza.

$Z_{\beta}$ : Valor asociado al poder estadístico.

Soto y Cvetkovic (2020) también señalan que la eficiencia estadística aumenta cuando se incrementa la cantidad de controles. Por ello, cuando la cantidad de controles es diferente a la cantidad de casos ( $n_e$ ) se usa la siguiente expresión:

$$n_e = \frac{n_c(1+c)}{2c}$$

Donde  $c$  es el cociente entre el número de controles ( $m$ ) y el número de casos ( $n$ ):

$$c = \frac{m}{n}$$

## 5.4. Tamaño de la muestra para variable cuantitativa

### 5.4.1. Tamaño de la muestra para estimar la media

Para calcular el tamaño de la muestra cuando se requiere estimar la media, es decir, cuando la variable es cuantitativa, se utiliza una expresión similar a la fórmula para el cálculo del tamaño de la muestra para variable cualitativa. Martínez (2012) resalta que para el cálculo del tamaño de la muestra se consideran a la varianza, el nivel de confianza y la precisión de la estimación como los componentes.

#### 5.4.1.1. Para una población infinita

$$n = \frac{Z^2 \sigma^2}{e^2}$$

$\sigma$ : Desviación estándar poblacional.

$e$ : Limite aceptable de error de estimación.

$Z$ : Valor relacionado al nivel de confianza.

### Descripción de los elementos de la fórmula

En la fórmula mostrada hay elementos ya conocidos como el error de estimación ( $e$ ) y el valor relacionado con el nivel de confianza ( $Z$ ). La desviación estándar es un nuevo elemento que se explicará a continuación.

#### Desviación estándar poblacional $\sigma$

La desviación estándar es una medida de variabilidad, es decir, es un valor que nos indica que tan parecidos o diferentes pueden ser los elementos de una muestra o población con respecto a su media. El cuadrado de la desviación estándar poblacional se denomina varianza poblacional. Es decir,  $\sigma^2$  es la varianza poblacional.

Para entender mejor la definición de desviación estándar observe la siguiente tabla que contienen el registro de las edades de un grupo de niños:

**Tabla 5.6** Edades de los niños por grupo.

Grupo A	Grupo B
5	2
6	2
7	8
6	6
6	7

Para hacer la comparación entre los grupos calculamos las medias respectivas:

$$\bar{X}_A = \frac{5 + 6 + 7 + 6 + 6}{5} = 6$$

$$\bar{X}_B = \frac{2 + 2 + 8 + 6 + 7}{5} = 5$$

Luego, calculamos la desviación estándar para cada muestra ( $S$ ) usando la raíz cuadrada de la media cuadrática de los datos.

$$S = \sqrt{\frac{(x_1 - \bar{X})^2 + (x_2 - \bar{X})^2 + (x_3 - \bar{X})^2 + \dots + (x_n - \bar{X})^2}{n - 1}}$$

$$S_A = \sqrt{\frac{(5 - 6)^2 + (6 - 6)^2 + (7 - 6)^2 + (6 - 6)^2 + (6 - 6)^2}{5 - 1}} = 0.71$$

$$S_B = \sqrt{\frac{(2 - 5)^2 + (2 - 5)^2 + (8 - 5)^2 + (6 - 5)^2 + (7 - 5)^2}{5 - 1}} = 2.83$$

Observe que la desviación estándar del grupo “B” es mayor que la desviación estándar del grupo “A”. En otras palabras, el grupo “B” presenta mayor variabilidad que el grupo “A”. Esto se debe a que los elementos del grupo “B” son muy diferentes entre sí y con respecto a la media calculada para ese grupo, en cambio, los elementos del grupo “A” son muy cercanos entre sí o parecidos a la media del grupo respectivo.

En la tabla mostrada en el ejemplo se aprecia con facilidad que el grupo “B” presenta la mayor variabilidad sin necesidad de hacer ningún cálculo porque la muestra es pequeña. Sin embargo, la observación directa de la variabilidad se complica cuando la muestra es grande. Por ello es necesario realizar el cálculo mostrado, pero con apoyo de software estadístico.

La desviación estándar poblacional es un valor que se debe conocer de antemano con el apoyo de investigaciones anteriores.

**Ejemplo:** si se quiere realizar un estudio sobre el IMC de los peruanos, se puede utilizar la información presentada en el gráfico donde se aprecia que  $\sigma = 4.4 \text{ kg/m}^2$ .

**Tabla 5.7** Promedio y desviación estándar de las variables estudiadas de acuerdo al género en la población peruana.

	Total		Femenino		Masculino		p
	X	DE	X	DE	X	DE	
Edad (años)	46,2	16,8	46,1	16,7	46,4	17,1	0,16
Peso (kg)	64,1	12,5	60,8	11,8	68,1	12,2	<0,01
Talla (cm)	156,0	8,8	150,7	6,3	162,6	6,7	<0,01
IMC (kg/m <sup>2</sup> )	26,3	4,4	26,8	4,7	25,7	3,8	<0,01
CC (cm)	88,6	11,1	87,9	11,5	89,4	10,3	<0,01

IMC: índice de masa corporal. CC: circunferencia de cintura

Fuente: “El sobrepeso, la obesidad y la obesidad abdominal en la población adulta del Perú”. Anales de la Facultad de Medicina, 80(1), 21-27

### ¿Qué hacer cuando no se conoce $\sigma$ ?

Cuando no se conoce el valor de  $\sigma$ , se debe realizar una prueba piloto para estimar el valor buscado. Mason y Lind (1992) indican que se puede realizar una prueba piloto con 50 individuos, por ejemplo.

**Ejemplo:** Se quiere realizar un análisis del consumo semanal de pollo (en Kg) en el distrito de Magdalena. Por ello, se realiza una prueba piloto encuestando a 60 personas y con ayuda de un programa estadístico se obtiene la desviación estándar de la muestra:

**Tabla 5.8** Media y desviación estándar de la muestra.

n	Media	Desviación estándar
60	2.687	0.485

Esta desviación estándar de la muestra nos puede servir para estimar la desviación estándar poblacional.

En el siguiente gráfico se muestra la fórmula que se usa en el programa Excel para calcular la desviación estándar de los valores obtenidos en la prueba piloto.

Individuo	Consumo (Kg)
1	2
2	2.5
3	2
4	2.7
5	2.5
6	2
7	2
8	2.9
9	3
10	2

n=	60
Media=	2.687
Desv. Estándar=	0.485

**Figura 5.7** Cálculo de la desviación estándar muestral usando Excel.

Martínez (2012) señala que el tamaño de la muestra para la prueba piloto ( $n_{piloto}$ ) puede ser calculada como un porcentaje del tamaño de la población (N), pero no debe exceder al 10%. Si tenemos que  $N=800$  y elegimos 3%, entonces:

$$n_{piloto} = 3\%(800) = 24$$

Mason y Lind (1992) resaltan que cuando una población tiene poca variación entonces se requiere una muestra más pequeña. Esta afirmación se desprende de la misma fórmula:

$$n = \frac{Z^2 \sigma^2}{e^2}$$

Observe que  $n$  es directamente proporcional a la desviación estándar. Por lo tanto, si la población tiene una desviación estándar grande, el tamaño de la muestra necesario será grande, y si la población presenta una desviación estándar pequeña, no se requiere un tamaño de muestra grande.

### Ejemplos sobre cálculo del tamaño de muestra

**Ejemplo 1:** Mediante un estudio se desea conocer el consumo diario de calorías por parte de los estudiantes universitarios de la ciudad de Lima. En una investigación anterior se obtuvo que el consumo promedio de calorías en una población similar fue de 2980 calorías diarias con una desviación estándar de 420 calorías. Determine el tamaño de muestra necesario para realizar el estudio. Considere un error de 40 calorías y un nivel de confianza de 95%.

Solución:

$$\sigma = 420$$

$$z = 1.96$$

$$e = 40$$

Reemplazamos los datos en la fórmula:

$$n = \frac{Z^2 \sigma^2}{e^2}$$

$$n = \frac{1.96^2 \times 420^2}{40^2}$$

$$n = 423.536 \approx 424$$

**Ejemplo 2:** El administrador de una cadena de restaurantes desea conocer el consumo promedio (en S/) de los clientes a inicios del 2021. Para ello, planifica realizar un muestro aleatorio con base en el consumo promedio registrado el año anterior donde se obtuvo un consumo promedio de S/ 120. Además, la desviación estándar obtenida fue de S/ 30. Con esta información, determine el tamaño de muestra necesario considerando un error de S/ 4 y un nivel de confianza de 95%.

Solución:

$$\sigma = 30$$

$$z = 1.96$$

$$e = 4$$

Reemplazamos los datos en la fórmula:

$$n = \frac{Z^2 \sigma^2}{e^2}$$

$$n = \frac{1.96^2 \times 30^2}{4^2}$$

$$n = 216.09 \approx 217$$

**Ejemplo 3:** Un grupo de estudiantes realiza un estudio con la finalidad de conocer la cantidad de visitas que realizan los jóvenes a las diferentes salas de cines de la capital. No se cuenta con información previa y desea



utilizar un nivel de confianza de 99% y un error que represente el 10% de la desviación estándar. Determine el tamaño de muestra necesario para realizar el estudio.

Solución:

$\sigma$  se desconoce

$$z = 2.58$$

$$e = 0.1\sigma$$

Reemplazamos los datos en la fórmula:

$$n = \frac{Z^2 \sigma^2}{e^2}$$

$$n = \frac{2.58^2 \times \sigma^2}{(0.1\sigma)^2}$$

$$n = 665.64 \approx 666$$

#### 5.4.1.2. Para población finita

La siguiente expresión se obtiene multiplicando la fórmula para el cálculo de la población infinita por el factor de corrección  $\frac{N-n}{N-1}$ .

$$n = \frac{NZ^2\sigma^2}{(N-1)e^2 + Z^2\sigma^2}$$

N: Tamaño de la población.

$\sigma$ : Desviación estándar poblacional.

e: Limite aceptable de error muestral.

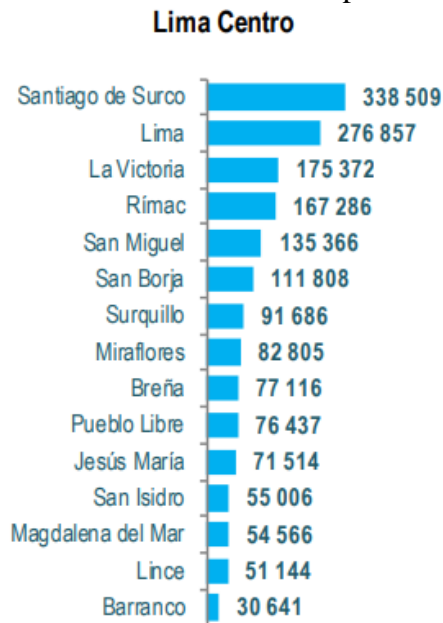
Z: Valor relacionado al nivel de confianza.

#### Ejemplos sobre cálculo del tamaño de muestra

**Ejemplo 1:** Mediante una investigación se quiere conocer el consumo mensual familiar promedio de bebidas gaseosas en los habitantes del distrito de Magdalena. Calcule el tamaño de la muestra necesaria para realizar dicha investigación utilizando un nivel de confianza del 95% y un error de 0.12 litros.

