

Probability distributions IV: Normal distribution and determination of sample size (sampling theory).

Distribuciones de probabilidad IV: Distribución normal y determinación de tamaño de muestra (Teoría del muestreo).

Juan Luis Soto Espinosa ¹  <https://orcid.org/0000-0003-2600-9292>.

¹ FES Zaragoza, UNAM

Dirección (autor principal): Batalla de 5 de Mayo esq. Fuerte de Loreto, Col Ejército de Oriente, Alcaldía Iztapalapa, C.P.09230, Ciudad de México.

Correo electrónico de contacto: soej@unam.mx

Fecha de envío: 27/09/2025

Fecha de aprobación: 18/12/2025

DE

Abstract

Sampling theory is defined as the study of the relationships between a population and the different subsets of elements (samples) drawn from it. This theory is applicable to various fields of knowledge since it allows us to estimate population characteristics based on population parameters such as measures of central tendency (mean, median) and measures of dispersion (standard deviation and variance).

Sampling theory aims to determine, as closely as possible, the value of population parameters by studying the corresponding sample characteristics (sample statistics), always ensuring that the samples considered are significant; that is, that the statistics obtained from them are as close as possible to the population parameters.

Another relevant use of sampling theory is that it allows us to determine whether the observed differences between two samples are due to random variations or whether, on the other hand, they are statistically significant.

Keywords: Sampling theory, sample, statistical difference, population parameters, representativeness

Resumen

La Teoría del Muestreo se define como el estudio de las relaciones existentes entre una población y los diferentes subconjuntos de elementos (muestras) extraídos de ella. Esta teoría es aplicable a diversos campos de conocimiento ya que permite estimar características de las poblaciones a partir de parámetros poblacionales tales como las medidas de tendencia central (media, mediana) y las medidas de dispersión (desviación estándar y varianza).

La teoría del muestreo pretende conocer lo más cercano posible el valor de los parámetros poblacionales, a partir del estudio de las características muestrales correspondientes (estadísticos muestrales), buscando siempre que las muestras consideradas sean significativas, es decir, que los estadísticos que de ellas se obtengan sean lo más cercano posible a los parámetros poblacionales.

Otro uso relevante de la teoría de muestreo es que permite determinar si las diferencias observadas entre dos muestras se deben a variaciones relacionadas con el azar o si, por otro lado, resultan estadísticamente significativas.

Palabras clave: Teoría de muestreo, muestra, diferencia estadística, parámetros poblacionales, representatividad

Introducción

Una característica importante para que un estudio de investigación se considere relevante es que el número de elementos o participantes considerados sea suficientemente grande y diverso para reflejar las características de la población, para con ello obtener un error máximo permisible que tenga un valor muy

bajo y los resultados obtenidos puedan ser extrapolados a la población objetivo.

Con el fin de garantizar que las muestras consideradas en una investigación cumplan con los requisitos necesarios para que los análisis y resultados presenten el menor nivel de error posible y tengan una confianza elevada, se utiliza la Teoría del muestreo.

Documentos Educativos

La Teoría del Muestreo se define como el estudio de las relaciones existentes entre una población y los diferentes subconjuntos de elementos (muestras) extraídos de ella.

Esta teoría es aplicable a diversos campos de conocimiento ya que permite estimar características de las poblaciones a partir de parámetros poblacionales tales como las medidas de tendencia central (media, mediana) y las medidas de dispersión (desviación estándar y varianza).

Cuando una medida de tendencia central es determinada en la población de estudio, recibe el nombre de PARÁMETRO, mientras que si estas medidas son determinadas en las muestras, reciben el nombre genérico de ESTADÍSTICOS.

La teoría del muestreo pretende conocer lo más cercano posible el valor de los parámetros poblacionales, a partir del estudio de las características muestrales correspondientes (estadísticos muestrales).

Otro uso relevante de la teoría de muestreo es que permite determinar si las diferencias observadas entre dos muestras se deben a variaciones relacionadas con el azar o si, por otro lado, resultan estadísticamente significativas.

Etapas de planeación

En la etapa de planeación de un estudio se deben tener en cuenta dos fases de suma relevancia, la primera tiene que ver con la determinación de la cantidad de elementos o participantes van a ser considerados dentro del estudio, considerando sus características, las variables de estudio y la variabilidad que existe entre los individuos participantes respecto a las variables que se están estudiando.

La segunda etapa considera los procedimientos y algoritmos a través de los cuales se selecciona a los

participantes dentro de la población que conformarían la muestra de estudio.

Un proceso de muestreo tiene siempre como propósito obtener valores de parámetros poblacionales obteniendo estadísticos de la o las muestras que se obtengan de la población objetivo. la más cercano 'posible al valor real. Si los elementos que componen la muestra son adecuados para el estudio en cuestión la diferencia entre los estadísticos muestrales y los parámetros poblacionales será lo más pequeña posible y por lo tanto se tendrá un resultado confiable.

Las estimaciones obtenidas en un proceso de muestreo pueden ser, por casualidad, cercanas al valor real o ser muy diferentes del valor real del parámetro estudiado, en otras palabras, un proceso de muestreo deficiente puede dar resultados inconsistentes, diferentes al valor real de la variable, extremadamente complicados de reproducir y difíciles de explicar.

Un sistema de muestreo deficiente puede, en ocasiones por efecto del azar, proporcionar estimaciones muy exactas; de la misma manera, un excelente sistema de muestreo puede proporcionar estimaciones muy alejadas del verdadero valor del parámetro estudiado.

