

Ejercicios Unidad I

1. Suponga que estamos investigando sobre el porcentaje de alumnos que trabajan de una población de 20 alumnos de la Universidad de Talca. Base de datos de la población:

Nombre Alumno	¿Trabaja?
Juan	SI
Alicia	NO
Pedro	NO
Marcos	NO
Alberto	SI
Jorge	SI
José	NO
Carlos	NO
Miguel	NO
Victoria	SI

Nombre Alumno	¿Trabaja?
María	NO
Fernanda	NO
Julio	SI
Rosa	NO
Fabián	NO
Ana	NO
Laura	NO
Enrique	NO
Carmen	SI
Marcelo	SI

- Elija una muestra aleatoria simple de tamaño $n=4$ de esta población. Use una tabla de números aleatorios o Excel para la generación de números aleatorios.
 - Indique cuál es el Parámetro y cuál es el Estadístico en (a).
 - Elija una muestra estratificada de tamaño $n=4$ de esta población. Use una tabla de números aleatorios o Excel para la generación de números aleatorios. Indique los pasos para elegir la muestra.
2. Se tiene a la siguiente población de personas clasificadas como consumidores de drogas:

Nombre	Droga
Felipe	Alcohol
Wilma	Marihuana
José	Pasta Base
Viviana	Cocaína
Pablo	Alcohol
Rodrigo	Marihuana
Carlos	Extasis
Catherine	Alcohol
Claudia	Marihuana
Valentina	Neoprén
Enrique	Marihuan

Nombre	Droga
Antonio	Alcohol
Gerardo	Marihuana
Carmen	Alcohol
Pamela	Cocaína
María	Pasta Base
Alejandra	Extasis
Eduardo	Neoprén
Ronal	Relevon
Susana	Heroína
Hugo	Marihuana
Hernán	Alcohol

- Elija una muestra aleatoria simple de tamaño $n=6$ de esta población. Use una tabla de números aleatorios o Excel para la generación de números aleatorios.
 - Suponga que estamos investigando sobre el porcentaje de personas consumidoras de alcohol, calcule el parámetro y el estadístico adecuado.
 - Elija una muestra estratificada de tamaño $n=6$ de esta población. Use una tabla de números aleatorios o Excel para la generación de números aleatorios. Indique los pasos para elegir la muestra.
- Determinar el número de profesionales a encuestar en una región donde se estima en 4500 el número de ellos. El objetivo del estudio es determinar entre otras cosas, la intencionalidad de seguir estudios de maestría, con una prueba piloto de 20 profesionales, se determinó que la proporción de profesionales con afán de continuar sus estudios era del 25%. La confiabilidad del estudio, dado que sus resultados serán validados con otras fuentes se definió en el 90%, el error puede estar entre el 4 y el 6%, dependiendo de los costos se definirá cual tamaño seleccionar.
 - Una Institución de Salud tiene 6100 empleados, se quiere determinar cómo es el clima laboral en la organización, usando una confiabilidad del 95%, un error admisible de 6% y considerando que la proporción de empleados no satisfechos es del 30%. Calcule el número de empleados a consultar por categoría, si se tiene en cuenta, que las diferentes categorías de empleados que pueden influir en la opinión de los trabajadores, tienen la siguiente distribución: Contabilidad y Costos 80 empleados, Administración 150, operativos 5600, seguridad 180 y otros cargos 90.
 - Suponga que el director de personal cree que el ingreso promedio de gerentes comerciales es de 51800 dólares y una desviación estándar de 4000, el director general cree que el promedio de la muestra dista menos de 500 dólares del ingreso medio nacional.
 - Trace la distribución muestral de \bar{x} cuando se usan muestras aleatorias simples de tamaño 60.
 - Que sucede a la distribución muestral de \bar{x} si se usan muestras aleatorias simples de tamaño 120.