Solución:

En primer lugar, podemos utilizar la información brindada por el INEI sobre la población en la zona centro de la capital:



**Figura 5.8** Población en los distritos de Lima Centro

Fuente: Instituto Nacional de Estadística e Informática - Proyecciones de Población por Distritos, 2014

De la figura seleccionamos 54 566 como tamaño de la población y se estima que la desviación estándar es de 1.2 litros.

$$N = 54\,566$$

$$\sigma = 1.2$$

$$z = 1.96$$

$$e = 0.12$$

Reemplazamos los datos en la fórmula:

$$n = \frac{NZ^2\sigma^2}{(N-1)e^2 + Z^2\sigma^2}$$

$$n = \frac{54\,566 \times 1.96^2 \times 1.2^2}{(54\,566 - 1)0.12^2 + 1.96^2 \times 1.2^2}$$

$$n = 381.481 \approx 382$$

**Ejemplo 2:** En un estudio se desea conocer la satisfacción laboral de los trabajadores del Ministerio de Salud en la región Apurímac en el 2019. Calcule el tamaño de la muestra necesario para realizar dicha investigación utilizando un nivel de confianza del 95% y un error de 2 puntos. La satisfacción se midió con un test cuyo resultado varía de 20 a 100 puntos.

Solución:

En primer lugar, podemos utilizar la información brindada por el Ministerio de Salud:

**Tabla 5.9** Recursos Humanos del MINSa en la región Apurímac en el año 2018

Personal	Cantidad
Profesionales asistenciales	2528
Profesionales administrativos	220
Técnicos asistenciales	1459
Técnicos administrativos	442
Auxiliares asistenciales	85
Auxiliares administrativos	219
<b>Total</b>	<b>4953</b>

Fuente: Ministerio de Salud (2019)

De la tabla 5.9 seleccionamos 4 953 como tamaño de la población y para la desviación estándar se estimó 19 puntos.

$$N = 4\,953$$

$$\sigma = 19$$

$$z = 1.96$$

$$e = 2$$

Reemplazamos los datos en la fórmula:

$$n = \frac{NZ^2\sigma^2}{(N-1)e^2 + Z^2\sigma^2}$$
$$n = \frac{4\,953 \times 1.96^2 \times 19^2}{(4\,953 - 1)2^2 + 1.96^2 \times 19^2}$$
$$n = 324.08 \approx 325$$

#### 5.4.2. Tamaño de muestra para la diferencia de dos medias en grupos independientes

Existen investigaciones en la que se quiere comparar las medias de grupos diferentes. Los grupos en comparación son seleccionados de acuerdo a alguna característica de interés. En las investigaciones experimentales los grupos han sido sometidos a diferentes tratamientos. La finalidad de esta comparación es determinar si existe diferencia significativa entre las medias de los grupos seleccionados. Por ello se utiliza la siguiente expresión para el cálculo del tamaño de muestra para cada grupo:

$$n = (\sigma_1^2 + \sigma_2^2) \left( \frac{(Z_{\alpha/2} + Z_{\beta})}{D} \right)^2$$

$\sigma_1^2$  y  $\sigma_2^2$  son las varianzas de cada grupo. Pero si asumimos que los grupos tienen varianzas iguales se obtiene la siguiente expresión:

$$n = 2 \left( \frac{(Z_{\alpha/2} + Z_{\beta}) \sigma}{D} \right)^2$$

$n$ : tamaño de cada grupo.

$\sigma$ : desviación estándar.

$D$ : diferencia entre medias.

$Z_{\alpha/2}$ : valor asociado al nivel de confianza.

$Z_{\beta}$ : valor asociado al poder estadístico.

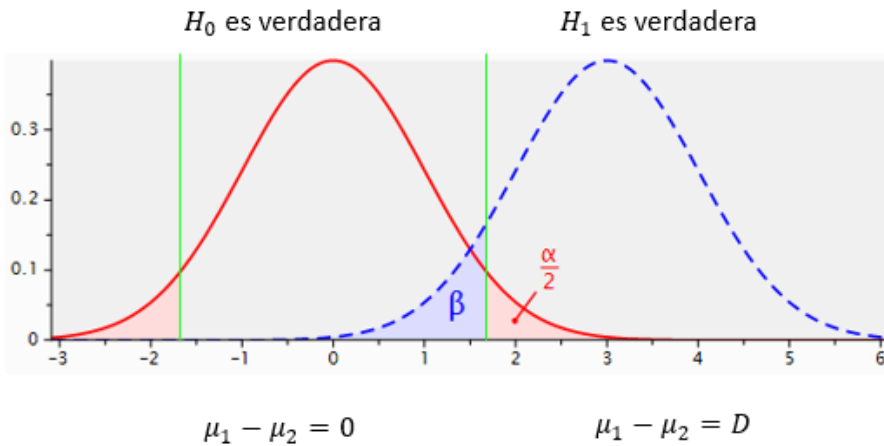
El poder estadístico  $(1-\beta)$  es la probabilidad de que la hipótesis alternativa sea aceptada cuando la hipótesis alternativa es verdadera.

Además:

$\alpha$ : probabilidad de rechazar  $H_0$  cuando  $H_0$  es verdadera.

$\beta$ : probabilidad de aceptar la hipótesis nula cuando esta es falsa.

En muchas investigaciones se usa  $\beta = 20\%$  y a este valor se le asocia  $Z_{\beta} = 0.842$ .



**Figura 5.9** Distribución de posibles resultados de la diferencia de medias según la veracidad de las hipótesis

Martínez et al. (2006) resaltan que el riesgo  $\beta$  siempre es unilateral en comparación con el riesgo  $\alpha$  que puede ser bilateral o unilateral. Note en el gráfico que si la diferencia de medias se ubica a la derecha de la línea de corte se rechaza  $H_0$ , en cambio, si la diferencia de medias se ubica a la izquierda de la línea de corte se acepta  $H_0$ .

Para hacer más sencillo el procedimiento para el cálculo del tamaño de muestra en este tipo de estudios podemos usar el cuadro elaborado por Camacho (2008) donde se establecen los valores  $Z_\alpha$  y  $Z_\beta$ .

**Tabla 5.10** Valores de  $Z_\alpha$  y  $Z_\beta$  para distintos niveles de significancia y potencia estadística

Significancia $\alpha$	Valor de $Z_\alpha$		Potencia $1 - \beta$	Valor de $Z_\beta$
	Unilateral	Bilateral		
			0.8	0.842
0.01	2.326	2.576	0.9	1.282
0.05	1.645	1.960	0.95	1.645
0.10	1.282	1.645	0.99	2.326

Fuente: Tamaño de muestra en estudios clínicos. Acta Médica Costarricense.

## La diferencia entre medias D

Pintado (2014) resalta que la diferencia entre medias es un valor que se establece usando como referencia investigaciones previas con características similares, de no contar con esa información el investigador debe establecer el valor más adecuado para esta diferencia basándose en su marco teórico.

Martínez et al. (2006) consideran que la diferencia a considerar debe ser relevante para establecer diferencias significativas. Por ejemplo, Si se utilizan dos metodologías distintas de enseñanza para analizar como esto afecta la calificación promedio de dos grupos de estudiantes con una escala de calificación de 0 a 20, no es muy práctico considerar una diferencia  $D=0.2$  puntos. En cambio, una diferencia  $D=2$  puntos nos indicarían que si hay una diferencia significativa entre las calificaciones promedio según la metodología utilizada en la enseñanza.

**Ejemplo 1:** En una investigación se desea conocer si existen diferencias significativas en el nivel de hemoglobina en dos grupos de niños. El primer grupo los niños están al cuidado directo de sus padres y en el segundo grupo los niños se encuentran al cuidado de otras personas designadas por los padres debido a que estos trabajan para solventar los gastos del hogar. Calcula el tamaño de muestra necesario para hacer la comparación utilizando un nivel de confianza de 95%, y un poder estadístico de 80%. Además, en investigaciones anteriores con grupos similares se obtuvo una desviación estándar de 1.06 g/dL y una diferencia de 1.1 g/dL entre los promedios de cada grupo.

Solución:

$$Z_{\alpha/2} = 1.96$$

$$Z_{\beta} = 0.842$$

$$\sigma = 1.06$$

$$D = 1.1$$

Reemplazamos los valores en la fórmula:

$$n = 2 \left( \frac{(1.96 + 0.842)1.06}{1.1} \right)^2$$

$$n = 14.581$$

$$n = 15$$

Finalmente, se entiende que se debe utilizar una muestra mínima de 15 niños de cada grupo.

**Ejemplo 2:** Para una investigación en la que se quiere conocer la efectividad de dos tipos de dietas se asume una diferencia mínima de 1.2 kg en la masa corporal promedio entre los grupos a analizar. Se realizará una comparación entre dos grupos seleccionados según su tipo de dieta. Además, por estudios anteriores se sabe que la desviación estándar de las masas corporales es de 4.5 kg. Calcula el tamaño de muestra adecuado asumiendo que los grupos tienen igual desviación estándar.

Solución:

Utilizaremos un nivel de confianza de 95%, y un poder estadístico de 80%

$$Z_{\alpha/2} = 1.96$$

$$Z_{\beta} = 0.842$$

$$\sigma = 4.5$$

$$D = 1.5$$

Reemplazamos los valores en la fórmula:



$$n = 2 \left( \frac{(1.96 + 0.842)4.5}{1.5} \right)^2$$

$$n = 141.322$$

$$n = 142$$

Entonces, se debe utilizar una muestra mínima de 142 personas de cada grupo.

### 5.4.3. Tamaño de muestra para la diferencia de dos medias en muestras relacionadas

En este caso la variable se mide en dos momentos diferentes, antes y después de cierto tratamiento. La finalidad de este proceso es determinar si existen diferencias significativas entre los datos obtenidos de manera individual antes y luego del tratamiento. Asumiendo que los grupos tienen varianzas iguales se obtiene la siguiente expresión:

$$n = \left( \frac{(Z_{\alpha/2} + Z_{\beta})\sigma}{D} \right)^2$$

$n$ : tamaño de cada grupo.

$\sigma$ : desviación estándar.

$D$ : media de las diferencias individuales.

$Z_{\alpha/2}$ : valor asociado al nivel de confianza.

$Z_{\beta}$ : valor asociado al poder estadístico.

**Ejemplo 1:** En una investigación se quiere verificar si cierto tratamiento permite la reducción de los niveles altos de glucosa en la sangre en un grupo de adultos. Por ello se realizan dos mediciones de glucosa, una antes del tratamiento y otra después de efectuado el tratamiento. De acuerdo con investigaciones anteriores se estima que el

promedio de las diferencias entre los niveles de glucosa en la sangre antes y después del tratamiento es de 15 mg/dl. Mediante un estudio piloto se sabe que la desviación estándar de ambos grupos es de 20 mg/dl aproximadamente. Calcula el tamaño de muestra adecuado.

Solución:

Utilizaremos un nivel de confianza de 95%, y un poder estadístico de 80%

$$Z_{\alpha/2} = 1.96$$

$$Z_{\beta} = 0.842$$

$$\sigma = 20$$

$$D = 15$$

Reemplazamos los valores en la fórmula:

$$n = \left( \frac{(1.96 + 0.842)20}{15} \right)^2$$

$$n = 13.958$$

$$n \approx 14$$

Entonces, se debe utilizar una muestra mínima de 14 personas.

