

## Actividades y Lecturas complementarias Capítulo 7

### Finalidad

Se plantean una serie de actividades para que los estudiantes generen datos. Se propone resolver un rompecabezas del mapa de la República Mexicana, se le puede dar el enfoque administrativo económico. Estas ideas se pueden extender a otros países o América Latina. Varias actividades se plantean con temas relacionados con la economía y habilidades de percepción. Con la información generada se puede construir un curso de estadística.

Se proponen una serie de lecturas que, principalmente, se pueden obtener por internet. Ellas se irán actualizando.

### *Actividades y Lecturas complementarias Capítulo 7*

#### **Ejercicios de aplicación práctica**

Desde el capítulo 1 se han propuesto una serie de problemas prácticos con la finalidad de que los estudiantes realicen actividades tales que les permita realizar encuestas, así como algunas estrategias para recabar información. Cada una de éstas les permitirá generar datos, así como aprender a usar éstos para realizar análisis estadísticos. En particular para el tema desarrollado aquí, les permitirá extender las aplicaciones y formalizar los conceptos.

Se anexa un archivo en una hoja de cálculo: Excel, MR-CAP7.EXCEL, este contiene los resultados de una muestra de  $n = 150$  casos. Con ésta información, reportada por un entrevistador, se puede elaborar una práctica como la indicada más adelante.

#### **Consultar**

1. <http://www.calest.com/> En la opción: Apoyo Didático  
En esta dirección aparecen una serie de lecciones y artículos con el fin de complementar algunas ideas en el proceso enseñanza y aprendizaje. Algunas lecciones están en construcción y varias se irán mejorando y actualizando año con año.
2. <http://www.tuguiacountable.com/2012/05/el-muestreo-en-la-auditoria.html> Ver capítulo 7.
3. <https://explorable.com/es/tecnicas-de-muestreo-estadistico> Ver capítulo 7.

#### **Sobre el rompecabezas del mapa de la República Mexicana**

1. De la aplicación que ha realizado sobre el rompecabezas sobre la República Mexicana, considere los 120 datos capturados en el capítulo 2. Tome en cuenta que esta es una población, calcule la media, la varianza y la desviación estándar, considere estas tres medidas como parámetro. Siga las actividades que se proponen a continuación para estimar la media y la varianza.
  - a) Verifique si los datos siguen una distribución normal, use el papel de probabilidad normal.
  - b) Si se cumple el supuesto de normalidad, seleccione una muestra de tamaño  $n = 5, 10, \text{ y } 20$ , en cada caso obtenga la media, la varianza y la desviación estándar. Construya los intervalos de confianza del 90 %, 95 % y 99 %, discuta sus resultados.
  - c) Junte las muestras de tamaño  $n = 5$  del inciso anterior con 50 compañeros, si no, selecciónelas hasta obtener 50. Calcule en cada caso la media, la varianza y la desviación estándar; con esos 50 datos elabore un histograma para los tres estadísticos e interprete sus resultados.
  - d) Repita el inciso c para los tamaños de muestra  $n = 10, \text{ y } 20$ . En relación con sus resultados, obtenga sus conclusiones.
  - e) Si no se cumple el supuesto de normalidad, seleccione una muestra de tamaño  $n = 45$  y construya los intervalos de confianza del 85 %, 90 % y 97 %; discuta sus resultados.
  - f) Desde el planteamiento inicial se ha propuesto considerar diferentes grupos, ya determinados por la edad, por el grado escolar o alguna otra categorización que haya considerado. Para la situación o situaciones consideradas obtenga la media, la varianza y la desviación estándar. Con esa información construya el intervalo de confianza del 95 % para la media, varianza y desviación estándar. Interprete los resultados en el contexto referido.

