

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO



COLEGIO DE CIENCIAS Y HUMANIDADES
PLANTEL VALLEJO

ÁREA DE MATEMÁTICAS

ESTADÍSTICA Y PROBABILIDAD II



UNIDAD II.

DISTRIBUCIONES MUESTRALES

En esta unidad trataremos las distribuciones muestrales, que nos sirve para estudiar una población desconocida tomándole muestras, y a través del estudio de las mismas poder hacer inferencias acerca de toda la población.

Introducción

Algunas de las razones para efectuar un muestreo a una población, en lugar de estudiarla directamente, pueden ser;

- *El tamaño de la población es infinito.*
- *El muestreo es de tipo destructivo.*
- *La población es finita, pero demasiado grande.*
- *Sería muy caro estudiar a toda la población y basta con deducciones aproximadas.*
- *Tomaría demasiado tiempo analizar la población total.*

Cuando las muestras son lo suficientemente grandes, se pueden hacer inferencias bastante exitosas, con pocos recursos, en comparación con técnicas más refinadas de la Estadística.

La *teoría del muestreo* tiene por objetivo, el estudio de las relaciones existentes entre la distribución de una característica en dicha población y las distribuciones de dicha característica en todas sus muestras.

Las ventajas de estudiar una población a partir de sus muestras son principalmente:

Costo reducido:

Si los datos que buscamos los podemos obtener a partir de una pequeña parte del total de la población, los gastos de recolección y tratamiento de los datos serán menores. Por ejemplo, cuando se realizan encuestas previas a una votación, es más barato preguntar a 400 personas su intención de voto, que a 250 000.

Mayor rapidez:

Actualmente nos estamos acostumbrados a ver los resultados del escrutinio de las casillas electorales, se obtiene una aproximación bastante buena del resultado final de unas elecciones, muchas horas antes de que el recuento final de votos haya finalizado;

Más posibilidades:

Para hacer cierto tipo de estudios, por ejemplo el de duración de cierto tipo de focos, no es posible en la práctica destruirlos todos para conocer su vida media, ya que no quedaría nada que vender. Es mejor destruir sólo una pequeña parte de ellas y sacar conclusiones sobre la duración de los demás.

Es decir, la teoría del muestreo es el estudio de las relaciones entre una población y las muestras que se extraen de ella.

Del análisis de las muestras se pueden *estimar o inferir* datos de la población como su media (μ), varianza (σ^2), etc., llamados *parámetros poblacionales*, denotados usualmente con letras griegas, a partir de los valores obtenidos de la muestra, tales como la media muestral \bar{X} , la varianza muestral s^2 .

Por ejemplo, en Bioquímica el verdadero nivel de glucosa de un paciente μ es siempre desconocido y además variable con el tiempo, por esto se le extrae una muestra de sangre para poder estimarlo a través de una medición. Así, el valor \bar{X} obtenido (ya sea con una o más muestras de sangre) es el que se pone como resultado del análisis clínico.

Cuando la población es finita y de un tamaño manejable en tiempo y costo, los valores poblacionales se calculan directamente, sin necesidad del muestreo.

Por ejemplo, si se trata de revisar diamantes, a nadie se le ocurriría tomar muestras sino que se controlarían uno por uno.

Las muestras extraídas deben ser representativas de la población en estudio para que las conclusiones obtenidas sean válidas. El estudio de los diferentes métodos de muestreo y las distintas variantes ocasionadas se llama *diseño de experimentos*.

Para extraer una muestra representativa se usa el método del *muestreo al azar o aleatorio*, que *asigna a todo integrante de la población la misma probabilidad de ser elegido*.

Cuando por lo menos uno de los elementos de la población tiene una probabilidad mayor (o menor) de ser seleccionado que el resto de los integrantes de la misma, entonces se trata de un método de *muestreo no aleatorio*.

Un sorteo es un método que hace equiprobables a todos los elementos de la población en el proceso de selección. Por ejemplo, se identifica cada elemento mediante un número, luego se colocan estos números en una urna, se mezclan bien, para luego ir sacando los números de la misma hasta completar el tamaño de muestra deseado.

DISTRIBUCIONES MUESTRALES

Como sabemos, cuando una medida tal como la media, la desviación estándar, la proporción, etc. describe cierto aspecto de una población, esta medida toma el nombre de **parámetro** de la población considerada. De igual manera, cuando una medida describe cierto aspecto de una muestra, se le da el nombre de **estadístico** de la muestra.

Un parámetro es una **cantidad constante**. Una vez determinado su valor, este valor es permanente, **mientras no cambie la población**.

De una muestra aleatoria de n valores, varios estadísticos podrían medirse. Usando estos n valores, se podría calcular una media, una mediana, un rango, una varianza, una desviación estándar, etc.

Ahora si en lugar de una muestra consideramos todas las muestras posibles del mismo tamaño, n , de una población y al calcular la media aritmética de cada muestra, algunas de las medias muestrales serían iguales; otras medias muestrales serían diferentes. Agrupando todas las medias muestrales que son semejantes y utilizando las reglas de probabilidad, podríamos determinar la probabilidad de obtener una muestra con una media particular.