#### 5.4.4. Tamaño de la muestra para asociar dos variables cuantitativas utilizando el coeficiente de correlación de Pearson

Hulley et al. (2013) nos proporcionan la expresión para poder calcular el tamaño de muestra mínimo para estimar la correlación entre dos variables analizadas en una misma población.

$$n = \left( \frac{z_{1-\frac{\alpha}{2}} + z_{1-\beta}}{\frac{1}{2} \ln\left(\frac{1+\rho}{1-\rho}\right)} \right)^2 + 3$$

$n$ : tamaño de la muestra.

$\rho$ : valor del coeficiente de correlación asumido para la población.

$Z_{\alpha/2}$ : valor asociado a la significancia estadística.

$Z_{\beta}$ : valor asociado al poder estadístico.

**Ejemplo 1:** Los directivos del centro educativo “San Nicolas” realizan una investigación entre sus estudiantes para conocer la correlación entre las horas de apoyo paterno durante las clases virtuales y la calificación obtenida en una prueba de aptitud diseñada por los mismos directivos. Estima el tamaño de muestra necesario para establecer la correlación usando un nivel de confianza de 95% y un poder estadístico de 80%. Además, asuma que el coeficiente de correlación de Pearson mínimo es de 0.5.

Solución:

$$Z_{\alpha} = 1.96$$

$$Z_{\beta} = 0.842$$

$$\rho = 0.5$$

Reemplazamos los valores en la fórmula:

$$n = \left( \frac{1.96 + 0.842}{\frac{1}{2} \ln\left(\frac{1+0.5}{1-0.5}\right)} \right)^2 + 3$$

$$n = 29.0200$$

$$n = 30$$

El valor calculado nos indica que se debe utilizar una muestra mínima de 30 estudiantes. Cabe resaltar que las variables apoyo paterno durante las clases virtuales y la calificación obtenida en la prueba de aptitud serán medidas en los 30 estudiantes.

**Ejemplo 2:** Un estudiante de posgrado ha decidido realizar una investigación para establecer la correlación entre el clima laboral y el nivel de estrés en los empleados de una empresa ¿Qué tamaño de muestra debe utilizar para el desarrollo de su investigación?

Solución:

Usando los valores convencionales para el nivel de confianza de (95%) y el poder estadístico (80%). Para el coeficiente de correlación de Pearson mínimo usaremos el valor 0.4.

$$Z_{\alpha} = 1.96$$

$$Z_{\beta} = 0.842$$

$$\rho = 0.4$$

Reemplazamos los valores en la fórmula:

$$n = \left( \frac{1.96 + 0.842}{\frac{1}{2} \ln\left(\frac{1 + 0.4}{1 - 0.4}\right)} \right)^2 + 3$$

$$n = 46.745$$

$$n = 47$$

El valor calculado nos indica que se debe utilizar una muestra mínima de 47 estudiantes.

## 5.5. Tamaño de la muestra para diseños experimentales

Para entender los diseños experimentales primero se debe conocer la definición de los elementos que se usan en el proceso de experimentación, la terminología. Por ello, a continuación, se definen de forma simple los principales elementos utilizados en la planificación de los experimentos factoriales y de diseño de bloques aleatorizados.

**Factores:** Son las variables que manipula el experimentador.

**Niveles:** Son los valores que se le asigna a cada factor. El nivel de intensidad (subcategoría distinta) de un factor es su nivel.

**Tabla 5.11** Ejemplos de niveles para cada factor.

Factor	Niveles
Cantidad de azúcar	10 g
	15 g
	20 g
Temperatura	12°C
	14°C
	16°C
Cantidad de agua	70 ml
	90 ml

**Tratamiento:** Es una combinación específica de los niveles de los factores.

Si tenemos un solo factor, cada nivel representa un tratamiento.

**Tabla 5.12** Tratamientos para un solo factor.

Consumo de kcal diarias	Tratamiento
2100	1
2300	2
2500	3

Si tenemos más de un factor, la combinación de niveles genera el número de tratamientos. En la tabla se observa que el factor Consumo de kcal/día tiene 2 niveles y el factor Horas de ejercicio diario también tiene 2 niveles. Por lo tanto, el número de tratamientos está dado por el producto  $2 \times 2 = 4$ .

**Tabla 5.13** Tratamientos para dos factores.

Factores		Tratamiento
Consumo de kcal/día	Horas de ejercicio diario	
2100	1	1
2100	2	2
2300	1	3
2300	2	4

**Réplica:** Se llama réplica a cada experimentación que considera a todos los tratamientos. Es decir, basándonos en el ejemplo de la tabla 5.13, si se ejecuta el experimento con cada uno de los 4 tratamientos, se ha realizado una réplica. Si luego se ejecuta nuevamente el experimento con cada tratamiento se tiene la segunda réplica.

Por ejemplo, si en un experimento se tienen dos factores, un factor  $A$  tiene  $a$  niveles, el otro factor  $B$  tiene  $b$  niveles y se realizan  $n_0$  réplicas, el número total de datos experimentales es:

$$n = a \times b \times n_0$$

Por ejemplo, si consideramos los siguientes factores y sus respectivos niveles:

**Tabla 5.14** Factores con dos niveles.

Factor	Niveles
Levadura (g)	7
	9
Fermentación (min)	40
	30

En la tabla se observa que hay dos factores cada uno con dos niveles y hay 4 tratamientos. Los niveles del factor *A* son 7 g y 9 g, los niveles del factor *B* son 40 min y 60 min, y se han realizado 4 réplicas. Por lo tanto, se verifica que el número total de datos experimentales es 16.

$$n = 2 \times 2 \times 4$$

$$n = 16$$

**Tabla 5.15** Cantidad total de datos experimentales.

<b>Factor A</b>	<b>Factor B</b>	<b>Efecto</b>
<b>Levadura (g)</b>	<b>Fermentación (min)</b>	<b>Altura (cm)</b>
7	40	3
7	40	3
7	40	3,2
7	40	3,5
7	60	4,9
7	60	4,5
7	60	5
7	60	4,8
9	40	3,2
9	40	3,2
9	40	4
9	40	3,5
9	60	4,8
9	60	4,2
9	60	3,5
9	60	4

Gutiérrez y De La Vara (2008) señalan que según la experiencia el número de réplicas en muchos experimentos con un solo factor varía entre 5 y 10, pero llegando en algunos casos hasta 30. Además, para orientarnos en cuanto al número de réplicas a utilizar describen algunas consideraciones:

**Tabla 5.16** Relación entre las características de la muestra y la cantidad de réplicas.

Características esperadas	Cantidad de réplicas
Diferencias pequeñas entre tratamientos.	Mayor
Gran variabilidad en cada tratamiento.	Mayor
Más de tres tratamientos.	Menor

Gutiérrez y De La Vara (2008) también muestran la fórmula para calcular el número de réplicas  $n$ , basada en la diferencia mínima significativa entre tratamientos:

$$n = \frac{2(t_{(0.025, k \times n_0 - k)})^2 \sigma^2}{(D_T)^2}$$

Donde:

$D_T$ : diferencia mínima significativa entre tratamientos.

$k$ : cantidad de tratamientos.

$n_0$ : propuesta inicial para el número de réplicas.

$\sigma$ : desviación estándar.

### Diseños balanceados y no balanceados

Un diseño es balanceado cuando cada tratamiento se aplica a igual número de unidades experimentales. Es decir, cuando el número de observaciones es igual para cada tratamiento.



**Tabla 5.17** Diseño balanceado.

Observación	Temperatura	Tiempo
1	30	20
2	30	20
3	35	15
4	35	15
5	30	15
6	30	15
7	35	20
8	35	20

Note que hay dos observaciones para cada tratamiento.

Un diseño es no balanceado cuando por lo menos un tratamiento presenta una cantidad diferente de datos experimentales en comparación con los demás. Es decir, cuando el número de observaciones no es igual para cada tratamiento.

**Tabla 5.18** Diseño no balanceado.

Observación	Temperatura	Tiempo
1	30	20
2	30	20
3	35	15
4	35	15
5	30	15
6	30	15
7	30	15
8	35	20
9	35	20
10	35	20

En la tabla se observa que los dos primeros tratamientos presentan dos observaciones cada uno. En cambio, los dos últimos tratamientos presentan 3 observaciones cada uno.

Las fórmulas para el cálculo del tamaño muestral que se han desarrollado anteriormente también son útiles para las investigaciones con diseño experimental. Por ejemplo:

**Tabla 5.19** Fórmulas que se pueden usar en los diseños experimentales.

Objetivo del experimento	Fórmula
Comparación de medias	Tamaño de muestra para la diferencia de medias en dos grupos independientes. Tamaño de muestra para la diferencia de medias en dos grupos relacionados.
Comparación de proporciones	Tamaño de muestra para la diferencia de proporciones en dos grupos.

### 5.5.1. Tamaño de muestra para una regresión lineal simple

En las investigaciones de alcance explicativo se trata de establecer la relación de causalidad entre dos variables o dos grupos de variables. Es decir, se analiza como los valores de la variable independiente determinan los valores de la variable dependiente.

De manera más formal mediante las técnicas de regresión de una variable  $y$  sobre una variable  $x$ , se busca una función que sea una buena aproximación al diagrama de dispersión generado por los valores de las variables. La siguiente expresión corresponde al modelo lineal donde  $a$  y  $b$  son constantes que se deben estimar y  $E$  es el error de estimación.

$$y = a + bx + E$$

Establecer el tamaño de muestra necesario para una regresión lineal simple significa un proceso de cálculo muy complejo que se aleja de los fines de este libro. Por ello, solo mostraremos parte del trabajo de Tianyuan Guan, M. Khorshed Alam y M. Bhaskara Rao (2019) quienes a través de una serie de algoritmos y simulaciones establecen el tamaño de muestra con base en la significancia, el tamaño del efecto, el poder

estadístico, la media del poder estadístico en la simulación y su respectiva desviación estándar.

**Tabla 5.20** Tamaño de muestra de acuerdo según los resultados de una simulación.

Significancia	Tamaño del efecto	Poder estadístico	n	Media	Desviación estándar
0.05	0.2	80%	199	0.797	0.0133
		90%	272	0.9039	0.0094
		95%	330	0.9497	0.0069
		99%	450	0.9891	0.0033
	0.3	80%	91	0.7978	0.0124
		90%	123	0.9028	0.0094
		95%	150	0.9505	0.0067
		99%	220	0.992	0.0028
	0.4	80%	53	0.7973	0.0128
		90%	70	0.8966	0.0096
		95%	87	0.9494	0.0071
		99%	121	0.9891	0.0034
	0.5	80%	36	0.8051	0.0124
		90%	48	0.9095	0.0091
		95%	58	0.95	0.0068
		99%	79	0.9888	0.0033

Fuente: Extracto de la tabla 3 del artículo “*Sample Size Calculations in Simple Linear Regression: Trials and Tribulations*” 2019.

En las referencias bibliográficas de este libro se comparte el enlace para que puedan acceder al artículo completo donde se encuentran las demás

tablas con los tamaños de muestra calculados para niveles de significancia de 0.1, 0.5 y 0.01.