Un criterio importante para evaluar un sistema de muestreo consiste en realizar varios experimentos con elementos obtenidos de la misma población, realizar un análisis descriptivo de los estadísticos obtenidos y observar la distribución de frecuencias de las estimaciones obtenidas. Si se trata de un buen sistema de muestreo, el análisis de frecuencia de las estimaciones obtenidas presentará una varianza baja y el promedio estadístico de las estimaciones se encontrará próximo al verdadero valor del parámetro poblacional.

DE

Documentos Educativos

Sesgos

Cuando existe diferencia entre el valor promedio de las estimaciones muestrales y el valor verdadero del parámetro poblacional, se dice que se presenta un **SESGO**. Se define como sesgo a la diferencia entre el valor verdadero y el promedio de mediciones en muchas determinaciones obtenidas **POR EL MISMO SISTEMA DE MEDICIÓN**. (El término «sesgo» se usa también refiriéndose al proceso por el cual se producen las diferencias.)

El tamaño del sesgo que se presenta en un proceso de estimación y la varianza de un sistema de muestreo son, en términos generales, independientes entre sí; un sistema puede dar estimaciones con una varianza muy pequeña (con poca variación entre valores) pero con una gran diferencia respecto al valor real.

En un ejemplo:

Se realiza un estudio acerca de la calificación de 100 estudiantes en una evaluación final, con los siguientes resultados:

Alumno	Calificación	Alumno	Calificación
1	6	51	6
2	7	52	10
3	10	53	5
4	5	54	10
5	5	55	10
6	10	56	5
7	6	57	10
8	7	58	5
9	6	59	10
10	6	60	10
11	10	61	6
12	7	62	10
13	5	63	8
14	10	64	9
15	9	65	10
16	5	66	5
17	8	67	5
18	5	68	5
19	10	69	8

Alumno	Calificación	Alumno	Calificación
20	10	70	8
21	5	71	5
22	10	72	7
23	9	73	5
24	5	74	10
25	10	75	8
26	5	76	5
27	5	77	5
28	8	78	5
29	9	79	5
30	6	80	10
31	7	81	6
32	6	82	10
33	5	83	5
34	10	84	5
35	5	85	6
36	10	86	10
37	9	87	5
38	10	88	9
39	8	89	5
40	5	90	10
41	8	91	6
42	5	92	10
43	5	93	5
44	7	94	9
45	5	95	5
46	6	96	10
47	10	97	6
48	7	98	10
49	10	99	5
50	10	100	6

Los parámetros estadísticos de esta población son:

Parámetros	
Media	7.3
Varianza	4.47

Para realizar un análisis estadístico se toman cinco muestras de cinco estudiantes, considerando tres sistemas de muestreo diferentes, cuyo resultado se presenta a continuación:

DE

Documentos Educativos

Sistema de muestreo	Muestra A	Muestra B	Muestra C	Muestra D	Muestra E
1	8	6	6	5	6
	6	5	6	8	6
	5	6	6	6	7
	6	7	8	6	6
	5	6	5	6	5
2	9	6	10	8	8
	8	10	8	9	9
	8	5	5	6	10
	10	10	6	10	6
	10	5	8	9	10
3	5	8	7	5	10
	5	5	10	6	6
	6	9	10	10	10
	9	5	5	9	5
	8	6	5	9	7

Los parámetros estadísticos de cada sistema de muestreo se presentan a continuación:

Sistema 1	Muestra A	Muestra B	Muestra C	Muestra D	Muestra E	Promedio
Media	6	6	6.2	6.2	6	6.08
Varianza	1.5	0.5	1.2	1.2	0.5	0.98

Sistema 2	Muestra A	Muestra B	Muestra C	Muestra D	Muestra E	Promedio
Media	9	7.2	7.4	8.4	8.6	8.12
Varianza	1	6.7	3.8	2.3	2.8	3.32

Sistema 3	Muestra A	Muestra B	Muestra C	Muestra D	Muestra E	Promedio
Media	6.6	6.6	7.4	7.8	7.6	7.2
Varianza	3.3	3.3	6.3	4.7	5.3	4.58

De los tres métodos de muestreo, el que presenta estadísticos más cercanos al valor real del parámetro poblacional es el sistema 3, por lo que un contraste de hipótesis tendría mayor confiabilidad estadística en aquel conjunto de dato cuyos estadísticos sean lo

más cercano posible al valor del parámetro poblacional.

Al estudio de las deducciones (inferencias) realizadas a partir de muestras obtenidas de una población y generalizadas a la misma, junto con las estimaciones de precisión a través de la Teoría de Probabilidad se le conoce como Inferencia Estadística o Estadística inferencial.

La Teoría del Muestro permite determinar si las diferencias observadas entre dos muestras tomadas de la misma población son producidas por efecto del azar o si efectivamente se presentan independientemente del azar y son estadísticamente significativas.

Para que los resultados obtenidos en un experimento estadístico, se debe asegurar, en la medida de lo posible que las muestras obtenidas sean representativas, es por eso que el Diseño del Experimento es tan importante, ya que a través de él es posible estudiar los diferentes métodos de muestreo y los diferentes aspectos positivos y negativos que implican.

Muestreo Aleatorio

El Muestreo Aleatorio es un proceso a través del cual cada elemento o integrante de una población tiene las mismas probabilidades de ser seleccionado como parte de una muestra.

Este método es fundamental para garantizar la representatividad de la muestra, lo que permite a los investigadores inferir y generalizar sus conclusiones a toda la población con un mayor grado de precisión.

El principio en el que el muestreo aleatorio se fundamenta es la imparcialidad de la selección, este principio minimiza el riesgo de que el investigador introduzca sesgo en el proceso de selección de la muestra; al tener cada integrante de la población la misma probabilidad de ser incluido en el estudio, se garantiza que la muestra refleje de mejor manera las

DE

Documentos Educativos

características y la diversidad que se presenta en la población.