Repitiendo este procedimiento para cada media posible de una muestra, se construye una distribución de probabilidad de la variable aleatoria “media muestral”.

*Una distribución de muestreo o muestral es una **distribución de probabilidad** donde la variable aleatoria es una estadística.*

Es decir: un estadístico es una **cantidad variable** ya que un mismo tipo de estadístico obtenido en diferentes muestras de una misma población, puede variar su valor de muestra a muestra, por tanto podemos decir que un estadístico es una **variable aleatoria** y su valor depende de los elementos que conforman la muestra.

Dado que un estadístico es una variable aleatoria que depende únicamente de la muestra observada, debe tener una distribución de probabilidad.

Definición. *La distribución de probabilidad de un estadístico recibe el nombre de **distribución muestral**.*

La distribución de probabilidad de \bar{X} se llama “distribución muestral de la media”.

La distribución de probabilidad de σ^2 se llama “distribución muestral de la varianza”.

La distribución muestral de un estadístico depende del tamaño de la población, del tamaño de la muestra y del método de selección de las muestras.

Tenemos que cualquier distribución de probabilidad (y por lo tanto cualquier distribución muestral) puede ser descrita por su media, su desviación estándar y su forma.

Antes de continuar establezcamos la simbología a utilizar:

Parámetros	Estadísticos
N: Tamaño de la población	n: tamaño de la muestra
μ : media de la población	\bar{x} : media de la muestra
σ : desviación estándar de la población	s: desviación estándar de la muestra
$\mu_{\bar{x}}$: media de la distribución muestral de la media (media de todas las muestras posibles de tamaño n)	$\sigma_{\bar{x}}$: desviación estándar de la distribución muestral de la media (error estándar de la media)

¿Cómo podemos obtener la descripción de una distribución muestral de un estadístico?

Para explicar este concepto vamos a obtener la descripción de la distribución muestral de la media para una población con pocos elementos.

DESCRIPCIÓN DE UNA DISTRIBUCIÓN MUESTRAL DE LA MEDIA

Para describir la distribución muestral de la media, es necesario seguir los siguientes pasos:

- 1.- Obtener todas las muestras posibles de tamaño n de la población de tamaño N
- 2.- Para cada muestra se calcula la media.
- 3.- Se construye la distribución de probabilidad para los valores de las medias
- 4.- Se calcula la media y la desviación estándar de la distribución de probabilidad.
- 5.- Se construye la gráfica de la distribución de probabilidad

Ejemplo: Una población está formada por los valores:

12, 6, 7, 10, 8, 11, 9, 13

Describir la distribución muestral de la media, al tomar todas las muestras de tamaño 3 **sin reemplazo**,

1 y 2.- Para determinar todas las muestras necesitamos primero saber el número de muestras posible: Si el muestreo es sin reemplazo el número de muestras está dado por:

$${}_N C_n = \frac{N!}{n!(N-n)!}$$

Para este ejemplo: $N=8$, $n = 3$, por lo que el número de muestras es:

$${}_8 C_3 = \frac{8!}{3!(8-3)!} = 56$$

Ordenando los datos tenemos: 6, 7, 8, 9, 10, 11, 12, 13

Entonces todas las muestras de tamaño 3 así como sus respectivas medias son:

No.	Muestra	media	No.	muestra	Media	No.	muestra	Media	No.	muestra	media
1	6 7 8	21/3	15	6 9 13	28/3	29	7 9 12	28/3	43	8 10 13	31/3
2	6 7 9	22/3	16	6 10 11	27/3	30	7 9 13	29/3	44	8 11 23	31/3
3	6 7 10	23/3	17	6 10 12	28/3	31	7 10 11	28/3	45	8 11 13	32/3
4	6 7 11	24/3	18	6 10 13	29/3	32	7 10 12	29/3	46	8 12 13	33/3
5	6 7 12	25/3	19	6 11 12	29/3	33	7 10 13	30/3	47	9 10 11	30/3
6	6 7 13	26/3	20	6 11 13	30/3	34	7 11 12	30/3	48	9 10 12	31/3
7	6 8 9	23/3	21	6 12 13	31/3	35	7 11 13	31/3	49	9 10 13	32/3
8	6 8 10	24/3	22	7 8 9	24/3	36	7 12 13	32/3	50	9 11 12	32/3
9	6 8 11	25/3	23	7 8 10	25/3	37	8 9 10	27/3	51	9 11 13	33/3
10	6 8 12	26/3	24	7 8 11	26/3	38	8 9 11	28/3	52	9 12 13	34/3
11	6 8 13	27/3	25	7 8 12	27/3	39	8 9 12	29/3	53	10 11 12	33/3
12	6 9 10	25/3	26	7 8 13	28/3	40	8 9 13	30/3	54	10 11 13	34/3
13	6 9 11	26/3	27	7 9 10	26/3	41	8 10 11	29/3	55	10 12 13	35/3
14	6 9 12	27/3	28	7 9 11	27/3	42	8 10 12	30/3	56	11 12 13	36/3