### 5.5.1. Tamaño de muestra para un ANOVA

El análisis de la varianza es una prueba estadística que se utiliza para comparar medias entre más de dos grupos. Pero por la complejidad de la notación y los cálculos, optamos por mostrar una de las funciones del software Minitab que permite determinar el tamaño de muestra ingresando como parámetros la cantidad de niveles, la potencia estadística, la diferencia entre medias y la desviación estándar agrupada.

La ruta para el cálculo del tamaño de muestra para un ANOVA de un factor en Minitab es la siguiente: Estadísticas-Potencia y tamaño de la muestra-ANOVA de un solo factor. En general, Minitab también ofrece funciones para el cálculo del tamaño muestral para diseños factoriales (más de un factor) con dos niveles.

En la figura 5.10 se aprecia que para calcular el tamaño de muestra se ha considerado 4 niveles, la potencia estadística es la usual de 0.8 y los valores para la diferencia de medias y la desviación estándar son aproximados, dependen de trabajos anteriores o de una prueba piloto. La casilla para el tamaño de muestra se deja en blanco porque es el valor que se desea calcular. También se debe tener en cuenta el 0.05 de significancia, este valor se puede editar pulsando el botón de opciones.

**Figura 5.10** Cuadro de ingreso de datos para el cálculo del tamaño de muestra usando Minitab.

Luego de aceptar, el programa brinda una tabla con los resultados. En la figura 5.11 se aprecia el tamaño de muestra calculado igual a 12 elementos por cada nivel. Entonces, como se tiene un solo factor con 4 niveles, el total de observaciones será  $12 \times 4 = 48$ .

ANOVA de un solo factor  
 $\alpha = 0.05$  Desviación estándar asumida = 2.8  
 Factores: 1 Número de niveles: 4

**Resultados**

Diferencia máxima	Tamaño de la muestra	Potencia objetivo	Potencia real
4	12	0.8	0.811658

*El tamaño de la muestra es para cada nivel.*

**Figura 5.11** Cuadro de resultados para el cálculo del tamaño de muestra usando Minitab.

## 5.6. Diseño, objetivo, método estadístico y tamaño de la muestra

Toda la información mostrada sobre el cálculo del tamaño de la muestra, todavía no es suficiente para que el investigador logre la selección ideal del método de muestreo de acuerdo a las características de su investigación. Esto se debe a que los objetivos de cada investigación tienen diversas características de acuerdo a la problemática que se desea analizar.

Además, muchos objetivos de investigación se reducen, estadísticamente a estimar parámetros como la media y la proporción, realizar comparaciones entre grupos o verificar la relación entre variables. Pero esta simplificación es aparente. Por ejemplo, en el caso de la estimación de la media, se puede estimar la media de un grupo, se pueden comparar las medias de dos grupos o las medias de más de dos grupos. La situación se torna más compleja cuando se debe establecer si la comparación se realizará en grupos independientes o en grupos relacionados.

Otro aspecto a considerar son las pruebas estadísticas que se usan comúnmente para contratar hipótesis y su relación con el cálculo del tamaño de muestra. Entre estas pruebas tenemos dos grandes grupos: Las pruebas paramétricas, que requieren la verificación de una serie de condiciones, y las pruebas no paramétricas o de distribución libre, que no requiere la verificación de las condiciones anteriores.

**Tabla 5.21** Algunas pruebas estadísticas de uso frecuente.

Pruebas paramétricas	Pruebas no paramétricas
Prueba t para muestras independientes.	Prueba U de Mann-Whitney.
Prueba t para muestras relacionadas. ANOVA	Prueba de Wilcoxon.
ANOVA para muestras relacionadas.	Prueba de Kruskal-Wallis.
Coefficiente de correlación de Pearson.	Prueba de Friedman
	Coefficiente de correlación de Spearman.
	Prueba Ji cuadrado

La selección del método estadístico para la investigación depende de los objetivos y estos deben denotar claramente si se hará uso de una estimación de parámetro o una prueba de hipótesis. Por ejemplo:

**Tabla 5.21** Objetivo de la investigación y objetivo estadístico.

Objetivo	Inferencia estadística	Objetivo estadístico
Establecer las principales características del consumo de frutas en los empleados de la empresa X.	En algunas investigaciones descriptivas no se hace uso de la inferencia.	Interpretar tablas y gráficos estadísticos.
Identificar el consumo de frutas en los empleados de la empresa X.	Estimación de parámetro.	Estimar la media.
Comparar el consumo de frutas entre los empleados varones y mujeres de la empresa X.	Prueba de hipótesis.	Estimar la diferencia de medias.
Comparar el consumo de frutas entre los empleados del área de ventas, el área contable y el área de logística de la empresa X.	Prueba de hipótesis.	Estimar la diferencia de medias.
Establecer la correlación entre la masa corporal y el consumo de frutas en los empleados de la empresa X.	Prueba de hipótesis.	Determinar el coeficiente de correlación.
Explicar la influencia del consumo de frutas sobre la masa corporal en los empleados de la empresa X.	Prueba de hipótesis.	Analizar la regresión.

En la tabla 5.21 se observa la diferencia entre el objetivo de la investigación y el objetivo estadístico. El objetivo estadístico permite seleccionar la prueba adecuada para alcanzar el objetivo de la investigación. Por ejemplo, si el objetivo estadístico es estimar la diferencia de medias, entonces se podría utilizar la prueba t de Student siempre que la cantidad de medias a comparar sea igual a dos y se cumplan las condiciones para las pruebas paramétricas.

Debido a las razones expuestas, las siguientes orientaciones son solo algunas de las que el investigador se puede servir para reconocer y seleccionar el tamaño de muestra adecuado. Se resalta que no se abarcan todos los tipos de diseño ni todas las pruebas estadísticas que se utilizan en investigación, lo que se muestra son solo ejemplos.

**Tabla 5.22** Diseño de investigación y objetivo estadístico.

Tipo	Característica	Objetivo estadístico
Descriptivos	Transversal	Estimar la media.
		Estimar la proporción.
		Estimar la diferencia de medias.
		Estimar la diferencia de proporciones.
	Longitudinal	Estimar la diferencia de medias.
		Estimar la diferencia de proporciones.
Analíticos	Cohortes	Comparar las proporciones.
	Casos y controles	Comparar las proporciones.
	Correlacional	Estimar el coeficiente de correlación.
	Correlacional causal	Analizar la regresión

Los objetivos estadísticos mostrados en la tabla 5.22 se han utilizado a lo largo del desarrollo del capítulo 5. Por tanto, el lector puede guiarse de estos objetivos para seleccionar la fórmula para el cálculo del tamaño muestral.



## Capítulo 6: Preguntas frecuentes sobre la muestra

### ¿Se debe incluir el proceso del cálculo del tamaño de muestra en la investigación o tesis?

El procedimiento realizado por el investigador para calcular el tamaño de muestra, generalmente no se muestra en el informe de investigación. Pero dicho proceso a utilizar para el muestreo aparece mencionado dentro de los protocolos de investigación y los cálculos se pueden mostrar en los anexos de la tesis. García et al. (2013) indican que el tamaño de la muestra se brinda como un valor determinado y se sobreentiende que el cálculo para obtener dicho tamaño de muestra se efectuó con estricto orden metodológico. Aquí se resalta la ética del investigador para ceñirse objetivamente al método estadístico que permita el cálculo del tamaño de muestra a pesar de las dificultades que puedan presentarse en el proceso.

Cárdenas y Arancibia (2016) resaltan que en la actualidad muchas revistas científicas en psicología publican artículos en los cuales no muestran procedimientos importantes como el tamaño del efecto y el cálculo del tamaño de la muestra. También explican dos razones fundamentales para la omisión de los procesos señalados: la ausencia de paquetes estadísticos y la falta de exigencias editoriales.

Las posibles causas presentadas para la omisión del proceso de cálculo del tamaño del efecto y el tamaño de la muestra, en la actualidad, se reducen solo a la falta de exigencias editoriales, puesto que ahora se cuenta con una gran cantidad de paquetes estadísticos para realizar

dichos cálculos. Cabe resaltar que no es indispensable que el investigador comprenda dichos cálculos de modo riguroso, sino que es más importante que comprenda el significado de los resultados obtenidos. Por ello, se requiere el asesoramiento de especialistas del área estadística.

### ¿El tamaño de la muestra puede ser igual al tamaño de la población?

El uso de muestras en la investigación se justifica cuando no se puede acceder a la totalidad de los elementos de la población. Pero cuando se tiene pleno acceso a toda la población en el proceso de recolección de datos no es necesario el uso de muestras y el proceso se denomina censo.

Además, cuando los datos se obtienen de toda la población, sólo se requiere establecer conclusiones a partir de la estadística descriptiva (Medenhall et al., 2010). En este tipo de recolección de datos no es necesario hacer uso de estadística inferencial. Es decir, no se debe estimar el valor de parámetros como la media poblacional o la proporción poblacional puesto que ya se han considerado a todos los elementos de la población.

Por ejemplo, en una investigación sobre clima institucional en una clínica que cuenta con 120 personas entre médicos, enfermeros, personal administrativo y de mantenimiento se organiza una encuesta en coordinación con los directivos de la clínica de modo que todo el personal esté obligado de responder la encuesta. En este caso la población es igual a la muestra ya que se encuestará a la totalidad de personas que laboran en dicha clínica.

## ¿Qué sucede cuando en una muestra se quieren analizar características cuantitativas y cualitativas?

Este caso es bastante común ya que, generalmente, cuando se elabora un instrumento de recolección de datos se trata de que este permita recolectar la mayor cantidad de información sobre un determinado hecho de interés. Por ejemplo, cuando se quiere conocer los hábitos alimentarios de cierta población se elaboran preguntas como las siguientes:

**Tabla 6.1** Preguntas para variables cuantitativas y cualitativas.

Ítem	Tipo de característica
¿Qué cantidad de vegetales consumes por semana?	Cuantitativa
¿Cuántos vasos con agua consumes por día?	
¿Cuántas frutas consumes al día?	
¿Qué tipo de comida consumes con mayor frecuencia?	Cualitativa
¿Qué tipo de snacks consumes con mayor frecuencia?	

Como se puede apreciar en la tabla muchas encuestas tienen como objetivo la recolección de información de diferentes características en una misma muestra y estas características pueden ser de naturaleza cualitativa o cuantitativa. Al respecto, Cochran (1980) señala que en este caso se pueden calcular los tamaños de muestra para cada característica, si los tamaños de muestra son similares se opta por el mayor valor cuando se cuenta con los recursos necesarios. Si los tamaños de muestra calculados para cada característica son diferentes puede utilizarse un tamaño de muestra más pequeño. Sin embargo, cuando los tamaños de muestra para cada característica son muy diferenciados se puede pensar en excluir dicha característica del instrumento de investigación.

## ¿Qué hacer si no se puede completar el tamaño de la muestra?

En ocasiones sucede que siguiendo el procedimiento para el cálculo del tamaño de la muestra de manera estricta se obtiene una cantidad de elementos cuyo número es difícil de completar. Es decir, el investigador sabe que no puede completar el tamaño de muestra que ha calculado por falta de recursos o simplemente porque es muy difícil el acceso a dicha muestra. En estos casos Martínez et al, (2006) señalan que la mejor decisión es no realizar dicha investigación.

La decisión de no seguir con la investigación por causa de la escasa posibilidad de completar el tamaño de muestra es lógica. Sin embargo, los estudiantes que necesitan terminar su investigación para obtener su grado o título académico no encontrarán mucha satisfacción en esta respuesta. Por ello, a continuación, se brindan algunas opciones a considerar cuando nuestra investigación peligra por motivo del tamaño muestral incompleto.