Como para obtener los elementos de una muestra se toma en cuenta la probabilidad que tiene cada uno en ser seleccionado, este tipo de muestreo es considerado como Muestreo Probabilístico.

El muestreo aleatorio es muy utilizado en estudios donde se pretende llegar a conclusiones que puedan ser extrapoladas hacia los parámetros de la población, como por ejemplo medias, varianzas o proporciones, a partir de los datos obtenidos de la muestra.

Otra ventaja que se tiene es que a través del uso del muestreo aleatorio es posible estimar el margen de error y el nivel de confianza de las conclusiones obtenidas de los datos muestrales y determinar la precisión y fiabilidad de los resultados.

Existen principalmente cuatro tipos de muestreo aleatorio:

1. *Muestreo aleatorio simple*
2. *Muestreo aleatorio estratificado*
3. *Muestreo aleatorio por conglomerados y*
4. *Muestreo aleatorio sistemático.*

Cada tipo de muestreo uno se utiliza para diferentes situaciones

Muestreo Aleatorio Simple

También conocido como Muestreo Aleatorio Puro, es un procedimiento estadístico a través del cual es posible seleccionar subconjuntos (muestras) representativos de una población a través de técnicas que favorezcan la selección al azar de los diferentes participantes en un estudio.

Los pasos que se deben seguir para llevar a cabo un muestreo aleatorio simple se pueden resumir en:

- 1) *Establecer la población de estudio:*
En este punto se debe delimitar de la

forma más clara posible el tamaño de la población de estudio.

- 2) *Definir el marco muestral: El marco muestral es una lista completa de todos los elementos de la población, este listado es un insumo de gran relevancia para la selección de la muestra.*
- 3) *Elegir un método de selección: Para garantizar que los elementos en la población tengan la misma probabilidad de ser elegidos, se debe elegir una técnica de selección completamente al azar, con esto se procura que la variabilidad que existe en la población de origen se vea reflejada en la muestra electa.*
- 4) *Realizar la selección al azar: Se debe utilizar la técnica seleccionada para elegir tantos elementos al azar como demande el tamaño de muestra elegido.*

Las técnicas utilizadas consideran la aleatoriedad como un principio fundamental en el proceso de selección, algunos ejemplos de este proceso se presentan a continuación:

- *Lotería aleatoria: A cada integrante de la población se le identifica a través de un número o cadena de texto. En un contenedor se depositan los identificadores donde son mezclados y extraídos de forma aleatoria. Los elementos seleccionados se seguirán extrayendo hasta completar el tamaño de muestra deseado, luego de lo cual, se revisan los elementos extraídos.*
- *Métodos físicos: Para poblaciones pequeñas, es posible utilizar elementos como dados, monedas al aire, boletos o ruletas. Una vez repartidos entre los miembros de la población, se realiza la selección*

DE

Documentos Educativos

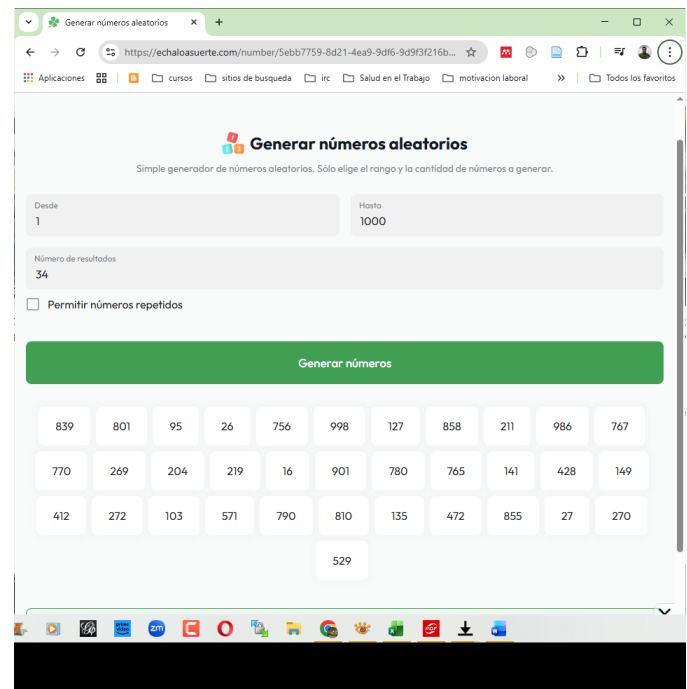
estableciendo un criterio de selección aleatoria basado en alguna característica de los objetos repartidos (color, tamaño, forma, texto, números, etc).

- *Tabla de números aleatorios: Es posible utilizar herramientas estadísticas preconstruidas, como por ejemplo, las tablas de números aleatorios que se pueden consultar en libros de estadística, revistas especializadas e incluso en sitios de Internet. El uso de este tipo de herramientas garantiza que el proceso de selección tenga la aleatoriedad requerida en la selección de los elementos integrantes de la muestra.*
- *Generadores de números aleatorios: En la actualidad existen herramientas digitales que pueden ser descargadas a diferentes dispositivos, tales como teléfonos celulares inteligentes, tabletas o computadoras; además es posible acceder a sitios en línea donde es posible establecer los valores mínimo y máximo a considerar y solicitar el tamaño de la muestra a seleccionar, con esto el sistema en cuestión generará de forma aleatoria los números correspondientes al tamaño de muestra que se requiere. Un ejemplo de este tipo de aplicaciones es ÉCHALOASUERTE, esta aplicación se puede acceder en la URL:*

<https://echaloasuerte.com/number>

En esta interfaz se debe ingresar el valor inferior, el valor superior y la cantidad de valores aleatorios que se desea generar entre los dos números de referencia. Cabe señalar que además de generar números aleatorios, esta aplicación también permite generar letras aleatorias (si se desea utilizar el alfabeto como criterio de selección), seleccionar elementos directamente de una lista de forma aleatoria o asociar elementos de una lista,

todo utilizando herramientas de selección aleatoria.