### **Estudio sobre el consumo de energía en los hogares**

El consumo de energía es un servicio por el cual se debe pagar cada bimestre, a partir de este pago se crea una serie de cuestiones que resultan de interés para la economía familiar, porque representa un gasto. Por otro lado, es importante para el gobierno ya que se requiere de una estrategia para generar energía. Considerando los recibos de facturación se puede generar datos que permiten contestar varias preguntas, la primera de ellas (que es la más relevante) es el consumo de energía en cada uno de los hogares, indagar si en los seis pagos bimestrales existe un consumo regular de energía o hay grandes cambios, precisar si el consumo de energía en el periodo de verano es mayor que en el otro. En este problema también se puede plantear una encuesta para saber si hay mayor consumo de energía en las casas que tienen horno de microondas de las que no, en esa misma dirección se puede preguntar sobre la calefacción, el número de personas que habitan la casa, el área construida, zonas residenciales, entre otras preguntas.

El ahorro de energía eléctrica es una finalidad importante para mejorar el bienestar del ser humano. También en los hogares, empresas e industria en general, existe interés en disminuir el consumo de energía eléctrica con fines económicos. Para llevar a cabo las actividades que se citan abajo y realizar

un estudio para conocer el consumo, medido en kWh, utilice los recibos que genera la Comisión Federal de Electricidad.

### Actividades

1. Que una persona (alumno o estudiante) seleccione 10 personas y les pregunte sobre el consumo de energía del último recibo, haga un diagrama de caja e interprete el resultado. Estime las medidas de tendencia central, dispersión y posición e interprétalas.
2. Junte este resultado con el de otros 10 alumnos y realice un histograma, una tabla de frecuencias, un diagrama de puntos, un diagrama de tallo y hoja, y un diagrama de caja e interprete sus resultados. El polígono de frecuencia acumulado, ¿se aproxima a una distribución acumulada? A partir de este último diagrama indique ¿cuál es el consumo del 10 %, 20 %, 40 %, 60 % y 80 % de las personas entrevistadas?

### Aplique la siguiente encuesta

Pregunte a una persona por:

1. Número de personas que vive en su casa.
2. Área de construcción de la casa en la que habita.
3. Si tienen y usan horno de microondas.

Que el alumno responda a las siguientes preguntas:

1. a) ¿Son diferentes los diagramas de caja entre las personas que tienen horno de microondas y las que no lo tienen?  
 b) Construya un intervalo de confianza del 95 % para la media y varianza del consumo de energía.  
 c) Construya los intervalos de confianza del 95 % para la media y varianza del consumo de energía, considerando por separado los hogares que tienen horno de microondas y las que no lo tienen.

### Problema de habilidad de percepción: hombre en granos de café<sup>1</sup>

En la transparencia que se describe en la figura de este ejercicio se muestran granos de café, en uno de los cuales aparece la cara de un hombre. El problema que se plantea a partir de esta foto es conocer el desarrollo cerebral de las personas; los neurólogos y sicólogos utilizan este tipo de representación para realizar diversos estudios sobre la habilidad de percepción en diferentes individuos. Algunas de estas preguntas se plantean los ejercicios de este problema. La *actividad* consiste en mostrar a una

---

<sup>1</sup>Esta práctica la puede encontrar en [http://www.calest.com/Apoyo Didáctico](http://www.calest.com/Apoyo%20Didáctico)

persona de manera individual la transparencia, a partir de ese momento tomar el tiempo que le lleva encontrar al hombre en los granos de café. *Variable de respuesta:* tiempo en que se tarda una persona en responder. Tome el tiempo usando un cronómetro.

#### Actividades

1. Una persona (alumno o estudiante) selecciona 15 personas; a cada una, en forma individual, le muestra la transparencia y registra el tiempo de respuesta. (En este caso sin considerar la edad).
  - a) Hacer un diagrama de caja e interpretar. Calcule los estadísticos básicos e interprete.
  - b) Juntar la información con 10 personas y elaborar un histograma, tabla de frecuencias, un polígono de frecuencias, un diagrama de tallo y hoja, una gráfica de puntos, un diagrama de caja. En cada caso interprete sus resultados.
  - c) Tome una muestra de tamaño  $n = 30$  de sus 150 datos y calcule un intervalo de confianza de 90 % para la media, varianza y desviación estándar del tiempo de respuesta de las personas entrevistadas. Interprete sus resultados.