Para obtener un tamaño de muestra más pequeño se puede optar por reducir el alcance de la investigación. Es decir, que la investigación ya no se oriente a una población grande, sino a una población más específica.

Pintado (2014) señala que cuando no se alcanza el tamaño muestral requerido no es necesario abandonar la investigación, sino que se debe reformular el tipo de conclusiones que se obtienen con el tamaño de muestra conseguido.

Demey et al. (2004) resalta que cuando no es posible aumentar el tamaño de la muestra se puede reconsiderar la precisión de las estimaciones dentro de intervalos de confianza para obtener conclusiones válidas sobre los experimentos. Por ejemplo, al utilizar una muestra más pequeña que la requerida lo que se obtiene es una menor precisión en las conclusiones, si establecemos esta precisión en intervalos de confianza se puede llegar a resultados como los siguientes.

**Tabla 6.2** Valores del error según el tamaño de la muestra.

Tamaño de la muestra	Intervalo de confianza para la media	Error de estimación (máx)
50	[65.50; 68.50]	1.50
40	[65.33; 68.67]	1.67
35	[65.21; 68.79]	1.79
30	[65.07; 68.93]	1.93
25	[64.88; 69.12]	2.12

En la tabla 6.2 se ha analizado la masa corporal de un grupo de personas. Los intervalos de confianza para la media han sido calculados considerando un nivel de confianza de 95%, una desviación estándar de 5.4 kg y una media muestral de 67 kg. Observe que a medida que se reduce el tamaño de la muestra, el intervalo de confianza tiene mayor amplitud. Un intervalo de amplitud pequeña brinda una mayor precisión y esto también se evidencia observando el error de estimación máximo para la media.

De todo lo anterior se debe concluir que, si el investigador decide proseguir con el estudio a pesar de no completar el tamaño muestral, debe señalar en su informe que sus conclusiones están limitadas y que existe un nuevo margen de error de tamaño considerable.

### ¿Qué dificultades se deben considerar en experimentos con seres vivos?

Existe una gran diferencia entre la experimentación con materiales inanimados y la experimentación con seres vivos. Por ejemplo, en una investigación sobre resistencia de materiales pueden presentarse algunas incidencias durante el experimento de modo que dicho experimento deba realizarse otra vez. Esto no genera un inconveniente mayor puesto que muchas incidencias están contempladas en el plan de contingencia, a lo más se tendrá que invertir recursos extra. En cambio, en un experimento sobre el efecto de un tipo de sustancia en focas, la situación se complica cuando suceden algunas incidencias porque

realizar nuevamente el experimento requiere el uso de más animales y estos son considerados como recursos escasos.

Complementando la idea del párrafo anterior se debe tener una planificación del proceso experimental de tal modo que se prevean, controlen y, de ser posible, se eviten cualquier tipo de contingencias que impliquen el aumento en la cantidad de réplicas del experimento. El fin ideal de esta planificación es que se alcancen los objetivos trazados utilizando el menor número de animales.

Demey et al. (2004) resaltan que en muchas investigaciones en biotecnología una de las principales limitaciones es la escasa disponibilidad del material experimental. Utilizar un tamaño de muestra pequeño genera el riesgo de que se puedan establecer conclusiones o generalizaciones erróneas, pero estos inconvenientes se pueden reducir con un adecuado proceso de muestreo.

Rojo (2014) señala que la experimentación con animales requiere un esfuerzo mayor dado que los procesos se tienen que regir por el código de ética a nivel local e internacional. Esto implica que el trabajo del investigador experimental está sujeto a reglas muy estrictas con respecto al tratamiento de seres vivos. Por ello, se debe tener en consideración los protocolos establecidos por las instituciones respectivas para la experimentación de este tipo.

### ¿Se deben considerar muestras de reserva?

Durante la etapa de recolección de datos pueden presentarse una serie de errores en los resultados obtenidos. Estos errores pueden ser resultado de diferentes situaciones que muchas veces escapan del manejo del investigador. Por ejemplo, en una encuesta puede suceder que un encuestado no responda a una pregunta en especial o responda incoherencias.

Por ello es muy acertado contar con más elementos muestrales aparte de los que se han considerado según el cálculo del tamaño muestral.

Pintado (2014) resalta que en los ensayos clínicos se suele disponer de una cantidad extra de individuos con respecto a los elementos de la muestra. Esta cantidad extra de individuos representa el 10% de cada grupo. Por ejemplo, si el tamaño muestral calculado es de 90 elementos se debe considerar la inclusión de  $10\% \times 90 = 9$  elementos más para reemplazar en caso de haber respuestas defectuosas.

Rojo (2014) menciona que en el Comité de Ética en Experimentación Animal del Vall d'Hebron Institut de Recerca se utiliza la siguiente fórmula, basada en la experiencia profesional de los integrantes de la institución, para establecer la cantidad de elementos necesarios para la muestra cuando se tiene en cuenta el desarrollo de incidencias que merman la cantidad de unidades experimentales durante el proceso mismo de experimentación.

$$X = \frac{n}{A \times B \times C \times \dots}$$

Esta es la fórmula para el cálculo de la muestra en base a incidencias seriadas en los procedimientos.

Donde:

$X$ : número total de animales necesarios

$n$ : número mínimo de animales calculado de modo estadístico.

$A = 1 - \%$  de incidencia 1.

$B = 1 - \%$  de incidencia 2.

$C = 1 - \%$  de incidencia 3.

Por ejemplo, si para un experimento con conejos se ha establecido un tamaño muestral mínimo de 12 animales. Además, se estima que hay un 20% de probabilidad de que los conejos incrementen su masa corporal de forma desigual durante el experimento y se estima una

probabilidad de 10% de que los conejos cambien sus hábitos de alimentación en el transcurso de dicho experimento.

Con la información anterior podemos establecer los siguientes valores:

$n$ : 12

$$A = 1 - 20\% = 80\%$$

$$B = 1 - 10\% = 90\%$$

Reemplazando estos valores en la fórmula se establece la nueva cantidad de elementos muestrales:

$$X = \frac{12}{80\% \times 90\%}$$

$$X = 16.667$$

$$X \approx 17$$

Entonces, son 17 los conejos que deben ser parte del experimento.

### ¿Cómo se realiza el muestreo en alimentos?

El muestreo en alimentos se rige por diferentes normas, estándares y protocolos establecidos por instituciones internacionales, regionales y a nivel local. Las instituciones establecen un plan de muestreo idóneo que permita el tratamiento homogéneo en la extracción de muestras en el sector alimentario.

A nivel internacional existe el Programa Conjunto FAO/OMS sobre Normas Alimentarias con la finalidad de proteger la salud de los consumidores y promover buenas prácticas en el comercio de alimentos. Este programa brinda a la comunidad internacional el Codex Alimentarius, o Código Alimentario, que es un conjunto de normas, y códigos de prácticas aprobados por una comisión especial.



En la tabla 6.3 se muestra solo una parte de una de las directrices del Código Alimentario denominada “Métodos recomendados de muestreo para la determinación de residuos de plaguicidas”. En esta tabla se sugiere la cantidad mínima de producto para ser analizada en el laboratorio.

**Tabla 6.3** Muestra para productos alimenticios primarios de origen animal.

Despojos de mamíferos	Naturaleza de las muestras	Tamaño mínimo de cada muestra de laboratorio
Hígado	Hígado o hígados enteros, o parte de hígado.	0.4 kg
Riñón	1 o ambos riñones de uno o más animales	0.2 kg
Corazón	Corazón o corazones enteros, o sólo porción del ventrículo, si éste es grande.	0.4 kg
Otros	Parte o unidad entera de uno o más animales, o sección transversal tomada del producto congelado a granel	0.5 kg

Los productos pueden ser frescos, refrigerados o congelados.

Fuente: FAO Codex Alimentarius 1999

Sin embargo, el manejo de los alimentos es una empresa compleja y la finalidad del muestreo determina el uso de diferentes orientaciones. Por ejemplo, la finalidad del muestreo puede ser el control de calidad del producto, la medición del grado de aceptación del producto, establecer la durabilidad del producto o analizar las características de un nuevo producto alimentario.

También existen normas nacionales como la Norma Técnica Peruana 700.002-2012 “Lineamientos y procedimientos de muestreo del pescado y productos pesqueros para inspección” donde se establecen orientaciones para el establecimiento de planes de muestreo y anexos

en los que ya se establece el tamaño de muestra requerido bajo ciertas condiciones.

A continuación, se muestra algunos de los anexos de la norma mencionada donde se sugiere el tamaño de la muestra de acuerdo al nivel de inspección:

**Tabla 6.4** Tamaño de la muestra de acuerdo al nivel de inspección cuando el peso neto del producto es igual o menor a 1 kg.

Tamaño del lote (N)	Tamaño de la muestra (n)	
	Nivel de inspección I	Nivel de inspección II
4 800 o menos	6	13
4 801 - 24 000	13	21
24 001 – 48 000	21	29
48 001 – 84 000	29	48
84 001 – 144 000	48	84
144 001 – 240 000	84	126
Más de 240 000	126	200

Fuente: Norma Técnica Peruana 700.002-2012

El nivel de inspección depende de las condiciones del lote a analizar. Cuando la calidad del lote no está cuestionada como en las inspecciones iniciales se utiliza el nivel de inspección I. El nivel II de la inspección se utiliza cuando la calidad del producto está cuestionada y se requiere realizar una nueva inspección del lote.

En el caso de la evaluación de las características de un nuevo producto alimentario se recurre a la evaluación sensorial, donde un grupo de personas o jueces degustan y expresan su opinión respecto a las propiedades del producto. Surco y Alvarado (2011) resaltan que la evaluación sensorial permite encontrar atributos valiosos para los consumidores, por ello este tipo de procedimiento es de uso generalizado en la industria alimentaria. En cuanto a la cantidad de jueces, Catania y Avagnina (2007) señalan que esta varía de acuerdo con el tipo de prueba sensorial a desarrollar.

Debemos establecer que en este tipo de análisis el tamaño de la muestra no lo conforman la cantidad de jueces porque de acuerdo a la definición de muestra, esta es la agrupación de elementos donde se medirá la variable de interés. Mediante la evaluación sensorial se analiza el producto usando como instrumento la opinión de los jueces. Sin embargo, podemos afirmar que la cantidad de jueces si determina la cantidad de muestras del producto a evaluar. Por ejemplo, para evaluar la dulzura de un tipo de bizcocho se han preparado muestras utilizando diferentes cantidades de Stevia (4.5 g, 4.9 g y 5.3 g) y la degustación se realiza con 10 jueces. Entonces:

**Tabla 6.5** Cantidad de muestras de acuerdo a los tipos de bizcocho y la cantidad de jueces.

Cantidad de jueces	Cantidad de Stevia	Tipos de bizcocho	Total de muestras a degustar
10	4.5 g	3	30
	4.9 g		
	5.3 g		

### ¿Cómo considerar el tamaño del efecto en el cálculo del tamaño de la muestra?

En el capítulo 2 se mencionó el significado del tamaño del efecto, sin embargo, en este capítulo se hará referencia a su importancia en el cálculo del tamaño de la muestra. Las pruebas de significancia permiten determinar si existen diferencias significativas entre las características que se analizan, pero no siempre nos indican la magnitud de esta diferencia. Cárdenas y Arancibia (2016) resaltan que los valores elevados en las pruebas de significancia sólo indican probabilidades altas de que las diferencias se deban al azar.