- *Uso de hojas de cálculo: Dentro del software de productividad que actualmente se utiliza, es posible acceder a hojas de cálculo, que son programas de cómputo cuya función en capturar, procesar, representar e imprimir datos esencialmente numéricos. Estos programas cuentan con funciones incorporadas para generar números aleatorios, entre las que se pueden mencionar:*

En **Microsoft Excel**, la función $=ALEATORIO.ENTRE(a,b)$, esta función genera números aleatorios entre dos valores, un valor mínimo (a) y un valor máximo (b), es una función que puede repetirse tantas veces como sea necesario y estará generando continuamente números aleatorios cada vez que se realice una acción en Excel. La sintaxis es la siguiente:

Supongamos que desea generar un número aleatorio entre 1 y 100, la instrucción sería:

DE

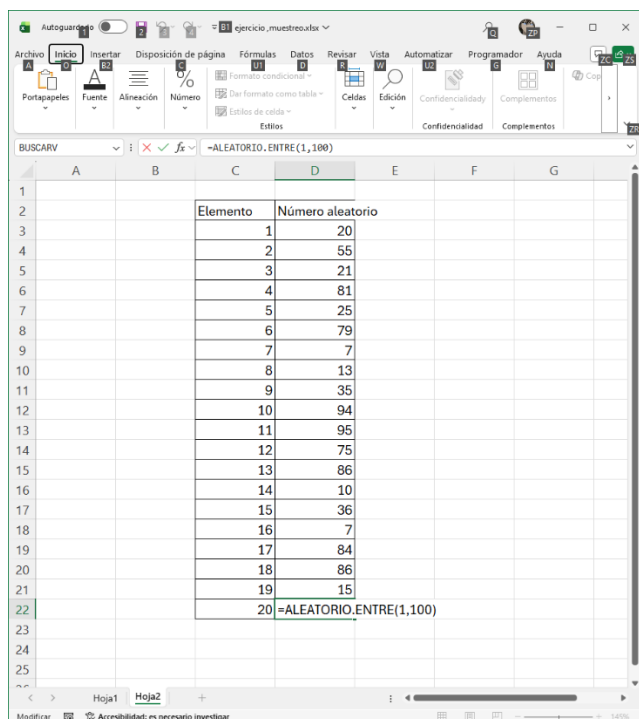
Documentos Educativos

=ALEATORIO.ENTRE(*a*,*b*)

Donde *a* representa el límite inferior y *b* representa el límite superior, sustituyendo se tiene:

=ALEATORIO.ENTRE(1,100)

Como resultado la función devolverá un número al azar entre 1 y 100; si se requieren 20 elementos al azar, basta copiar y pegar la función en 20 celdas para que realice la determinación en una sola intención, o bien, presionar la tecla F2 y <ENTER> 20 veces anotando cada vez el número resultante.



Elemento	Número aleatorio
1	20
2	55
3	21
4	81
5	25
6	79
7	7
8	13
9	35
10	94
11	95
12	75
13	86
14	10
15	36
16	7
17	84
18	86
19	15
20	=ALEATORIO.ENTRE(1,100)

En Google Sheets existe la función equivalente =RANDBETWEEN(*a*,*b*) que tiene la misma lógica de funcionamiento, por lo que en este caso *a* también representa el valor inferior y *b* el superior. Si desea obtener 34 números aleatorios entre 200 y 750, se tiene que:

Valor inferior *a* = 200

Valor superior *b* = 750

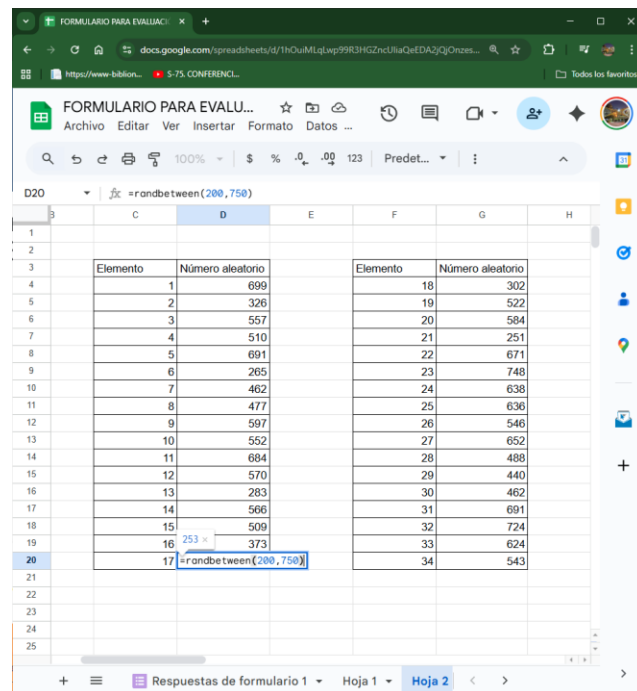
La instrucción en Google Sheets sería:

=RANDBETWEEN(*a*,*b*)

Sustituyendo valores se tiene:

=RANDBETWEEN(200,750)

Con lo que se obtienen números al azar entre el valor 200 y el 750.



Elemento	Número aleatorio
1	699
2	326
3	557
4	510
5	691
6	265
7	462
8	477
9	597
10	552
11	684
12	570
13	283
14	566
15	509
16	253
17	373
18	302
19	522
20	584
21	251
22	671
23	748
24	638
25	636
26	546
27	652
28	488
29	440
30	462
31	691
32	724
33	624
34	543

Muestreo aleatorio estratificado

El muestreo aleatorio estratificado es una técnica de probabilística de muestreo que es utilizada cuando dentro de una población es posible identificar claramente subgrupos de elementos (estratos) con base en alguna característica de sus elementos.