3.- Tomando como base las medias de las muestras podemos ahora construir la siguiente distribución de probabilidad

\bar{x}	$P(\bar{x})$
7	1/56
22/3	1/56
23/3	1/28
8	3/56
25/3	1/14
26/3	5/56
9	3/28
28/3	3/28
29/3	3/28
10	3/28
31/3	5/56
32/3	1/14
11	3/56
34/3	1/28
35/3	1/56
12	1/56
	1/1

4.- Calculemos ahora la media y la desviación estándar de la distribución muestral de la media

\bar{x}	$P(\bar{x})$	$\bar{x} \cdot P(\bar{x})$	\bar{x}^2	$\bar{x}^2 \cdot P(\bar{x})$
7	1/56	1/8	49	7/8
22/3	1/56	11/84	484/9	121/126
23/3	1/28	23/84	529/9	529/252
8	3/56	3/7	64	24/7
25/3	1/14	25/42	625/9	625/126
26/3	5/56	65/84	676/9	845/126
9	3/28	27/28	81	243/28
28/3	3/28	1	784/9	28/3
29/3	3/28	29/28	841/9	841/84
10	3/28	15/14	100	75/7
31/3	5/56	12/13	961/9	4805/504
32/3	1/14	16/21	1024/9	512/63
11	3/56	33/56	121	363/56
34/3	1/28	17/42	1156/9	289/63
35/3	1/56	5/24	1225/9	175/72
12	1/56	3/14	144	18/7
	1/1	19/2		183/2

Entonces tendremos que:

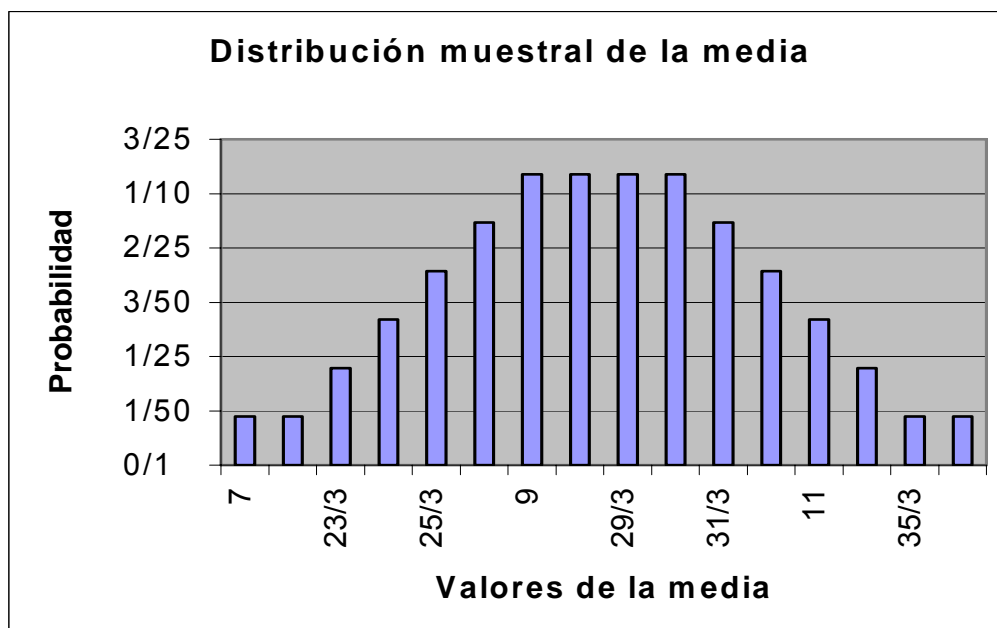
la media es igual a: $\mu_{\bar{x}} = \frac{19}{2}$

la varianza de la distribución es: $\sigma_{\bar{x}}^2 = \frac{183}{2} - \left(\frac{19}{2}\right)^2 = \frac{5}{4}$

y la desviación estándar es: $\sigma_{\bar{x}} = \sqrt{\frac{5}{4}} = \frac{\sqrt{5}}{2}$

5.- Tomando como base la distribución de probabilidad para la media construimos la gráfica de la distribución muestral de la media.

La gráfica de la distribución muestral de la media es:



Podemos observar que la grafica tiene una forma semejante (aproximada) a una distribución normal. Como se establecerá más, adelante esta aproximación será cada vez mejor a medida que aumente el tamaño de la muestra.

Si este problema lo intentamos resolver cuando el **muestreo es con reemplazo**, entonces tendremos que el número de muestras a extraer será de N^n , y por tanto para nuestro ejemplo, el número de muestras sería de: $8^3 = 512$. Realizar esta tarea requiere de mucho más de esfuerzo y tiempo.