En una prueba de hipótesis es difícil encontrar que la diferencia sea exactamente cero, por tanto, siempre se encuentran diferencias. En ese

sentido, tal y como se plantean las hipótesis, es común que se rechace la hipótesis nula debido a la escasa probabilidad de encontrar  $d=0$ .

El tamaño del efecto se puede entender como la mínima diferencia que el investigador está dispuesto a aceptar como relevante (Prajapati, Dunne y Armstrong, 2010, como se citó en Cárdenas y Arancibia, 2016). También se puede precisar que el tamaño del efecto nos indica cuánto de la variable dependiente se puede determinar por la variable independiente (Snyder & Lawson, 1993, como se citó en Cárdenas y Arancibia, 2016).

Sobre el tamaño del efecto Pardo y San Martín (2010) señalan que el investigador puede elegir un tamaño del efecto pequeño, medio o grande. Aunque el investigador no sepa con exactitud el tamaño del efecto que es adecuado para su investigación existe información que permite seleccionar los valores para el cálculo del tamaño de la muestra en función del tamaño del efecto.

Pardo y San Martín también nos brindan la forma de interpretar el tamaño del efecto, de modo que, los valores en torno a 0,20 indican un efecto pequeño; los valores en torno a 0,50, un efecto medio; valores en torno a 0,80 y mayores, indican un efecto grande. Sin embargo, estas interpretaciones varían de acuerdo a la prueba estadística a utilizar.

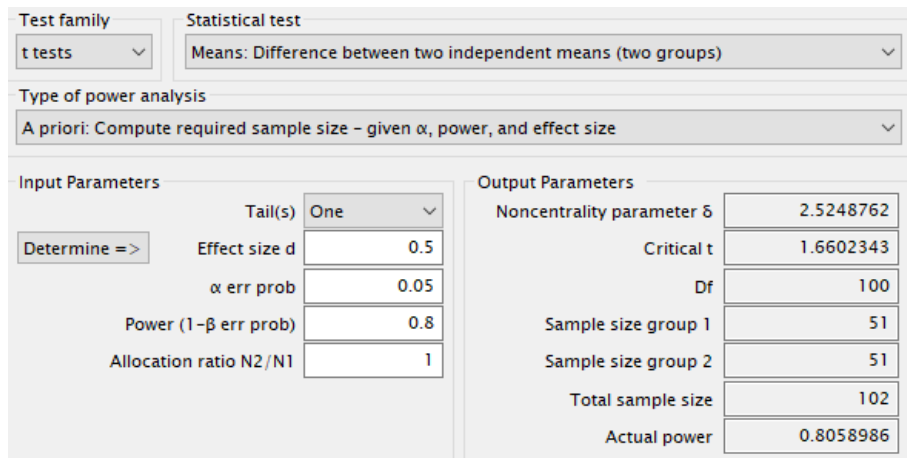
Cárdenas y Arancibia (2016) analizan los beneficios de utilizar G\*Power que es un programa estadístico diseñado para realizar estimaciones de la potencia estadística y del tamaño del efecto. También nos brindan una tabla con los valores para interpretar el tamaño del efecto en diferentes pruebas estadísticas.

**Tabla 6.6** Valores referenciales para el tamaño del efecto en diferentes pruebas estadísticas.

Prueba	Pequeño	Mediano	Grande
Prueba t	0.20	0.50	0.80
ANOVA unifactorial	0.10	0.25	0.40
ANOVA factorial	0.01	0.06	0.14
Chi cuadrado	0.10	0.30	0.50
Regresión múltiple	0.02	0.15	0.35

Fuente: Cárdenas y Arancibia (2016)

En la figura 6.1 se muestra el cálculo del tamaño de muestra para un estudio donde se comparan dos medias en grupos independientes usando 0.5 como tamaño del efecto, nivel de significancia de 0.05 y 0.80 de poder estadístico. Los resultados, según G\*Power, indican que 51 es el tamaño de muestra calculado para cada grupo, por ello la muestra total se compone de 102 elementos.



**Figura 6.1** Tamaño de muestra para una prueba t para dos medias independientes.

En la figura 6.2 se muestra el cálculo del tamaño de muestra para un estudio donde se comparan más de dos medias en grupos independientes usando 0.4 como tamaño del efecto, nivel de significancia de 0.05 y 0.80 de poder estadístico. Los resultados, según G\*Power, indican que 72 es el tamaño de muestra calculado.

Test family	Statistical test	
F tests	ANOVA: Fixed effects, omnibus, one-way	
Type of power analysis		
A priori: Compute required sample size - given $\alpha$ , power, and effect size		
Input Parameters		Output Parameters
Determine =>	Effect size f	Noncentrality parameter $\lambda$
	0.4	11.5200000
	$\alpha$ err prob	Critical F
	0.05	3.1296440
	Power (1- $\beta$ err prob)	Numerator df
	0.85	2
	Number of groups	Denominator df
	3	69
		Total sample size
		72
		Actual power
		0.8534928

**Figura 6.2** Tamaño de muestra para una prueba ANOVA.

En la figura 6.3 se muestra el cálculo del tamaño de muestra para un estudio donde se precisa establecer la asociación entre variables categóricas con la prueba Ji Cuadrado usando 0.3 como tamaño del efecto, nivel de significancia de 0.05, 0.90 de poder estadístico y un grado de libertad. Los resultados, según G\*Power, indican que 117 es el tamaño de muestra calculado.

Test family	Statistical test	
$\chi^2$ tests	Goodness-of-fit tests: Contingency tables	
Type of power analysis		
A priori: Compute required sample size - given $\alpha$ , power, and effect size		
Input Parameters		Output Parameters
Determine =>	Effect size w	Noncentrality parameter $\lambda$
	0.3	10.5300000
	$\alpha$ err prob	Critical $\chi^2$
	0.05	3.8414588
	Power (1- $\beta$ err prob)	Total sample size
	0.90	117
	Df	Actual power
	1	0.9006096

**Figura 6.3** Tamaño de muestra para una prueba Ji Cuadrado.

En la figura 6.4 se muestra el cálculo del tamaño de muestra para un estudio correlacional usando un nivel de significancia de 0.05, 0.80 de poder estadístico y un coeficiente de correlación estimado de 0.5. Los

resultados, según G\*Power, indican que 29 es el tamaño de muestra calculado.

Test family		Statistical test	
Exact		Correlation: Bivariate normal model	
Type of power analysis			
A priori: Compute required sample size - given $\alpha$ , power, and effect size			
Input Parameters		Output Parameters	
Determine =>		Tail(s)	Two
Correlation $\rho$ H1	0.5	Lower critical r	-0.3672777
$\alpha$ err prob	0.05	Upper critical r	0.3672777
Power ( $1-\beta$ err prob)	0.80	Total sample size	29
Correlation $\rho$ H0	0	Actual power	0.8139420

**Figura 6.4** Tamaño de muestra para una investigación correlacional

En la figura 6.5 se muestra el cálculo del tamaño de muestra para un estudio correlacional usando un nivel de significancia de 0.05, 0.80 de poder estadístico y un coeficiente de correlación estimado de 0.3. Los resultados, según G\*Power, indican que 84 es el tamaño de muestra calculado.

Test family		Statistical test	
Exact		Correlation: Bivariate normal model	
Type of power analysis			
A priori: Compute required sample size - given $\alpha$ , power, and effect size			
Input Parameters		Output Parameters	
Determine =>		Tail(s)	Two
Correlation $\rho$ H1	0.3	Lower critical r	-0.2145669
$\alpha$ err prob	0.05	Upper critical r	0.2145669
Power ( $1-\beta$ err prob)	0.80	Total sample size	84
Correlation $\rho$ H0	0	Actual power	0.8003390

**Figura 6.5** Tamaño de muestra para una investigación correlacional

En la figura 6.6 se muestra el cálculo del tamaño de muestra para un análisis de regresión lineal múltiple usando un nivel de significancia de 0.05, 0.95 de poder estadístico, 0.15 de tamaño de efecto y 2 variables

independientes. Los resultados, según G\*Power, indican que 107 es el tamaño de muestra calculado.

Test family	Statistical test
F tests	Linear multiple regression: Fixed model, R <sup>2</sup> deviation from zero
Type of power analysis	
A priori: Compute required sample size - given $\alpha$ , power, and effect size	
Input Parameters	
Determine =>	Effect size f <sup>2</sup> 0.15
	$\alpha$ err prob 0.05
	Power (1- $\beta$ err prob) 0.95
	Number of predictors 2
Output Parameters	
	Noncentrality parameter $\lambda$ 16.0500000
	Critical F 3.0837059
	Numerator df 2
	Denominator df 104
	Total sample size 107
	Actual power 0.9518556

**Figura 6.6** Tamaño de muestra para un análisis de regresión lineal múltiple



## Capítulo 7: Ejemplos adicionales

**Ejemplo 1:** Un docente investigador desea analizar el nivel de comprensión lectora en estudiantes del nivel secundario en el distrito de Santa Anita. Para ello, selecciona los 5 colegios nacionales del distrito y obtiene la siguiente información. Realice el muestreo estratificado proporcional.

**Tabla 7.1** Tamaño de los estratos.

<b>Colegio</b>	<b>Cantidad de alumnos en el nivel secundario</b>
San Andrés	380
Tupac Amaru II	250
Alto Cénepa	410
Pachacutec	320
Politécnico II	280

Primero calculamos el tamaño de la población:

$$N = 380 + 250 + 410 + 320 + 280 = 1640$$

Luego utilizamos la fórmula para el cálculo del tamaño de la muestra considerando 95% de confianza, 4% como límite aceptable de error y  $p=0.5$ :

$$n = \frac{1640 \times 1.96^2 \times 0.5^2}{(1640 - 1)0.04^2 + 1.96^2 \times 0.5^2}$$

$$n = 439.62 \approx 440$$

Calculamos la constante de proporcionalidad:

$$c = \frac{n}{N}$$

$$c = \frac{440}{1640} = 0.2683$$

**Tabla 7.2** Tamaño de muestra de cada estrato.

<b>Colegio</b>	<b>Cantidad de alumnos</b>	<b>Tamaño de muestra por estrato</b>
San Andrés	380	$380 \times 0.2683 = 102$
Tupac Amaru II	250	$250 \times 0.2683 = 67$
Alto Cenepa	410	$410 \times 0.2683 = 110$
Pachacútec	320	$320 \times 0.2683 = 86$
Politécnico II	280	$280 \times 0.2683 = 75$

Verificamos el tamaño de la muestra:

$$102 + 67 + 110 + 86 + 75 = 440$$

Para elegir los elementos de la muestra de cada estrato se puede usar un muestreo aleatorio simple o un muestreo sistemático.

Para realizar el muestreo aleatorio se puede utilizar SPSS o EXCEL

### **Observación**

Existen diferentes procesos estadísticos para realizar muestreo estratificado. Incluso se puede calcular el tamaño muestral sin necesidad de establecer un nivel de confianza. En procesos más complejos, inclusive se considera el costo de muestreo para cada

estrato. Entonces, concluimos que existen fórmulas para el muestreo estratificado con afijación uniforme, con afijación proporcional y con afijación óptima. Además, se especifican fórmulas para estimar la media o la proporción en cada tipo de afijación. Sin embargo, en este apartado solo veremos la fórmula para el muestreo estratificado con afijación proporcional. La información complementaria se encuentra en las referencias bibliográficas.