Cuando se utiliza este método de muestreo, el proceso de selección de los elementos que constituirán la muestra se debe llevar a cabo de manera independiente en cada estrato, sin dejar ninguno sin considerar. Las ventajas que este tipo de muestreo tiene respecto al muestreo aleatorio simple son:

Documentos Educativos

- Facilita la planeación del proceso de toma de muestra, considerando los recursos requeridos para seleccionar elementos de cada estrato.
- Proporciona criterios para organizar el muestreo de forma diferencial, tomando en cada estrato la cantidad justa de datos, en otras palabras, se colecta una mayor cantidad de datos en los estratos de mayor tamaño y menos datos en los estratos más pequeños.
- Permite mantener en la muestra el mismo número y la misma proporción de los estratos que se presenta en la población estudiada, lo que garantiza un reflejo más fiel de la variabilidad poblacional.

El muestreo aleatorio estratificado es útil cuando:

- Se desea estudiar una población con un número de subgrupos claramente definidos.
- En cada uno de los grupos estudiados, es posible calcular la media de la variable de estudio (μ) y su desviación estándares (σ), de tal forma que en el estrato i la media estaría dada por μ_i y la desviación estándar por σ_i .
- En una población de tamaño conocido (N), cuando se conocen los tamaños de los distintos estratos, $N_1, N_2, \dots, N_K$, siendo

$$N = \sum_{i=1}^k N_i$$

Donde

N = Tamaño de la población

N_i = Tamaño del subgrupo o estrato i

K = Número de subgrupos o estratos identificados en la población.

En este caso, la proporción de elementos de la población está dada por:

$$\phi_i = \frac{N_i}{N}$$

Donde:

Φ = Proporción

N_i = Número de elementos en el estrato i

N = Número de elementos en la población

- En una población infinita se conocen las proporciones $\phi_1, \phi_2, \dots, \phi_k$ de elementos de la misma dentro de cada estrato, se tiene que

$$\sum_{i=1}^k \phi_i = 1.$$

Dónde:

K = Número de subgrupos o estratos

Φ_i = Proporción del estrato i

Es decir, la suma de todas las proporciones de los i estratos en la población siempre será igual a 1.

Ejemplo:

Se tiene una población de 4,000 trabajadores con diferentes puestos, distribuidos de la siguiente manera:

800 obreros

1200 ayudantes

600 choferes

200 mecánicos

400 vendedores

800 administrativos

Se desea realizar un estudio tomando una muestra representativa de 100 trabajadores, para lo cual se procede a calcular las proporciones de cada estrato.

Conociendo que:

DE

Documentos Educativos

$N = 4000$

N_1 (obreros) = 800

N_2 (Ayudantes) = 1200

N_3 (Choferes) = 600

N_4 (Mecánicos) = 200

N_5 (Vendedores) = 400

N_6 (Administrativos) = 800

$$\phi_1 = \frac{800}{4000} = 0.2$$

$$\phi_2 = \frac{1200}{4000} = 0.3$$

$$\phi_3 = \frac{600}{4000} = 0.15$$

$$\phi_4 = \frac{200}{4000} = 0.05$$

$$\phi_5 = \frac{400}{4000} = 0.1$$

$$\phi_6 = \frac{800}{4000} = 0.2$$

Conociendo las proporciones, es posible determinar el tamaño de muestra que debe considerar cada estrato, por lo que para una muestra de 100 participantes se tiene la siguiente distribución:

$$n_i = n * \phi_i$$

Dónde

n = Tamaño de muestra

n_i = Tamaño del estrato i

ϕ_i = Proporción del estrato i

Por lo que, para determinar el tamaño de cada estrato de una muestra de tamaño 100, se debe resolver:

$$n_1 = n * \phi_1 = 100 * 0.2 = 20$$

$$n_2 = n * \phi_2 = 100 * 0.3 = 30$$

$$n_3 = n * \phi_3 = 100 * 0.15 = 15$$

$$n_4 = n * \phi_4 = 100 * 0.05 = 5$$

$$n_5 = n * \phi_5 = 100 * 0.1 = 10$$

$$n_6 = n * \phi_6 = 100 * 0.2 = 20$$

De acuerdo con este cálculo, la muestra representativa deberá considerar 20 obreros, 30 ayudantes, 15 choferes, 5 mecánicos, 10 vendedores y 20 administrativos, dando un total de 100 participantes.

Una vez determinado el porcentaje de cada uno de los estratos se procede a realizar la selección en cada uno utilizando las técnicas descritas en el muestreo aleatorio simple hasta alcanzar el número de participantes por estrato.

Muestreo aleatorio por conglomerados

El muestreo aleatorio por conglomerados, también conocido como muestreo conglomerado o muestreo por racimos, es una técnica probabilística que consiste en seleccionar elementos de una población que se encuentra dividida en grupos más o menos homogéneos y bien definidos. Este tipo de muestreo es especialmente útil cuando se trabaja con poblaciones grandes y dispersas geográficamente.

Cada subgrupo o conglomerado debe representar las características de la población total. Este tipo de muestreo se aplica cuando los elementos de la población pueden ser clasificados en grupos y esta clasificación es exhaustiva y excluyente, es decir, se debe considerar que cada uno de los elementos de la población pertenece a una y solo una de las categorías seleccionadas. Cabe señalar que a los subgrupos en los que se divide una población también reciben el nombre de clústers (aglomerado, grupo, racimo o conjunto).