Podemos observar que cuando la población es finita y de tamaño moderado podemos construir una distribución muestral, obteniendo todas las muestras posibles de un tamaño dado, calculando para cada muestra el valor del estadístico que nos interesa y enumerando los diferentes valores calculados junto con sus probabilidades de ocurrencia.

Podemos hacer una aproximación de las verdaderas distribuciones muestrales basadas en poblaciones muy grandes o infinitas, sacando un gran número de muestras aleatorias es decir, de una población cualquiera se extraen k muestras; cada una permite calcular k estadísticos con los cuales se puede hacer un histograma como el de la derecha de la figura 1.

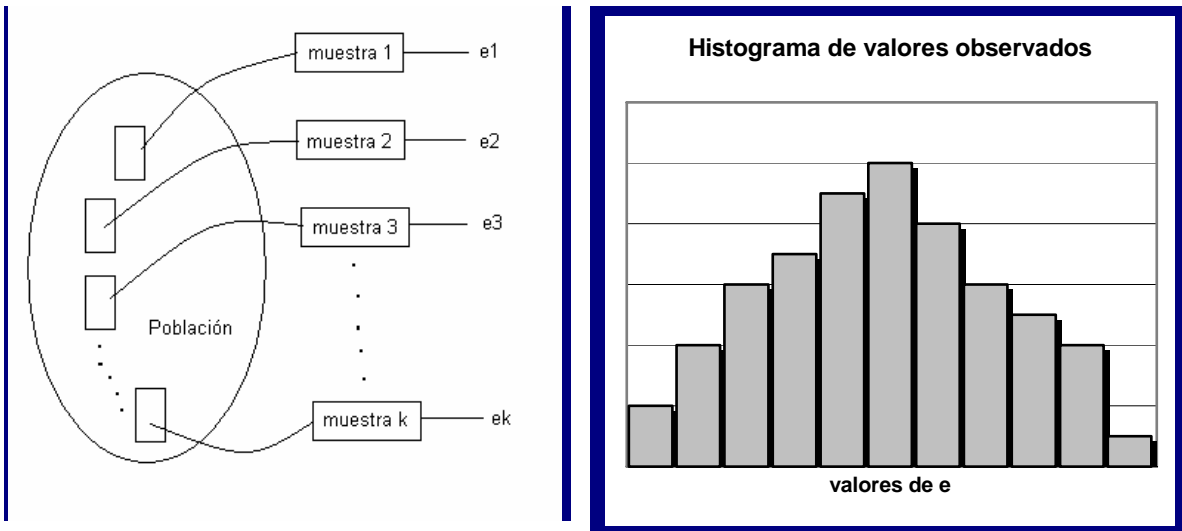


Figura 1

Se aprecia que el histograma adquiere forma de campana si se suavizan los escalones, al reducir los intervalos. La curva obtenida a partir de datos muestrales, observados a través del muestreo, tiende asintóticamente a otra curva teórica a medida que k aumenta, y los intervalos se hacen infinitesimales.

La pregunta sería ¿hay algún procedimiento más simple que nos permita obtener esta información, sin necesidad de realizar este laborioso proceso? La respuesta es si, dicho procedimiento esta basado en el Teorema del Límite Central

TEOREMA DEL LÍMITE CENTRAL

Este teorema es un principio fundamental de la teoría del muestreo, el llamado Teorema del Límite Central, establece que, a medida que crece el tamaño n de todas las posibles muestras que se pueden tomar de una población de tamaño N , que tiene como media μ y desviación estándar σ , la distribución muestral de las medias se aproxima a una distribución normal, con media μ y con desviación estándar $\frac{\sigma}{\sqrt{n}}$

De la interpretación del Teorema del Límite Central se pueden deducir los siguientes enunciados:

TEOREMA DEL LÍMITE CENTRAL

Dado que:

- 1.- La variable aleatoria X tiene una distribución (que puede ser o no normal) con media μ y desviación estándar σ .
- 2.- Se seleccionan muestras aleatorias de tamaño n de esa población.

Conclusiones:

- 1.- A medida que aumenta el tamaño de las muestras, la distribución de las medias de muestra \bar{x} se acercará a una distribución normal.
- 2.- La media de la distribución muestral de la media será igual a la media de la población μ

Es decir:

$$\mu_{\bar{x}} = \mu.$$

3.- La desviación estándar de las medias de muestra será:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

4.- Puesto que la distribución muestral de las medias de las muestras de tamaño n , se aproxima a una distribución normal, la probabilidad de que la media \bar{x} de una muestra cualquiera cumpla cierta condición, vienen dada por el área bajo la curva normal estándar, encontrándose el valor de z correspondiente a \bar{x} , por medio de la expresión:

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$$

5.- Al hacer un muestreo sin reemplazo, si el tamaño de la muestra n es **mayor al 5% del tamaño de la población finita** se debe ajustar la desviación estándar de las medias de la muestra multiplicándola por **el factor de corrección para población finita**:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

Reglas prácticas de uso común.