De acuerdo con Torres (2008) el proceso estadístico más adecuado para el cálculo del tamaño muestral en un muestreo estratificado proporcional con k estratos para estimar la media se obtiene usando la expresión:

$$n = \frac{\sum_{i=1}^k \frac{N_i}{N} S_i^2}{e^2 + \frac{1}{N} \sum_{i=1}^k \frac{N_i}{N} S_i^2}$$

Donde  $i=1, 2, 3, \dots, k$ ,  $N_i$  es el tamaño de cada estrato,  $S_i^2$  es la varianza de cada estrato y  $N$  es el tamaño de la población. Sin embargo, no siempre se cuenta con las varianzas de los estratos y se tiene que hacer una estimación.

Con los datos del ejemplo desarrollado anteriormente elaboramos la siguiente tabla considerando que el nivel de comprensión lectora se mide a través de la calificación vigesimal (0 a 20) y las varianzas que se muestran son estimadas.

**Tabla 7.3** Elementos para el cálculo del tamaño muestral.

Estrato	$N_i$	$S_i^2$	$\frac{N_i}{N} S_i^2$
San Andrés	380	20.25	4.692
Tupac Amaru II	250	25	3.811
Alto Cenepa	410	20.25	5.063
Pachacútec	320	25	4.878
Politécnico II	280	27.04	4.617
Total	1640		23.061

De la tabla tenemos la siguiente expresión:

$$\sum_{i=1}^k \frac{N_i}{N} S_i^2 = 23.061$$

Reemplazando este valor en la ecuación del tamaño de la muestra:

$$n = \frac{23.061}{0.4^2 + \frac{1}{1640} \times 23.061}$$

$$n = 132.488$$

$$n \approx 133$$

Este tamaño de muestra se obtiene considerando un error de estimación para la media de 0.4 puntos. Si utilizamos un error más pequeño, el tamaño de muestra aumenta su valor. Por ejemplo, si utilizamos un error de 0.2 tenemos:

$$n = 426.569$$

$$n \approx 427$$

Entonces, el tamaño de la muestra depende del error de estimación y de las varianzas de los estratos.

También se debe considerar la siguiente equivalencia en caso de contar con los valores de las varianzas de cada estrato a partir de las cuales se pueden calcular las varianzas muestrales de cada estrato:

$$S_i^2 = \frac{N_i}{N_i - 1} \times \sigma^2$$

Luego de obtener el tamaño de la muestra se procede con la estratificación proporcional.

**Ejemplo 2:** Un equipo investigador médico desea analizar las condiciones de trabajo de los profesionales de la salud en los establecimientos de salud de nivel de atención II y III a nivel nacional. Debido al alcance nacional de la investigación se opta por un muestro estratificado proporcional. Realice dicho procedimiento.

Para determinar la cantidad de hospitales que componen la población de interés se utiliza la información brindada por el MINSA en el 2011. De dicha información se extrae la siguiente tabla.

**Tabla 7.4** Establecimientos de salud, según nivel de atención, categoría y nivel de complejidad. Ministerio de Salud, Perú – 2007

Nivel de Atención	Categoría	Nivel de Complejidad	Establecimientos
I	I - 1	1	4 107
	I - 2	2	1 399
	I - 3	3	817
	I - 4	4	290
II	II - 1	5	59
	II - 2	6	26
III	III - 1	7	20
	III - 2	8	7
Sin nivel			308
Total			7 033

Fuente: MINSA “Experiencias de Planificación de los Recursos Humanos en Salud, Perú 2007-2010”.

Podemos reducir esta tabla considerando solo los estratos de interés para este estudio.

**Tabla 7.5** Tamaño de cada estrato.

Nivel	Categoría	Estrato	Establecimientos
I	II-1	1	59
	II-2	2	26
II	III-1	3	20
	III-2	4	7

Primero calculamos el tamaño de la población considerando solo los Niveles de atención II y III:

$$N = 59 + 26 + 20 + 7 = 112$$

Luego utilizamos la fórmula para el cálculo del tamaño de la muestra considerando 95% de confianza, 3% como límite aceptable de error y  $p=0.5$ :

$$n = \frac{112 \times 1.96^2 \times 0.5^2}{(112 - 1)0.03^2 + 1.96^2 \times 0.5^2}$$

$$n = 101.448 \approx 102$$

Aparentemente la reducción en la cantidad de hospitales a analizar ha sido mínima, pero hay que considerar que analizar un solo hospital ya representa un trabajo enorme.

Calculamos la constante de proporcionalidad:

$$k = \frac{n}{N}$$

$$k = \frac{102}{112} = 0.9107$$

Luego multiplicamos esta constante por cada uno de los tamaños de cada estrato:

$$0.9107 \times 59 = 53.7313$$

$$0.9107 \times 26 = 23.6782$$

$$0.9107 \times 20 = 18.2140$$

$$0.9107 \times 7 = 6.3749$$

Finalmente, aproximamos a números enteros los tamaños de muestra para cada estrato calculados en el proceso anterior:

**Tabla 7.6** Tamaño de muestra para cada estrato.

Nivel	Categoría	Estrato	Establecimientos
I	II-1	1	54
	II-2	2	24
II	III-1	3	18
	III-2	4	6
Total			102

**Observación**

Como en el ejemplo anterior utilizaremos otro método para el cálculo del tamaño de muestra para un muestreo estratificado proporcional.

Navarro (2011) nos Brinda la siguiente expresión para el cálculo de n cuando se quiere estimar la media en un muestreo estratificado proporcional:

$$n = \frac{\sum_{i=1}^k N_i \sigma_i^2}{ND + \frac{1}{N} \sum_{i=1}^k N_i \sigma_i^2}$$

Donde  $D = \frac{e^2}{4}$  y  $\sigma_i^2$  es la varianza de cada estrato.

Según el ejemplo desarrollado, se conoce el tamaño de cada estrato, pero no se conocen las varianzas de cada estrato, por ello se hará una

estimación de dichas varianzas. Esta estimación es solo para el ejemplo, el proceso correcto es basarse en información previa o en una prueba piloto.

Consideremos que lo que se trata de estimar es la media y los resultados se basan en una lista de cotejo tomada en cada hospital y el puntaje máximo acumulable es 100 puntos. Además, el error máximo aceptable será de 1 punto.

**Tabla 7.7** Elementos para el cálculo del tamaño muestral.

Estrato	$N_i$	$\sigma_i^2$	$N_i\sigma_i^2$
1	59	100	5900
2	26	81	2016
3	20	169	3380
4	7	144	1008
Total	$N = 112$		12304

De la tabla obtenemos:

$$\sum_{i=1}^k N_i\sigma_i^2 = 12304$$

Además:

$$D = \frac{1^2}{4} = 0.25$$

Reemplazando los valores en la fórmula:

$$n = \frac{12304}{112 \times 0.25 + \frac{1}{112} \times 12304}$$

$$n = 89.2518$$

$$n \approx 90$$



La interrogante sobre la fórmula adecuada para el muestreo estratificado, ahora depende de los recursos disponibles. Si es factible realizar la investigación en 102 hospitales ¿debería utilizar solo 90 hospitales? Siguiendo la lógica del muestreo, cuanto mayor sea el tamaño de la muestra tendremos una mejor estimación de las características de la población.

## Referencias

- Anderson, D., Sweeney, D. y Williams, T. (2012). *Estadística para negocios y economía*. 11<sup>o</sup> ed. México D.F. Cengage Learning Editores.
- Baltar, F. y Gorjup, M. T. (2012). *Muestreo mixto online: Una aplicación en poblaciones ocultas*. *Intangible Capital*, 8(1), 123-149. Disponible en: <https://www.redalyc.org/articulo.oa?id=54924517006>
- Blair, R.C. y Taylor, R. A. (2008). *Bioestadística*. México. Pearson Educación de México.
- Blanco, C. (2011). *Encuestas y estadísticas: Métodos de investigación cuantitativa en ciencias sociales y comunicación*. Córdoba, Argentina. Editorial Brujas.
- Blanco, M. (2020). *¿Cuántas universidades públicas y privadas hay en el Perú?* Blog Rankia. Disponible en: <https://www.rankia.pe/blog/mejores-universidades-escuelas-peru/4169587-cuantas-universidades-publicas-privadas-hay-peru#:~:text=Hay%20un%20total%20de%2051,fundada%20en%20el%20a%C3%B1o%201551>.
- Camacho, J. (2008). *Tamaño de muestra en estudios clínicos*. *Acta Médica Costarricense*, 50(1), 20-21. Disponible en: <https://www.redalyc.org/articulo.oa?id=43450104>
- Cárdenas, J. M., y Arancibia, H. (2016). *Potencia estadística y cálculo del tamaño del efecto en G\*Power: complementos a las pruebas de significación estadística y su aplicación en*

*psicología. Salud & Sociedad*, 5(2), 210-244.  
<https://doi.org/10.22199/S07187475.2014.0002.00006>

Catania, C. y Avagnina, S. (2007). *Curso superior de degustación de vinos*. EEA Mendoza. INTA. Disponible en: [https://inta.gob.ar/sites/default/files/script-tmp-29\\_\\_el\\_analisis\\_sensorial.pdf](https://inta.gob.ar/sites/default/files/script-tmp-29__el_analisis_sensorial.pdf)

Cochran, W. G. (1980). *Técnicas de muestreo*. México, D.F. Compañía Editorial Continental.