- c. Que afirmación general se puede hacer sobre lo que sucede con la distribución muestral de \bar{x} al aumentar el tamaño de la muestra? ¿Parece lógica esta generalización? Explique su respuesta.
6. La media del precio por galón de gasolina regular vendida en Estados Unidos es de 1.20 dólares (*The Energy Information Administration, 3 de marzo de 1997*). Suponga que la media de la población del precio por galón es 1.20 dólares, y que la desviación estándar de la población es de 0.10 dólares. También suponga que se selecciona una muestra aleatoria de 50 gasolineras, y que se calcula un precio de la media de la muestra con los datos reunidos en esas gasolineras.
- Muestre la distribución de muestreo de la media de la muestra \bar{x} para las 50 gasolineras.
 - ¿Cuál es la probabilidad de que la muestra aleatoria simple produzca una media de la muestra a menos de 0.02 dólares de la media de la población?
 - ¿Cuál es la probabilidad de que la muestra aleatoria simple produzca una media de la muestra a menos de 0.01 dólares de la media de la población?
7. El programa de pruebas a universitarios de la oficina universitaria estadounidense reportó una calificación SAT de la media de la población de 1017 (*The New York Times, 2000*) suponga que la desviación estándar de la población es 100.
- ¿Cuál es la probabilidad que una muestra aleatoria de 75 estudiantes produzca una media de la muestra de calificación SAT que quede a 10 o menos de la media de población?
 - ¿Cuál es la probabilidad de que una muestra aleatoria de 75 estudiantes produzca una media de la muestra de calificación SAT que quede a 20 o menos de la media de población?
8. El salario inicial promedio anual para expertos en comercialización es de 34000 dólares (*Times 8 de mayo de 2000*). Suponga que para una población de graduados con especialidad en comercialización el salario inicial promedio anual de 34000 y que la desviación estándar es 2000.
- ¿Cuál es la probabilidad de que una muestra aleatoria simple de expertos en comercialización tengan una media muestral dentro ± 250 dólares de la media poblacional para cada uno de los siguientes tamaños de muestra: 40, 80, 120, 200 y 400?
 - ¿Cuál es la ventaja de un tamaño de muestra más grande al intentar estimar la media poblacional?
9. La renta promedio mensual para un departamento de dos recamaras en Atlanta es de 982 dólares (*Elie, septiembre de 1988*). Suponga que la media poblacional es 982 dólares y la desviación estándar es 210 dólares
- ¿Cuál es la probabilidad de que una muestra aleatoria simple de 40 departamentos de dos recamaras de un renta mensual promedio dentro ± 100 dólares de la media poblacional?
 - ¿Cuál es la probabilidad de que una muestra aleatoria simple de 40 departamentos de dos recamas de una renta mensual promedio dentro de ± 25 dólares de la media poblacional?
 - Analice los resultados de los incisos a y b.
10. El precio promedio poblacional de una casa nueva monofamiliar es de 166500 dólares (*New One-Family Homes Sold oficina del censo de Estados Unidos 1997*) Suponga que la desviación estándar de la población es de 42000 dólares y que seleccionara una muestra de 100 casas monofamiliar nuevas.
- Presente la distribución muestral del precio de la media de la muestra de casas, basada en la muestra de 100.
 - ¿Cuál es la probabilidad de que la media de la muestra de las 100 compras quede a 10000 dólares o menos de la media de la población?
 - Repita el inciso b para los valores de 5000, 2500 y 1000 dólares.
 - Para estimar el precio de la media de población con ± 2500 y ± 1000 dólares de aproximación ¿Qué recomienda usted?
11. Se informa en la revista *Business Week* que, entre sus suscriptores, los que planean comprar un automóvil nuevo pretenden gastar un promedio de 27100 dólares (*Business Week, perfil del suscriptor 1996*) suponga que el precio del nuevo vehículo, para la población de suscriptores de *Business Week*, tiene una media de 27100 dólares y que su desviación estándar es de 5200 dólares.
- ¿Cuál es la probabilidad de que el precio de la media de la muestra del nuevo vehículo quede a 1000 dólares o menos de la media de población, si la muestra es de 30 suscriptores?
 - ¿Cuál es la probabilidad de que ese precio quede a 1000 dólares o menos de la media de población si la muestra es de 50 suscriptores?
 - ¿cuál es la probabilidad de que ese precio quede a 1000 dólares o menos de la media de población, si la muestra es de 100 suscriptores?
 - Recomendaría usted un tamaño muestral de 30, 50 o 100 si se desea tener una probabilidad mínima de 0.90 de que ese precio quede a 1000 dólares o menos de la media de la población?
12. Para estimar la edad media de una población de 4000 empleados, se selecciona una muestra aleatoria simple de 40 empleados.
- ¿Usaría usted el factor de corrección para población finita para calcular el error estándar de la media? Explique por que.
 - Si la desviación estándar de la población es de 8.2 años, calcule el error estándar, aplicando y sin aplicar el factor de corrección para población finita
 - ¿Cuál es la probabilidad de que la edad promedio de una muestra de los empleados tenga una aproximación de ± 2 años a la media de la edad de la población?
13. La duración de las bombillas de 100 W que fabrica una empresa sigue una distribución normal con una desviación típica de 120 horas de duración. Su vida media está garantizada durante un mínimo de 800 horas. Se escoge al azar una muestra de 50 bombillas de un lote y, después de comprobarlas, se obtiene una vida media de 750 horas. Construir un intervalo de confianza bilateral del 94% para la duración promedio, ¿habría que rechazar el lote por no cumplir la garantía?
14. Se ha tomado una muestra aleatoria de 100 individuos a los que se ha medido el nivel de glucosa en sangre, obteniéndose una media muestral de 110 mg/cc. Se sabe que la desviación estándar de la población es de 20 mg/cc.
- Obtener un intervalo de confianza, al 90%, para el nivel de glucosa en sangre en la población.
 - Obtener un intervalo de confianza, al 94%, para el nivel de glucosa en sangre en la población