DE

Documentos Educativos

Una técnica de muestreo aleatorio se utiliza entonces en los clústeres pertinentes para elegir qué grupos incluir en el estudio. En el muestreo por conglomerados en una sola etapa, se utilizan todos los elementos de cada uno de los grupos seleccionados. En el muestreo por conglomerados en dos etapas, una técnica de muestreo aleatorio se aplica a los elementos de cada uno de los grupos seleccionados.

La principal diferencia entre el muestreo por grupos y el muestreo estratificado es que en el muestreo por conglomerados el grupo se trata como la unidad de muestreo, de modo que la selección se realiza en una población de racimos (al menos en la primera etapa). dentro de cada estrato. En el muestreo estratificado, se toma una muestra aleatoria de cada uno de los estratos, mientras que en el muestreo por conglomerados sólo se toman muestras de los grupos seleccionados. Una motivación común del muestreo por conglomerados es reducir los costos aumentando la eficiencia del muestreo. Esto contrasta con el muestreo estratificado en el que la motivación es aumentar la precisión.

También hay muestreo de grupos de etapas múltiples, donde al menos dos etapas se toman en la selección de elementos de los grupos. De forma análoga al muestreo aleatorio estratificado, la población de estudio se encuentra dividida en subpoblaciones o subgrupos claramente definidos, llamados conglomerados. Mientras que en el muestreo estratificado se toman muestras de cada uno de los niveles (estratos) en el muestreo por conglomerados no se consideran todos los subgrupos, sino que se toma de forma aleatoria una muestra de los mismos, es decir, no se trabaja con todos los conglomerados de la población.

Una vez definidos los conglomerados a considerar, se puede proceder de dos formas:

- Se procede a obtener los datos de todos los integrantes del conglomerado de forma exhaustiva, en cuyo caso se dice que se trata de un muestreo por conglomerados en una etapa.
- Se procede a elegir una o varias muestras de los integrantes de un mismo estrato, en cuyo caso se dice que se tiene un muestreo por conglomerados en dos o más etapas.

Muestreo por conglomerados en una etapa

Una población con N número de integrantes:

$$U = \{x_1, x_2, x_3, \dots, x_N\}$$

Se encuentra dividida en i número de conglomerados identificados como $C_1, C_2, C_3, \dots, C_i$:

$$U = \{C_1, C_2, C_3, \dots, C_i\}$$

Cada conglomerado está conformado por j elementos:

$$C_i = \{i_1, i_2, i_3, \dots, i_j\}$$

En este tipo de muestreo se toman datos de los j elementos que conforman cada conglomerado C_i .

El muestreo en una etapa es la forma más sencilla pues en cada conglomerado se realiza la selección de elementos utilizando un muestreo aleatorio simple; se utiliza cuando los conglomerados son relativamente homogéneos y se puede suponer que cada uno de ellos es una réplica reducida de la población. Es especialmente útil en situaciones en las que se necesita un método rápido y rentable, y en las que la representación detallada de cada conglomerado es menos crítica.

El muestreo por conglomerados de una sola etapa toma una muestra aleatoria de conglomerados de la población y, dentro de éste, realiza el acopio de datos de TODOS los individuos dentro de ellos. Esta selección se realiza a través de técnicas de muestreo aleatorio simple. Este método de muestreo es sumamente útil cuando los conglomerados son

DE

Documentos Educativos

pequeños, pero puede ser menos preciso que otros tipos de muestreo, ya que los individuos dentro de un conglomerado tienden a ser similares, lo que genera información redundante.

Un ejemplo de muestreo por conglomerados en una etapa se presenta a continuación:

Se desea realizar una encuesta de calidad en la docencia de alumnos del tercer semestre de cuatro diferentes posgrados, para lo cual se realiza una selección de grupos al azar (conglomerados) correspondientes a cada posgrado y se aplica la encuesta de acopio de información a todos los estudiantes de cada grupo seleccionado. En este tipo de muestreo se realiza una selección aleatoria de los conglomerados a estudiar y al interior de ella se practica un CENSO a la totalidad de participantes

Muestreo por conglomerados en dos etapas

El muestreo por conglomerados en dos etapas inicia dividiendo la población en conglomerados a la población de estudio, igual que el muestreo de una etapa, pero posteriormente se realiza una segunda selección aleatoria entre los integrantes del conglomerado.

A diferencias del muestreo por conglomerados de una etapa, donde se recopilan datos de todos los integrantes de los conglomerados, el muestreo por conglomerados en dos etapas selecciona aleatoriamente a los individuos dentro de cada conglomerado seleccionado para recabar datos solamente en ellos. Esto es:

Una población con N número de integrantes:

$$U = \{x_1, x_2, x_3, \dots, x_N\}$$

Se encuentra dividida en i número de conglomerados identificados como $C_1, C_2, C_3, \dots, C_i$:

$$U = \{C_1, C_2, C_3, \dots, C_i\}$$

Cada conglomerado está conformado por j elementos:

$$C_i = \{i_1, i_2, i_3, \dots, i_j\}$$

En un muestreo de dos etapas se procede a seleccionar muestras al interior de cada conglomerado de k elementos.

$$M_i = \{i_1, i_2, i_3, \dots, i_k\}$$

Donde

$$k \in j \text{ (k pertenece a j)}$$

y además

$$k \leq j$$

k debe ser siempre menor o igual a j , dado que es un subconjunto de elementos de ese conjunto.

Un ejemplo de muestreo por conglomerados en dos etapas es el siguiente:

El estudio de calidad en la docencia del ejemplo anterior en una escuela que imparte 11 licenciaturas de 9 semestres cada una y entre 5 y 8 grupos por semestre, en la primera etapa se realiza un muestreo aleatorio simple de para seleccionar los semestres a considerar y al interior un segundo muestreo aleatorio simple para definir los grupos que han de participar en el estudio.