Demey, J. R., Zambrano, A. Y., Macchiavelli, R. y González, V. (2004). *Tamaño de muestra en experimentos biotecnológicos con suspensiones celulares*. *Interciencia*, 29(7), 396-400. [fecha de Consulta 26 de enero de 2022]. ISSN: 0378-1844. Disponible en: <https://www.redalyc.org/articulo.oa?id=33909409>

FAO (1999). *Métodos de muestreo recomendados para la determinación de residuos de plaguicidas a efectos del cumplimiento de los LMR*. CAC/GL 33-1999. Disponible en: [https://www.fao.org/fao-who-codexalimentarius/sh-proxy/es/?lnk=1&url=https%253A%252F%252Fworkspace.fao.org%252Fsites%252Fcodex%252Fstandards%252FCXG%2B33-1999%252FCXG\\_033s.pdf](https://www.fao.org/fao-who-codexalimentarius/sh-proxy/es/?lnk=1&url=https%253A%252F%252Fworkspace.fao.org%252Fsites%252Fcodex%252Fstandards%252FCXG%2B33-1999%252FCXG_033s.pdf)

FAO (2004). *Directrices generales sobre muestreo*. CAC/GL 50-2004. Disponible en: [https://www.fao.org/fao-who-codexalimentarius/sh-proxy/es/?lnk=1&url=https%253A%252F%252Fworkspace.fao.org%252Fsites%252Fcodex%252Fstandards%252FCXG%2B50-2004%252FCXG\\_050s.pdf](https://www.fao.org/fao-who-codexalimentarius/sh-proxy/es/?lnk=1&url=https%253A%252F%252Fworkspace.fao.org%252Fsites%252Fcodex%252Fstandards%252FCXG%2B50-2004%252FCXG_050s.pdf)

Flores, Z., Ordaz, F. y Ramírez, A. (2008). *Elementos de muestreo*. Universidad central de Caracas. Venezuela. Disponible en: <https://cienciassociales.webcindario.com/PDF/TecMuestreo.pdf>

- García, J. A., Reding, A. y López, J. C. (2013). *Cálculo del tamaño de la muestra en investigación en educación médica*. Departamento de Bioestadística y Bioinformática, Dirección de Investigación, Hospital General de México “Dr. Eduardo Liceaga”, México D.F., México.
- Google Data Studio. (2022). *Los distritos con mayor oferta para los estudios de tus hijos*. Disponible en: [https://datastudio.google.com/reporting/1\\_KgNbr8MsGtxzCotn-gj3BOyPb83h5UU/page/T7yFB](https://datastudio.google.com/reporting/1_KgNbr8MsGtxzCotn-gj3BOyPb83h5UU/page/T7yFB)
- Guan, T., Alam, M. K. and Rao, M. B. (2019). *Sample Size Calculations in Simple Linear Regression: Trials and Tribulations*. Department of Environmental Health, Division of Biostatistics and Bioinformatics, University of Cincinnati. Journal arXiv: Methodology. Disponible en: <https://arxiv.org/ftp/arxiv/papers/1907/1907.10569.pdf>
- Gutiérrez, H. y de la Vara, R. (2008). *Análisis y diseño de experimentos*. 2da Edición. México: McGraw-Hill.
- Hernández, R., Fernández, C. y Baptista, P. (2014). *Metodología de la Investigación*. México: McGraw-Hill.
- Hulley, S. B., Cummings, S. R., Browner, W. S., Grady, D., y Newman, T. B. (2013). *Designing clinical research: an epidemiologic approach*. 4.a ed. Philadelphia, USA. Lippincott Williams & Wilkins.
- IBM Corp. (2017). *IBM SPSS Statistics for Windows, Version 25.0*. Armonk, NY: IBM Corp.
- INEI (2014). *Una mirada a Lima Metropolitana*.
- INEI (2019). Perú: *Indicadores de Gestión Municipal 2019*. Disponible en: [https://www.inei.gob.pe/media/MenuRecursivo/publicaciones\\_digitales/Est/Lib1720/12.pdf](https://www.inei.gob.pe/media/MenuRecursivo/publicaciones_digitales/Est/Lib1720/12.pdf)

INS, (2021) Estado nutricional en adultos de 18 a 59 años, Perú: 2017 - 2018 Informe técnico de la Vigilancia Alimentaria y Nutricional por Etapas de Vida; Adultos 2017 2018. Disponible en:

[https://web.ins.gob.pe/sites/default/files/Archivos/cenan/van/sala\\_nutricional/sala\\_3/2021/Informe%20Tecnico-%20Estado%20nutricional%20en%20adultos%20de%2018%20a%2059%20a%C3%B1os%20CVIANEV%202017-2018.pdf](https://web.ins.gob.pe/sites/default/files/Archivos/cenan/van/sala_nutricional/sala_3/2021/Informe%20Tecnico-%20Estado%20nutricional%20en%20adultos%20de%2018%20a%2059%20a%C3%B1os%20CVIANEV%202017-2018.pdf)

Kleeberg, F. y Ramos, J. C. (2009). *Aplicación de las técnicas de muestreo en los negocios y la industria*. Ingeniería Industrial, (27),11-40. ISSN: 1025-9929. Disponible en: <https://www.redalyc.org/articulo.oa?id=337428493002>

Leyton, O. (2018). *La investigación en ciencias sociales. Establecimiento de hipótesis, métodos y técnicas de investigación, Información cualitativa y cuantitativa*. 2° ed. México D.F. Editorial Trillas.

Lind, D., Wathen, S.A. y Marchal, W. G. (2016). *Estadística aplicada a los negocios y la economía*. 16° ed. Madrid, España: Editorial McGraw Hill.

Lohr, S.L. (1999). *Muestreo: diseño y análisis*. México. International Thomson Editores.

Luna, V. M., Alva, A., Bernal, A. y Arteaga, M. (2011). Cuerpos de agua superficiales. En F. Bautista (Ed.), *Técnicas de muestreo para manejadores de recursos naturales*. 2da Ed. México. Universidad Nacional Autónoma de México.

Lwanga, S. y Lemeshov, S. (1991). *Determinación del tamaño de las muestras en estudios sanitarios: manual práctico*. Organización Mundial de la salud. Ginebra.

Manterola, C. y Otzen, T. (2014). *Estudios Observacionales. Los Diseños Utilizados con Mayor Frecuencia en Investigación*

*Clínica. International Journal of Morphology.* Vol.32 No.2 Temuco jun. 2014.  
[https://www.scielo.cl/scielo.php?script=sci\\_arttext&pid=S0717-95022014000200042](https://www.scielo.cl/scielo.php?script=sci_arttext&pid=S0717-95022014000200042)

Martínez, C. (2012). *Estadística y muestreo*. 13° Ed. Bogotá, Colombia. ECOE ediciones.

Martínez, M., Alonso, A. y Bes, M. (2006) *Estimación del tamaño Muestral*. En Martínez, M. (Ed), *Bioestadística Amigable* (pp. 373-396). 2da Edición. España. Ediciones Diaz de Dantos.

Mason, R. y Lind, D. (1992). *Estadística para Administración y Economía*. 7ma ed. México. Ediciones Alfaomega.

Mayor, J. (2007). *Muestreo Estadístico II. Diplomatura de Estadística. Curso Segundo*. Facultad de Matemáticas Universidad de Sevilla <https://personal.us.es/jmayor/ficheros/me204.pdf>

Medenhall, W., Beaver, R. y Beaver, B. (2010). *Introducción a la probabilidad y estadística*. 13° ed. México D.F. Cengage Learning Editores.

Minitab (2019) Versión 19.1.1. Disponible en: <https://www.minitab.com>

MINSA (2011) *Experiencias de Planificación de los Recursos Humanos en Salud, Perú 2007-2010*. Disponible en: <http://bvs.minsa.gob.pe/local/MINSA/1612-1.pdf>

MINSA (2019). *Compendio estadístico: Información de Recursos Humanos del sector salud Perú 2013-2018*. Lima. MINSA-DIGEP. Disponible en: <http://bvs.minsa.gob.pe/local/MINSA/10896.pdf>

- Moya, R. y Saravia, G. (2004). *Probabilidades e inferencia estadística*. Lima. 2da Edición. Editorial San Marcos.
- Navarro, J. (2011). *Introducción al diseño y análisis de poblaciones finitas*. En F. Bautista (Ed), *Técnicas de muestreo para manejadores de recursos naturales* (pp. 3-58). México. 2da Edición. Universidad nacional Autónoma de México.
- Nolberto, V. A. y Ponce, M. E. (2008). *Estadística inferencial aplicada*. Textos de la Maestría en Educación. Lima. Unidad de Post Grado de la Facultad de Educación de la Universidad Nacional Mayor de San Marcos.
- Norma Técnica Peruana 700.002 (2012). *Lineamientos y procedimientos de muestreo del pescado y productos pesqueros para inspección*. Disponible en: [http://www.sanipes.gob.pe/documentos/14\\_NTP700.0022012LineamientosyProcedimientosdeMuestreodePescadoyProductosPesquerosparaInspeccion.pdf](http://www.sanipes.gob.pe/documentos/14_NTP700.0022012LineamientosyProcedimientosdeMuestreodePescadoyProductosPesquerosparaInspeccion.pdf)
- Organización Mundial de la Salud (2004), *Codex Alimentarius - Directrices Generales sobre muestreo CAC/GL 50-2004*. Recuperado de: <https://www.fao.org/fao-who-codexalimentarius/codex-texts/guidelines/es/>
- Pajuelo, J., Torres, L., Agüero, R., y Bernui, I. (2019). *El sobrepeso, la obesidad y la obesidad abdominal en la población adulta del Perú*. Anales de la Facultad de Medicina, 80(1), 21-27. Disponible en: <https://dx.doi.org/10.15381/anales.v80i1.15863>
- Pardo, A. y San Martín, R. (2010). *Análisis de datos en ciencias sociales y de la salud II*. España. Editorial Síntesis.
- Pintado, M.B. (2014). *"Diseño experimental y consideraciones sobre el tamaño de muestra", "Animales de Laboratorio"*. En Revista de la Sociedad Española Para las Ciencias del Animal de Laboratorio, Número 62, Páginas 16-21. Disponible en:

[https://www.uib.cat/digitalAssets/303/303730\\_2014\\_animaleslaboratorio\\_num62\\_16\\_21.pdf](https://www.uib.cat/digitalAssets/303/303730_2014_animaleslaboratorio_num62_16_21.pdf)

Rand Corporation (1955) *Million Random Digits*. Disponible en: <http://www.rand.org/publications/classics/randomdigits>.

Rojo, A. (2014). "*Cálculo del tamaño muestral en procedimientos de experimentación con animales. Valoración de las incidencias*", "*Animales de Laboratorio*". En Revista de la Sociedad Española Para las Ciencias del Animal de Laboratorio, Número 62, Páginas 31-33. Disponible en: [https://www.uib.cat/digitalAssets/303/303729\\_2014\\_animaleslaboratorio\\_num62\\_31\\_33.pdf](https://www.uib.cat/digitalAssets/303/303729_2014_animaleslaboratorio_num62_31_33.pdf)

Soto, A. y Cvetkovic, A. (2020). *Estudios de casos y controles*. Rev. Fac. Med. Hum. Enero 2020; 20(1):138-143. DOI 10.25176/RFMH.v20i1.2555

Spiegel, M. y Stephens, L. (2009). *Estadística*. 4ta ed. México D.F. Editorial McGraw Hill.

Supo, J. (2014). *Cómo elegir una muestra. Técnicas para seleccionar una muestra representativa*. Arequipa, Perú. Editado e impreso por Bioestadístico EIRL.

Surco, J. C. y Alvarado, J. A. (2011). *Estudio estadístico de pruebas sensoriales de harinas compuestas para panificación*. Revista Boliviana de Química, 28(2),79-82. ISSN: 0250-5460. Disponible en: <https://www.redalyc.org/articulo.oa?id=426339676005>

Tamayo, M. (2003). *El proceso de la investigación científica. Incluye evaluación y administración de proyectos de investigación*. 4º ed. México D.F. Limusa Editores.

Torres, M. I. (2008). *Matemáticas Unex. "Muestreo"*. España Disponible en:



[http://matematicas.unex.es/~inmatorres/teaching/muestreo/assets/cap\\_4.pdf](http://matematicas.unex.es/~inmatorres/teaching/muestreo/assets/cap_4.pdf)

Vargas, V. (2007). *Estadística descriptiva para ingeniería ambiental con SPSS*. Cali, Colombia. Universidad Nacional de Colombia.

Webster, A. L. (2000) *Estadística aplicada a los negocios y a la economía*. 3ra Edición. Colombia. Editorial McGraw Hill.

Zappino, J. (2020). *Manual de estadística básica para no estadísticos*. Cuadernos del INAP No 13 año 1. Argentina. Instituto Nacional de Administración Pública.

ISBN: 978-612-00-7547-0



9 786120 075470