#### Problema: Memoria a corto plazo

Un problema que es de interés en diferentes áreas del conocimiento, consiste en conocer la capacidad de memoria a corto plazo. Con ese objetivo tienen desarrolladas varias actividades que les aplican a los participantes y así detectar esa capacidad.

Realice la siguiente práctica: ponga en una tarjeta blanca las 16 palabras que se indican en la tabla de abajo, luego muestre en un lapso de 20 segundos esa tarjeta a una persona, deje pasar un minuto y pídale que le diga las palabras que recuerda, pero no puede tardarse más de un minuto. Anote el número de palabras que recuerda, repita la actividad con otras 10 personas, junte sus datos con los de otros 10 estudiantes.

Candado	Sol	Pared	Cama
Botella	Lápiz	Cuchillo	Pulsera
Silla	Piedra	Manzana	Perro
Cuaderno	Rojo	Avión	Lluvia

Con los 100 datos obtenidos de la práctica:

1. Elabore un diagrama de puntos.
2. Estime un intervalo de confianza de 95 % para calcular la proporción de personas que recuerda más de 12 palabras.
3. Estime un intervalo de confianza del 95 % para estimar la proporción de personas que recuerda al menos 8 palabras?

### Actividad de la binomial relacionada con el ejemplo de la memoria

En esta actividad se presenta la distribución binomial con  $(n, p) = (10, 0.3)$ , esta situación se ilustra en la gráfica superior izquierda en la figura 7.E1. En esta misma se puede observar como varía esta distribución haciendo crecer  $n$ , por ejemplo:  $(20, 0.3)$ ,  $(30, 0.3)$ ,  $(60, 0.3)$ . Al final se percibe que con  $n$  suficientemente grande la distribución va aproximando a una forma simétrica.

Puede verificar esta actividad en **CalEst**, vea en la opción de distribuciones, en particular la binomial, ahí reproduzca los resultados indicados en figura 7.E1. A continuación practique considerando diferentes valores  $n$ , tal que  $n$  vaya creciendo para un valor fijo de  $p$ . Repita el procedimiento para otros valores de  $p$ .

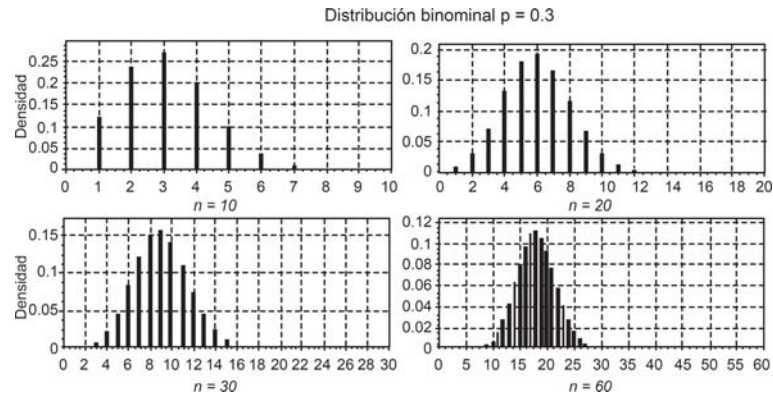


Figura 7.E1 Distribución binomial para  $p = 0.3$ , para diferentes valores crecientes de  $n$ .

### Tipos de muestreo

#### Muestreo aleatorio simple

Este tipo de procedimiento para seleccionar una muestra es el que se ha empleado durante lo expuesto en este capítulo. Una descripción más formal sobre esta técnica consiste en indicar que sea parte de una población lo más homogénea posible, cada elemento está en una lista e identificado por un número. A partir de ese número se recurre a una urna, tabla de números aleatorios para escoger una muestra de tamaño  $n$ . La población puede ser finita, tamaño  $N$ , o infinita. En el primer caso se puede obtener una muestra sin reemplazo, y cada una de las  ${}_N C_n$  muestras de tamaño  $n$  tiene la misma probabilidad de ocurrir. En una población infinita, la probabilidad de seleccionar un elemento dado es la misma cuando el muestreo es realizado con reemplazo o sin reemplazo.