Muestreo por conglomerados polietápico o multietápico

Cuando se tienen poblaciones de gran tamaño, alta variabilidad y con un número importante de variables a considerar, los muestreos de una o dos etapas resultan limitados, es por ello que es posible realizar muestreos que consideren tantas etapas como niveles de organización tenga la población.

Ejemplo de Muestreo por conglomerados polietápico:

Se desea realizar un estudio de pertinencia de los contenidos de los posgrados en salud en el trabajo

DE

Documentos Educativos

que se imparten en el país, para ello se realiza la definición de conglomerados, que en este caso serían las entidades federativas de la República Mexicana, al interior de cada conglomerado se realiza un muestreo aleatorio para definir los municipios que habrán de participar y dentro de ellos se realiza un tercer muestreo aleatorio para definir las escuelas que serán consideradas dentro del estudio. Si el estudio demanda un nivel superior de especificidad o de complejidad, se pueden considerar variables como el semestre, turno, grupos de edad, etc.

Muestreo aleatorio sistemático.

El muestreo aleatorio sistemático es un tipo de muestreo probabilístico que parte de la selección aleatoria del primer elemento del conjunto de elementos que conformarán la muestra y se selecciona el resto siguiendo un algoritmo de selección previamente establecido.

El algoritmo de selección consiste en la definición de un intervalo fijo que permita ir seleccionando elementos a partir del desplazamiento aplicado a los elementos que conforman la población hasta completar el tamaño de muestra requerido.

Para realizar un muestreo sistemático efectivo, es necesario definir claramente la población objetivo del estudio; de la misma manera, se debe establecer el tamaño de muestra requerido.

Posteriormente hay que definir lo que en estadística se conoce como MARCO MUESTRAL, que es un listado exhaustivo (o base de datos) de TODOS los elementos que conforman la población; este listado es de suma importancia ya que de él se obtendrá la muestra de estudio, es por ello que resulta indispensable que contenga todos los elementos de la población, los cuales deben ser identificados de manera única y estar ordenados de forma aleatoria para que todos los elementos tengan la probabilidad de ser electos y sea posible realizar la selección

adecuada y sin sesgo. A los elementos que conforman el marco muestral se les denomina UNIDADES MUESTRALES.

En ocasiones y dependiendo de las características de estudio, las unidades muestrales pueden organizarse en función de la variable de estudio, con lo que se genera una estratificación derivada de los diferentes valores que esta última puede adquirir. Una estratificación permite la selección de elementos de cada estrato para garantizar la selección de todos los valores de la variable en la proporción en la que se presenta en la población de estudio.

Conociendo el tamaño de la población (N) y el tamaño de muestra requerido (n), se procede a realizar la siguiente operación

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

Dónde:

i = Intervalo de selección

N = Número de elementos en el marco muestral

n = Número de elementos en la muestra

Ejemplo:

En una institución de salud se desea realizar un monitoreo de salud para determinar la edad a la que los pacientes adscritos han cumplido con el Programa Nacional de Vacunación, para lo cual se procede a construir un listado de derechohabientes, el cual está conformado por 450 personas con una desviación estándar de la edad igual a 0.25. Como se conoce el valor de la varianza poblacional de la variable edad, se determina el tamaño de muestra utilizando la fórmula de cálculo de tamaño de muestra para variables cuantitativas de una población finita, la cual establece:

DE

Documentos Educativos

$$n = \frac{N z_{\alpha/2}^2 s^2}{d^2 (N - 1) + z_{\alpha/2}^2 s^2}$$

Dónde:

N = Número de elementos en el marco muestral

Z = variable normal estandarizada Z para el nivel de confianza estadística deseado

S² = Varianza poblacional

d = error máximo permisible en el estudio

Recuerde que el valor de la variable estandarizada Z depende del porcentaje de confiabilidad estadística que debe tener el estudio, por lo que este varía dependiendo de la confianza deseada, siendo los valores más comunes los siguientes:

Confianza estadística (Potencia de la prueba)	Valor Z
90 %	1.645
95 %	1.96
99 %	2.575
99.5 %	2.807

Sustituyendo valores se tiene:

Considerando un nivel de confianza del 95% y un error máximo permisible de 5%, se tiene que el valor de la variable estandarizada Z para ese nivel de confianza es igual a 1.96, mientras que el valor del error máximo permisible es de 0.05.

Sustituyendo:

$$n = \frac{450 * 1.96^2 * 0.25^2}{0.05^2 (450 - 1) + 1.96^2 * 0.25^2}$$

$$n = \frac{450 * 0.2401}{1.1225 + 0.2401}$$

$$n = \frac{108.045}{1.3626}$$

$$n = 79.293 \approx 80$$

Recuerde que la desviación estándar, por definición, es la raíz cuadrada de la varianza, razón por la cual el valor de 0.75 se eleva al cuadrado. Si se conoce el valor de la varianza poblacional, este se sustituye en la ecuación en lugar del s² directamente, es decir, basta con conocer uno de los dos valores para resolver la ecuación.

Recuerde que el tamaño de muestra y el error máximo permisible son inversamente proporcionales, por lo que se recomienda redondear el valor del número de elementos en la muestra, cuando tenga una fracción decimal, al número entero superior.

Una vez conocidos los valores del tamaño poblacional (N) y el tamaño de muestra (n), se procede a calcular el intervalo de desplazamiento:

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

$$i = \frac{450}{80}$$

$$i = 5.625 \approx 5$$

En el caso del cálculo del desplazamiento, se recomienda redondear el valor decimal hacia el valor entero inmediato inferior, para garantizar la eficiencia del proceso de selección.