1.- Para muestras de tamaño n mayor que treinta ($n > 30$), la distribución muestral de la media se puede aproximar razonablemente bien con una distribución normal. La aproximación es más exacta a medida que aumenta el tamaño de la muestra n .

2.- Si la población original también está distribuida normalmente, las medias de muestra tendrán una distribución normal **para cualquier tamaño de muestra n .**

El Teorema Central del Límite permite establecer que, en condiciones muy generales, si la muestra es lo *suficientemente grande*, la distribución teórica de los valores obtenidos es aproximadamente normal.

Comprobemos con el teorema del límite central los resultados obtenidos en la distribución muestral de la media, del ejemplo anterior.

Calculemos la media y la desviación estándar de la población

La media es:
$$\mu = \frac{6+7+8+9+10+11+12+13}{8} = \frac{76}{8} = \frac{19}{2}$$

Luego la varianza es:
$$\sigma^2 = \frac{(6-\frac{19}{2})^2 + (7-\frac{19}{2})^2 + \dots + (13-\frac{19}{2})^2}{8} = \frac{21}{4}$$

Entonces su desviación estándar es:
$$\sigma = \sqrt{\frac{21}{4}} = \frac{\sqrt{21}}{2}$$

Utilizando el teorema del límite central tenemos que la media de la distribución muestral de la media es:

$$\mu_{\bar{x}} = \mu = \frac{19}{2}$$

La desviación estándar de la distribución muestral esta dada por (como la población es finita y el muestreo es sin reemplazo);

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} = \frac{\sqrt{21}/2}{\sqrt{3}} \sqrt{\frac{8-3}{8-1}} = \frac{1}{2} \sqrt{\frac{21(5)}{3(7)}} = \frac{\sqrt{5}}{2}$$

Como podemos observar los valores obtenidos en forma manual coinciden con los obtenidos con el teorema del límite central. Con esto verificamos o comprobamos la efectividad de dicho teorema.

Por lo que podemos concluir que la distribución muestral de medias, al realizar un muestreo aleatorio simple de tamaño 3 sin reemplazo sobre esta población, es

aproximadamente normal, con media $\mu_{\bar{x}} = \frac{19}{2}$ y desviación estándar de $\sigma_{\bar{x}} = \frac{\sqrt{5}}{2}$

Con esto hemos descubierto que: *el conocer la media y la varianza de la población el Teorema Central del Límite nos permitirá calcular la media y la varianza de la distribución muestral de las medias, sin la necesidad de generar la distribución muestral. Además sin importar la manera como se distribuya la población, la distribución de probabilidad de las medias muestrales, será siempre parecida a la distribución normal si el tamaño de la muestra es suficientemente grande. ($n \geq 30$).*

La semejanza con la distribución normal nos permite utilizar el modelo de distribución normal, para la describir aspectos de la distribución muestral de las medias, para muestras grandes.

El TCL nos da la razón para usar la tabla de distribución normal estándar cuando nuestra variable aleatoria es la media de la muestra \bar{X} .

Es decir si \bar{X} (media muestral) está distribuida normalmente (con $n \geq 30$) con media:

$\mu_{\bar{x}}$ y desviación estándar $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$, entonces:

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$$

está distribuida normalmente con una media igual a 0 y una desviación estándar igual a 1.

Por lo que si deseamos calcular la probabilidad, de que al obtener una muestra de una población finita o muy grande, de que la media muestral cumpla con ciertas condiciones o propiedades es necesario trabajar con una distribución normal con media y desviación estándar descritas por el Teorema del Límite Central.

APLICACIÓN DEL TEOREMA DEL LÍMITE CENTRAL PARA MEDIAS

Ejemplo 1.- Un investigador de mercados obtiene información respecto de una muestra aleatoria de 50 clientes de los 900 que adquieren una oferta especial determinada. Estas 50 personas gastaron en el almacén un promedio de \$ 7500 con una desviación estándar de \$ 900, ¿Cuál es la probabilidad de que en cualquier muestra de 50 de estas personas el gasto medio de compra esté entre \$ 7300 y \$ 7650?

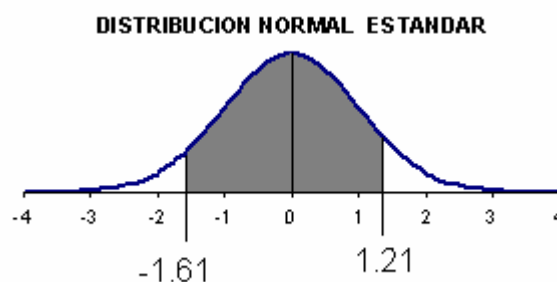
Solución:

$\mu = 7500$ Como la población es finita calculamos la desviación estándar de la
 $\sigma = 900$
 $N = 900$ distribución de medias como: $\sigma_{\bar{x}} = \frac{900}{\sqrt{50}} \sqrt{\frac{900-50}{900-1}} = 123.7619$
 $n = 50$

La probabilidad solicitada la podemos representar como: $P(7300 \leq \bar{x} \leq 7650)$, estandarizando los valores de 7300 y 7650 tenemos:

$$z_{7300} = \frac{7300 - 7500}{123.7619} = -1.61$$

$$z_{7650} = \frac{7650 - 7500}{123.7619} = 1.21$$



El área respuesta se encuentra entre estos dos puntos por lo tanto:

$$P(7300 \leq \bar{x} \leq 7650) = 0.44630 + 0.38686 = 0.83316$$

Entonces la probabilidad de que el gasto promedio de una muestra de 50 de estos consumidores, se encuentre entre el \$ 7300 y \$ 7650 es del 86.31%.