#### Muestreo sistemático

Ideas preliminares, la administración de una empresa tiene un programa para la reducción de costos y mejorar la calidad en el servicio. Una muestra de clientes registrados es seleccionada de manera continua para tener una auditoria detallada. El número total de registros de la población no se conoce con anticipación, no se puede utilizar el muestreo aleatorio simple para seleccionar los registros. Sin

embargo, se puede tener una idea del número aproximado de registros disponibles para su selección en un determinado periodo de tiempo y así escoger uno de esos  $k$  registros. En esta situación  $k$  es un número entero que permitirá generar un valor para determinar la unidad objeto de estudio.

Por ejemplo, un administrador tiene disponibles 12 nuevas facturas registradas por día y desea tener una muestra total de 275 de éstas por año. El número total de facturas disponibles al año se estiman en  $12 \times 365 = 4380$ . Para obtener alrededor de las 275 facturas en la muestra anual, el valor de  $k = \frac{4380}{275} = 15.93$ . Se toma el valor entero más cercano, así  $k = 16$ . Este valor se denomina el intervalo muestral y se escoge al azar un número entre 1 y 16. El procedimiento que se sigue a continuación para alcanzar la muestra es identificar cada factura, o en su caso un registro, asignando números consecutivos iniciando por el 1. En este ejemplo, se escoge un número entre 1 y 16, para fijar ideas, pensemos en 9, ese es el primer elemento de la muestra, los demás serán los que resulten de sumar consecutivamente 16, es decir 9, 25, 41, así sucesivamente.

Muestreo sistemático se tiene una lista aleatoria de los  $N$  individuos de una población relativamente homogénea y se desea obtener una muestra de tamaño  $n$ . Como punto de partida se toma el valor entero más grande del cociente entre  $N$  y  $n$ , así  $k = \frac{N}{n}$ . De esa manera se determina el intervalo de muestra  $k$ , se selecciona un número aleatorio  $j$  entre 1 y  $k$ , a continuación se van seleccionando los elementos identificados con  $j, j + k, j + 2k, \dots$  así hasta obtener  $n$  elementos.

Observaciones: Puede no conocerse el valor de  $N$  al inicio de la muestra, se puede conocer al final del proceso del muestreo sistemático, es decir  $k \times n$ . La muestra de tamaño  $n$  que se genera mediante el muestreo sistemático permite calcular la media y la proporción muestral de como en el caso del muestreo aleatorio simple.

### **Muestreo estratificado**

Motivación, en el escenario 1 se plantea la situación el efecto económico que provoca la obesidad. Al estudiar este problema es evidente que en este caso la población no es homogénea, por ejemplo, considerando la edad como una característica que provoca esta heterogeneidad. Así cuando la población está formada por elementos no similares en referencia a alguna característica, se dice que es no homogénea en ese caso se procede a organizar grupos o estratos relativamente homogéneos. En el camino de estimación de algunos de los parámetros de interés se muestrea cada estrato mediante el muestreo aleatorio simple o sistemático.

Un ejemplo adicional, la aplicación del impuesto a la renta se pueden crear cuatro grupos, los que ganan menos de 5 mil pesos al mes, los que ganan entre 5 y 15 mil pesos, los que ganan más de 15 mil pesos y los que ganan más de 25 mil pesos.

### **Muestreo por conglomerados**

En el muestreo por conglomerados las unidades a observar no la forman individuos sino que los elementos están integrados por grupos. Una diferencia con el muestreo estratificado es que los cúmulos deben ser tan heterogéneos dentro del conglomerado y lo más homogéneo posible entre ellos. Así que la unidad de muestreo en un conglomerado, entonces la población la constituyen esas unidades. En

el procedimiento una vez que un conglomerado es seleccionado se analizan todos los elementos que lo componen. Las cuadras en una ciudad forman conglomerados y los hogares integran el conglomerado.