15. La media de edad de los alumnos que se presentan a pruebas de acceso a la Universidad es de 18.1 años, y la desviación estándar de 0.6 años.
 - a. De los alumnos anteriores se elige, al azar, una muestra de 100. ¿Cuál es la probabilidad de que la media de la edad de la muestra esté comprendida entre 17.9 y 18.2 años?.
 - b. Calcular un intervalo de confianza del 99%.
16. En una gran ciudad española, la altura media de sus habitantes tiene una desviación estándar de 8 cm. Se pide:
 - a. Si la altura media de dichos habitantes fuera de 175 cm. ¿cuál sería la probabilidad de que la altura media de una muestra de 100 individuos tomada al azar fuera superior a 176 cm? Explica los pasos seguidos para obtener la respuesta.
 - b. Si se considera una muestra aleatoria de 100 individuos de esta ciudad, se obtiene una altura media de 178 cm. Determina un intervalo de confianza del 95% para la altura media de los habitantes de esta ciudad. Explica los pasos seguidos para obtener la respuesta
17. La media de las medidas de los diámetros de una muestra aleatoria de 200 bolas de rodamiento fabricadas por cierta máquina fue de 0.824 cm y la desviación estándar fue de 0.042 cm. Halla los límites de confianza al 95% para el diámetro medio de las bolas fabricadas por esa máquina.
18. Sea X una población normal. Se sabe que, a partir de una muestra de 40 unidades, se saca un media muestral igual a 100 y una varianza muestral igual a 4. Calcular el intervalo de confianza para la media al 95% y 99%, sabiendo que la varianza de la población es igual a 36.
19. Dos maquinarias A y B, usadas para el corte de pieles para zapato, tienen que ser comparadas para comprobar la calidad de la producción. Se extrae una muestra de 50 elementos por la maquinaria A y 45 para la maquinaria B. También, se sabe que la varianza de los cortes es de 9 cm para el grupo A y 16 para el grupo B. Sabiendo que las medias muestrales obtenidas son respectivamente iguales a 140 cm y 120 cm, hallar un intervalo de confianza al 93% para cada una de las dos poblaciones.
20. Sea X una población cuya distribución es desconocida. Sabiendo que $n=40$, media muestral igual a 100 y desviación estándar de la población es igual a 4, se calcule el intervalo de confianza al 95%.
 - a. Que cambia si $n=20$?