Se procede a elegir de forma aleatoria el primer elemento de la muestra dentro de los cinco (5) primeros elementos en el marco muestral, suponiendo que sea el tercer elemento:

Primer elemento (e_i) en la muestra: e₃

A partir de este primer elemento aleatorio, se aplica el desplazamiento, de forma que, para calcular el valor de cada elemento se utiliza la fórmula:

$$e_i = e_1 + 5 * i - 1$$

Donde e₁ representa la posición del primer elemento electo al azar en la muestra, mientras que e_i

DE

Documentos Educativos

representa a los elementos que ocupen una posición i determinada.

Como el valor de

$$e1 = 3$$

$$e2 = 3 + 5 = 8$$

$$e3 = 3 + 5 * 2 = 13$$

$$e4 = 3 + 5 * 3 = 18$$

$$e5 = 3 + 5 * 4 = 23$$

...

$$e78 = 3 + 5 * 77 = 388$$

$$e79 = 3 + 5 * 78 = 393$$

$$e80 = 3 + 5 * 79 = 398$$

Siguiendo el algoritmo definido, los elementos que serán considerados en la muestra son los que ocupan los lugares 3, 8, 13, 18, 23, 28, 33, 38, 43, 48, 53, 58, 63, 68, 73, 78, 83, 88, 93, 98, 103, 108, 113, 118, 123, 128, 133, 138, 143, 148, 153, 158, 163, 168, 173, 178, 183, 188, 193, 198, 203, 208, 213, 218, 223, 228, 233, 238, 243, 248, 253, 258, 263, 268, 273, 278, 283, 288, 293, 298, 303, 308, 313, 318, 323, 328, 333, 338, 343, 348, 353, 358, 363, 368, 373, 378, 383, 388, 393, 398, 403, 408, 413, 418, 423, 428, 433, 438, 443 y 448, un total de 80 elementos.

Retomando este mismo ejemplo, hay ocasiones en las que se desconoce el valor de los parámetros poblacionales 'pero se conoce el tamaño de la población, como la media o la desviación estándar, en esos casos, el tamaño de muestra se puede calcular utilizando la fórmula para el cálculo de tamaño de muestras considerando proporciones con población finita, la cual establece:

$$n = \frac{N z_{\alpha/2}^2 p q}{d^2 (N - 1) + z_{\alpha/2}^2 p q}$$

Dónde:

n = Tamaño de muestra

N = Tamaño de la población

Z = Variable normal estandarizada para el nivel de confianza estadística deseado

d = error máximo permisible

p = Probabilidad de ocurrencia de la característica estudiada

q = Probabilidad complementaria (probabilidad de no ocurrencia de la característica estudiada).

Considerando una confianza estadística del 95% y un error máximo permisible del 5% se procede a realizar el cálculo del tamaño de muestra:

$$N = 450$$

$$d = 0.05$$

$$p = 0.5$$

$$q = 0.5$$

Si se conoce la probabilidad de ocurrencia de la característica de estudio (p) se puede utilizar en la fórmula, recordando que q es igual a 1-p.

En caso de que se desconozca el valor de p, se recomienda utilizar el valor de 0.5 para p y 0.5 para q, ya que estos valores garantizan el mayor tamaño de muestra para el nivel de confianza que se elija.

Con estas consideraciones, se sustituyen valores en la fórmula:

$$\begin{aligned} n &= \frac{N z_{\alpha/2}^2 p q}{d^2 (N - 1) + z_{\alpha/2}^2 p q} \\ n &= \frac{450 * 1.96^2 * 0.5 * 0.5}{0.05^2 (450 - 1) + 1.96^2 * 0.5 * 0.5} \\ n &= \frac{450 * 0.9604}{0.05^2 * (450 - 1) + 0.9604} \\ n &= \frac{432.18}{2.2079} \end{aligned}$$

DE

Documentos Educativos

$$N = 195.74 \approx 196$$

Si se conoce el valor de p dentro de la población, supongamos 0.35 (probabilidad de 35% de ocurrencia), se tiene que el valor de q será:

$$q = 1 - p$$

$$q = 1 - 0.35$$

$$q = 0.65$$

Sustituyendo en la fórmula:

$$n = \frac{N z_{\alpha/2}^2 p q}{d^2 (N - 1) + z_{\alpha/2}^2 p q}$$

$$n = \frac{450 * 1.96^2 * 0.35 * 0.65}{0.05^2 (450 - 1) + 1.96^2 * 0.35 * 0.65}$$

$$n = \frac{393.28}{2.1214}$$

$$N = 185.38 \approx 186$$

Note que se obtiene un número de muestra similar con los dos métodos, aunque el conocimiento detallado de la población siempre permitirá el cálculo con mayor precisión.

Referencias:

Departamento de Didáctica de la Matemática. (2011). *Estadística con proyectos*. (C. Batanero, & C. Díaz, Eds.) Granada, España: Facultad de Ciencias de la Educación, Universidad de Granada.

García Pérez, A. (2008). *Estadística aplicada: conceptos básicos* (2a edición ed.). Madrid, España: Educación permanente / Universidad Nacional de Educación a Distancia.

Pértegas Díaz, S., & Pita Fernández, S. (2001). *La distribución normal*. *Cad Aten Primaria*, 8, 268-274.

Vargas Velasco, C. (2018). *Distribución normal estándar*.

Baird, D. (1991). *Propiedades matemáticas de la distribución normal o de Gauss*.

Rev. chil. anest. Vol. 43 Número 2 pp. 116-121 | <https://doi.org/10.25237/revchilanestv43n02.08>

Condori-Ojeda, Porfirio (2020). *Universo, población y muestra*. Curso Taller. <https://www.aacademica.org/cporfirio/18>

DE

Declaración de conflicto de intereses

El autor de este artículo expresa que no tuvo ningún conflicto de intereses durante la preparación de este documento ni para su publicación.

Obra protegida con una licencia Creative Commons