### Propiedades matemáticas de los estimadores

En este capítulo y en el anterior se ha expuesto algunos resultados sobre el proceso de estimación de parámetros. En general, los estimadores tienen varias propiedades matemáticas que es conveniente tener presentes a nivel de conocimiento, éstos deben ser insesgados, eficientes y consistentes. En esta sección se formaliza esta idea mediante algunas expresiones y en la figura 7.E2 se presentan e ilustran las características únicamente de los dos primeros. Inicialmente, se recuerda que en el capítulo 5 se planteó la expresión para la esperanza de una variable aleatoria, ésta es  $\mu = E(X)$ . Luego al principio de este capítulo se indicó de manera intuitiva que las medias  $\mu$   $\mu_{\bar{x}}$  son iguales, es decir:  $\mu = \mu_{\bar{x}}$ , de aquí se sigue que  $E(X) = \mu = \mu_{\bar{x}} = E(\bar{X})$ . Esto se justifica considerando una muestra aleatoria  $X_1, X_2, \dots, X_n$ , tal que  $\mu = E(X_1) = E(X_2) = \dots = E(X_n)$  y la varianza  $Var(X_1) = Var(X_2) = \dots = Var(X_n) = \sigma^2$ . Por lo tanto, se puede escribir:

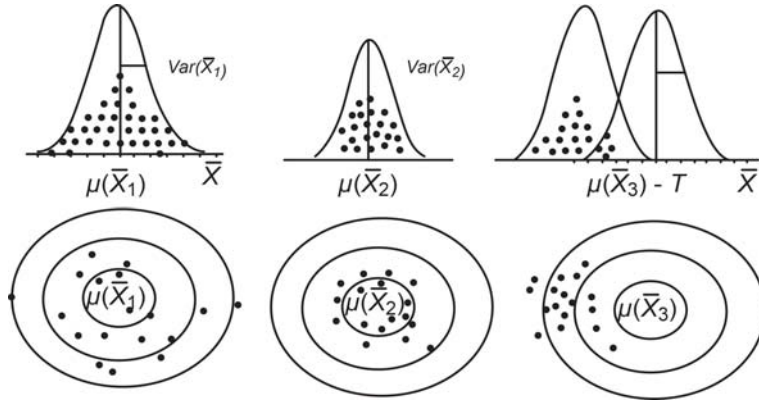
$$\begin{aligned}
 E(\bar{X}) &= E\left(\frac{X_1 + X_2 + \dots + X_n}{n}\right) \\
 &= \frac{1}{n} (E(X_1 + X_2 + \dots + X_n)) \\
 &= \frac{1}{n} (E(X_1) + E(X_2) + \dots + E(X_n)) \\
 &= \frac{1}{n} (n\mu) \\
 &= \mu
 \end{aligned} \tag{1}$$

Para la varianza

$$\begin{aligned}
 Var(\bar{X}) &= Var\left(\frac{X_1 + X_2 + \dots + X_n}{n}\right) \\
 &= \frac{1}{n^2} (Var(X_1 + X_2 + \dots + X_n)) \\
 &= \frac{1}{n^2} (Var(X_1) + Var(X_2) + \dots + Var(X_n)) \\
 &= \frac{1}{n^2} (nS^2) \\
 &= \frac{S^2}{n}
 \end{aligned} \tag{2}$$

Un estimador es *insesgado* si se cumple  $\mu = \mu_{\bar{x}} = E(\bar{X})$ , es decir que la media de la distribución muestral  $\bar{X}$  coincide con el parámetro  $\mu$ . En el primer caso de la figura 7.E2 se observa que los

valores de las medias de la variable  $\bar{X}_1$  se distribuyen como una normal con media  $\mu(\bar{X}_1)$  y varianza  $Var(\bar{X}_1)$ . En la segunda figura los valores de la media de la variable  $\bar{X}_2$  se distribuyen como una normal con media  $\mu(\bar{X}_2)$  y varianza  $Var(\bar{X}_2)$ . En ambos casos el estimador es insesgado, sin embargo en la segunda la varianza es menor al del primero,  $Var(\bar{X}_2) < Var(\bar{X}_1)$ , en esta situación se dice que el estimador  $\bar{X}_2$  es *más eficiente* que  $\bar{X}_1$ . En resumen: Un estimador se dice que es *más eficiente* si su varianza es menor que la varianza de otro estimador.



**Figura 7.E2** Estimadores insesgados, eficientes y sesgados.