La muestra de una muestra

El siguiente artículo sobre encuestas fue publicado en un número reciente de *The New York Times* en la red. Me pareció que es un artículo muy interesante y espero que ustedes también lo consideren así. Presentamos a continuación ese artículo en su totalidad.

Cómo se localiza al encuestado "típico"

Por MICHAEL KAGAY

"¿A qué clase de personas llama usted por teléfono para sus encuestas? Seguramente no está encuestando a ningún residente de Texas en este momento."

Los encuestadores reciben muchas llamadas como ésta, por lo general de personas que en verdad se sienten desconcertadas al ver que sus propios puntos de vista sobre algún tema de controversia, o incluso las opiniones de la mayoría de la gente de su respectiva localidad, no coinciden con la opinión registrada por las encuestas como la de la mayoría en todo el país.

Sin embargo, la mayor parte de los encuestadores tienen mucho cuidado y tratan de asegurarse de obtener una muestra representativa de entrevistados para sus encuestas. Es posible que la redacción correcta de las preguntas siga siendo un arte y que la conducción habilidosa de entrevistas sea todo un oficio, pero el muestreo apropiado es lo que introduce el elemento científico en las encuestas.

La marcación de dígitos al azar (o RDD por sus siglas en inglés) es

el procedimiento de muestreo estándar que todos los encuestadores profesionales usan en la actualidad. El objetivo es dar a todos los números telefónicos residenciales de los Estados Unidos la misma posibilidad de ser contactados para una entrevista. Casi todas las organizaciones de encuestas utilizan algún tipo de RDD.

Por ejemplo, la Encuesta de *The New York Times* y CBS News se basa en el sistema GENESYS, desarrollado y sostenido por el Marketing Systems Group de Filadelfia. Ese sistema consiste en una base de datos que contiene más de 42 000 conferencias telefónicas residenciales en todo el territorio del país, que se actualiza a intervalos de unos cuantos meses a fin de incluir los nuevos códigos de área y las conferencias realizadas en fechas más recientes.

Ese sistema incluye también software de computadora diseñado para extraer una muestra al azar de esas conferencias en cada nueva encuesta, y para formar al azar los cuatro últimos dígitos de cada uno de los números telefónicos que van a ser contactados.

Esa elección aleatoria del universo de todos los números telefónicos posibles es el factor que garantiza la representatividad de la muestra de las unidades familiares que son contactadas para una encuesta.

En el caso de la encuesta sobre la que preguntaba una persona al inicio de este artículo, los tejanos representan cerca de 7% del total de los estadounidenses, pero sólo tuvieron 6% de representación entre quienes respondieron a esa encuesta en particular. La cifra casi alcanzó la marca, pero quedó un poco baja.

La búsqueda insistente de entrevistados

Algunos tipos de personas son más difíciles de localizar y de entrevistar que otros. ¿Qué se hace para reducir o evitar la presencia de sesgos en los resultados?

El primer remedio consiste en la búsqueda insistente de las familias y luego del miembro designado de cada una de ellas que será encuestado.

En el curso de la mayoría de las encuestas hay que realizar múltiples llamadas a cada familia, lo cual incluye a menudo llamadas diurnas y nocturnas, entre semana y los fines de semana. Los entrevistadores dejan mensajes en máquinas contestadoras telefónicas sobre el propósito de la llamada, conciertan nuevas citas si la persona designada no está en la casa, y anuncian que volverán a llamar a una hora más conveniente para el entrevistado.

Cuando las personas se niegan a ser entrevistadas, en muchas encuestas se emplea también un escuadrón especial de entrevistadores dispuestos a hacer una llamada adicional para tratar de convertir una negativa en una entrevista completa. Esa tarea requiere destrezas especiales, habilidad para convencer a la persona que responde sobre la importancia de la entrevista, poseer la personalidad apropiada y ser capaz de tolerar un posible rechazo.

Ponderación o balanceo después de la entrevista

Sin embargo, casi todas las encuestas incluyen un número ligeramente excesivo de mujeres, personas jubiladas, egresados de la universidad e individuos blancos, en comparación con las proporciones de éstos que según la

(continúa)



La muestra de una muestra

(continuación)

Oficina del Censo existen en la población de Estados Unidos en un momento cualquiera.