Ejemplo 2.- Un fabricante de cierto tipo de foco afirma que su producto tiene una duración media de 700 horas con una desviación estándar de 120 horas. Una persona ha comprado un ciento de focos de esta marca con la idea de comprar más si la duración media de la muestra excede de 680 horas. ¿Qué probabilidad hay de que dicha persona no vuelva a hacer sus compras con este fabricante?

La probabilidad de que no le compre es igual a la probabilidad de que la muestra de los 100 focos tenga una duración media por debajo de 680 horas, por tanto la respuesta es igual a esta probabilidad.

Solución:

$\mu = 700$ Debido a que la población la podemos considerar como infinita el valor de la
 $\sigma = 120$ desviación estándar de la distribución muestral la calcularemos como:
 $n = 100$ $\sigma_{\bar{x}} = \frac{120}{\sqrt{100}} = 12$

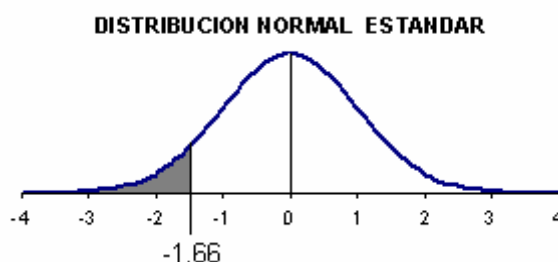
La probabilidad buscada la podemos representar como:

$$P(\bar{x} \leq 680)$$

estandarizando el valor de 680 tenemos:

$$z_{680} = \frac{680 - 700}{12} = -1.66$$

El área bajo la curva normal a calcular esta la izquierda de -1.66, entonces:



$$P(\bar{x} \leq 680) = 0.5 - 0.45154 = 0.04846$$

Podemos concluir que la probabilidad de que esta persona no vuelva a comprarle a este fabricante es del 4.84%

DISTRIBUCIÓN MUESTRAL DE PROPORCIONES

Cuando el estadístico obtenido en cada muestra es el porcentaje de ocurrencia de un suceso dado, se trata de una distribución de proporciones. Por ejemplo a un investigador de mercados puede estar interesado en conocer la proporción de consumidores de alguna zona prefieren los productos de su empresa. A un candidato a algún puesto político le gustaría conocer la proporción de votantes que van a votar por él, en las próximas elecciones.

Supóngase que tenemos una variable nominal dicotómica, es decir tenemos una población donde cada elemento puede ser clasificado como un éxito o un fracaso. Si el tamaño de una población finita es N , y en esa población hay K éxitos, habrá $(N - k)$ fracasos, La proporción de la población π , se define simplemente como la proporción de éxitos

$$\pi = \frac{k}{N}.$$

Si de esta población se selecciona una muestra de n elementos, la variable aleatoria x indicará el número de éxitos en la muestra, notemos que x puede tomar solamente valores enteros. La proporción de la muestra p es simplemente la proporción de éxitos en la muestra $p = \frac{x}{n}$.

Si x es el número de éxitos en una muestra en una muestra aleatoria, entonces:

$$0 \leq p \leq 1 .$$

Debido a que la proporción es una variable aleatoria, esperamos también que la distribución muestral de proporciones muestrales debe presentar características similares a la distribución muestral de medias.

Para comprobar esto realicemos un ejemplo donde describamos una distribución muestral de proporciones.

DESCRIPCIÓN DE UNA DISTRIBUCIÓN MUESTRAL DE LA PROPORCIÓN

Para describir la distribución muestral de la proporción, es necesario seguir los siguientes pasos:

- 1.- Obtener todas las muestras posibles de tamaño n de la población de tamaño N
- 2, Para cada muestra se calcula la proporción de la característica de interés.
- 3.- Se construye la distribución de probabilidad para los valores de las proporciones
- 4.- Se calcula la media y la desviación estándar de la distribución de probabilidad.
- 5.- Se construye la gráfica de la distribución de probabilidad

Ejemplo: Una población está formada por 4 hombres y 3 mujeres

Describir la distribución muestral de la proporción de hombres, al tomar todas las muestras de tamaño 3 **sin reemplazo**,

1 y 2.- Para determinar todas las muestras necesitamos primero saber el número de muestras posible: Si el muestreo es sin reemplazo el numero de muestras está dado por:

$${}_N C_n = \frac{N!}{n!(N-n)!}$$

Para este ejemplo: N=7, n = 3, por lo que el número de muestras es:

$${}_8 C_3 = \frac{7!}{3!(7-3)!} = 35$$

Entonces todas las muestras de tamaño 3 así como sus respectivas proporciones de hombres son:

No	Muestra	p	No	muestra	p	No	muestra	p	No.	muestra	P
1	H H H	1	10	H H M	2/3	19	H H M	2/3	28	H H M	2/3
2	H H H	1	11	H H M	2/3	20	H H M	2/3	29	H M M	1/3
3	H H M	2/3	12	H H M	2/3	21	H H M	2/3	30	H M M	1/3
4	H H M	2/3	13	H M M	1/3	22	H H M	2/3	31	H M M	1/3
5	H H M	2/3	14	H M M	1/3	23	H M M	1/3	32	H M M	1/3
6	H H H	1	15	H M M	1/3	24	H M M	1/3	33	H M M	1/3
7	H H M	2/3	16	H H H	1	25	H M M	1/3	34	H H M	1/3
8	H H M	2/3	17	H H M	2/3	26	H H M	2/3	35	M M M	0
9	H H M	2/3	18	H H M	2/3	27	H H M	2/3			

3.- Tomando como base las proporciones de hombres en las muestras podemos ahora construir la siguiente distribución de probabilidad

p	P(p)
0	1/35
1/3	12/35
2/3	18/35
3/3	4/35
	1

4.- Calculemos ahora la media y la desviación estándar de la distribución muestral de proporciones

p	P(p)	x P(X)	X ²	x ² P(x)
0	1/35	0	0	0
1/3	12/35	4/35	1/9	4/105
2/3	18/35	12/35	4/9	8/35
1	4/35	4/35	1	4/35
	1.0	4/7		8/21

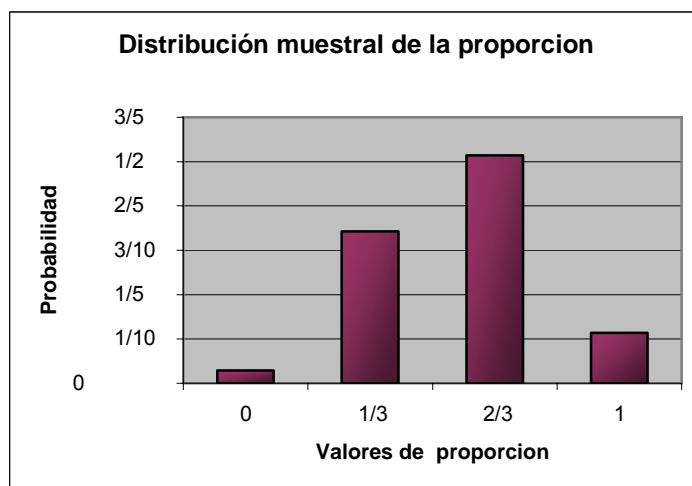
Entonces tendremos que:

la media de las proporciones es igual a: $\mu_p = \frac{4}{7}$

la varianza de la distribución de proporciones es: $\sigma_p^2 = \frac{8}{21} - \left(\frac{4}{7}\right)^2 = \frac{8}{147}$

y la desviación estándar es: $\sigma_{\bar{x}} = \sqrt{\frac{8}{147}}$

5.- Tomando como base la distribución de probabilidad para la media construimos la gráfica de la distribución muestral de la proporción.



Podemos observar que la grafica tiene una forma algo parecida a una distribución normal y esta aproximación será cada vez mejor a medida que aumente el tamaño de la muestra.

De nueva cuenta ¿Existe algún procedimiento sencillo y rápido que nos permita determinar esta información, sin necesidad de realizar este tedioso proceso?

Nuevamente la respuesta es sí, y dicho proceso está basado en Teorema del Límite Central para proporciones.

El teorema del límite Central para las proporciones establece:

TEOREMA CENTRAL DEL LÍMITE

Dado que:

- 1.- La variable aleatoria X tiene una distribución binomial con parámetro π , y
- 2.- Se seleccionan muestras aleatorias de tamaño n de esa población.

Conclusiones:

- 1.- A medida que aumenta el tamaño de las muestras, la distribución de las proporciones de muestra p se acercará a una distribución normal.
- 2.- La media de la distribución muestral de la proporción será igual a la proporción de la población π

Es decir:

$$\mu_p = \pi$$

- 3.- La desviación estándar de las proporciones de muestra será:

$$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}}$$

4.- Puesto que la distribución muestral de las proporciones de las muestras de tamaño n , se aproxima a una distribución normal, la probabilidad de que la media p de una muestra cualquiera cumpla cierta condición, viene dada por el área bajo la curva normal estándar, encontrándose el valor de z correspondiente a p , por medio de la expresión:

$$z = \frac{p - \mu_p}{\sigma_p}$$

5.- Al hacer un muestreo sin reemplazo, si el tamaño de la muestra n es **mayor al 5% del tamaño de la población finita** se debe ajustar la desviación estándar de las medias de la muestra multiplicándola por **el factor de corrección para población finita**:

$$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}} \sqrt{\frac{N-n}{N-1}}$$

Reglas prácticas de uso común.