Esto sucede así porque la frecuencia con la que las personas están en su casa es variable, así como su disposición para hablar por teléfono con desconocidos, y el grado de confianza que puedan sentir frente al hecho de compartir sus opiniones sobre temas políticos y sociales.

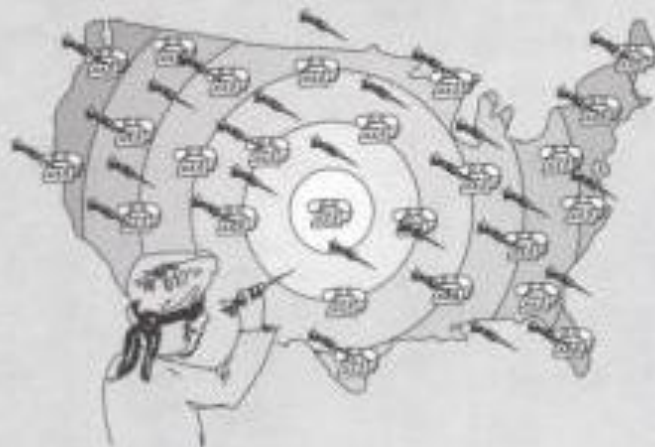
Por lo tanto, la mayoría de las encuestas ponderan ligeramente a sus encuestados —aumentando ligeramente la proporción de hombres, personas jóvenes, gente con menos instrucción y minorías raciales— para que sus proporciones tengan el equilibrio apropiado. Algunos encuestadores se refieren a este procedimiento como el balanceo.

Al mismo tiempo, la Encuesta del Times y CBS News aprovecha también la oportunidad y hace un ligero ajuste en cuanto al tamaño de la unidad familiar y el número de líneas telefónicas con que cuenta la vivienda. Con esto se equilibra la probabilidad de selección, en el caso de individuos en cuyas casas hay varias líneas telefónicas, y también el hecho de que en algunas viviendas familiares hay más residentes adultos que en otras.

Todos estos ajustes menores sirven para hacer que una buena muestra sea aún mejor y más representativa.

Una muestra que tenga la fisonomía del país

Entonces, ¿cómo son los entrevistados de una típica Encuesta del New



York Times y CBS News?

Después de todas las marcaciones telefónicas al azar, las segundas llamadas, la conversión de negativas en aceptaciones y la ponderación, la encuesta típica se compone de los siguientes elementos:

- En términos de geografía, 22% de los entrevistados viven en el noreste, 33% en el sur, 24% en el medio oeste y 21% en el oeste.
- En términos de género, 47% de ellos son hombres y 53% son mujeres.
- En términos de raza, 80% son blancos, 11% negros, 1% asiáticos y 6% de otras razas.
- En cuanto al nivel de educación, hay 24% que terminó la universidad, otro 27% que hizo algunos estudios en escuelas superiores o vocacionales, 37% que no pasó de la preparatoria y 12% que no terminó la preparatoria.
- En lo que se refiere a política, el ánimo cambia de vez en cuando, pero a últimas fechas 27% se considera republicano, al tiempo que 36% se identifica como demócrata y otro 30% asegura ser independiente.
- En cuanto a la filosofía política, 20% suele considerarse liberal,

mientras que 42% se identifica como moderado y 32% se proclama conservador.

Ésa es una muestra que sí refleja con bastante fidelidad a los Estados Unidos.

Bueno, cuando menos refleja a los Estados Unidos de habla inglesa, ya que la mayoría de las encuestas nacionales se realizan hoy únicamente en inglés. Ésta es una limitación de las encuestas que debe ser tomada en cuenta, ya que la población de habla española va en aumento en el país. Eso nos permite explicar también por qué la Encuesta del Times y CBS News se quedó un poco corta en cuanto a los tejanos incluidos en la encuesta a la que nos referíamos al principio del artículo; Texas es un estado en el cual la población de habla española es abundante.

Las encuestas tienen también otras limitaciones y problemas, y en futuras columnas de Poll Watch los iremos abordando también.

Fuente: Derechos de autor © 1999 por The New York Times Co. Reproducido con autorización.