1.- Para muestras de tamaño n con $np > 5$, y $np(1-p) > 5$ la distribución muestral de la proporción se puede aproximar razonablemente bien con una distribución normal. La aproximación es más exacta a medida que aumenta el tamaño de la muestra n .

Para comprobar la validez del Teorema del Límite Central, verifiquemos los resultados obtenidos manualmente en el problema anterior.

Como nuestra población está formada por 4 hombres y tres mujeres, tendremos:

$N = 7$, $x = 4$ por lo cual la proporción poblacional es: $\pi = \frac{4}{7}$

Como sabemos la desviación estándar de la población está dada por:

$$\sigma = \sqrt{\pi(1-\pi)} = \sqrt{\frac{4}{7}\left(1-\frac{4}{7}\right)} = \sqrt{\frac{12}{49}}$$

Aplicando el teorema del Límite Central para las proporciones, tendremos:

$$\mu_{\hat{p}} = \frac{4}{7} = \pi$$

Como realizamos un muestreo sin reemplazo tendremos que para calcular la desviación estándar de la distribución muestral de proporciones mediante la expresión que incluye el factor de corrección.

$$\sigma_{\hat{p}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

Sustituyendo valores:

$$\sigma_{\hat{p}} = \frac{\sqrt{\frac{12}{49}}}{\sqrt{3}} \sqrt{\frac{7-3}{7-1}} = \sqrt{\frac{12}{147}} \sqrt{\frac{4}{6}} = \sqrt{\frac{8}{47}}$$

Que efectivamente concuerdan con los resultados obtenidos anteriormente.

Por lo que podemos concluir que la distribución muestral de proporciones, al realizar un muestreo aleatorio simple de tamaño 3 sin reemplazo sobre esta población, es

aproximadamente normal, con media $\mu_{\hat{p}} = \frac{4}{7}$ y desviación estándar de $\sigma_{\hat{p}} = \sqrt{\frac{8}{47}}$

Con esto hemos comprobado que: *si conocemos la proporción de la población el Teorema del Límite Central nos permitirá calcular la media y la varianza de la distribución muestral de las proporciones, sin la necesidad de generar la distribución muestral.*

La semejanza con la distribución normal nos permite utilizar, el modelo normal, para la distribución muestral de las proporciones muestrales para muestras grandes.

El TLC nos da la razón para usar la tabla de distribución normal estándar cuando nuestra variable aleatoria es la proporción de muestra p .

Es decir si p esta distribuida normalmente con media π y desviación estándar σ_p , entonces:

$$z = \frac{p - \pi}{\sigma_{\hat{p}}}$$

está distribuida normalmente con una media igual a 0 y una desviación estándar igual a 1.

Por lo que si deseamos calcular la probabilidad, de que al obtener una muestra de una población finita o muy grande, la proporción muestral cumpla con ciertas condiciones o propiedades es necesario trabajar con una distribución normal con media y desviación estándar descritas por el Teorema del Límite Central para la proporción.

APLICACIÓN DEL TEOREMA CENTRAL DEL LÍMITE PARA PROPORCIONES

Ejemplo 1.- En un proceso para llenar botellas de refresco presenta una producción promedio en la que el 7% de las botellas no están completamente llenas. Si de este proceso se selecciona al azar una muestra de 225 botellas de un lote de 625 envases llenos, ¿Cuál es la probabilidad de que la proporción muestral de botellas parcialmente llenas

a) este entre 9% y 11%

b) menor al 8%

Solución:

$\pi = 0.07$ Debido a que la población es finita
 $N = 625$ calculamos la desviación estándar de la
 $n = 225$ distribución de proporciones como:

$$\sigma_p = \sqrt{\frac{0.07(1-0.07)}{225}} \sqrt{\frac{625-225}{625-1}} = 0.0136$$

La probabilidad solicitada la podemos representar como: $P(0.09 \leq p \leq 0.11)$,

Estandarizando los valores de 0.09 y 0.11 tenemos:

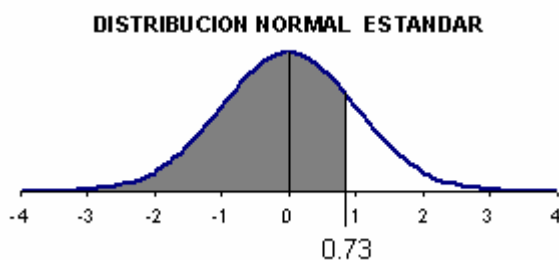
$$z_{.09} = \frac{0.09 - 0.07}{0.0136} = 1.47 \quad z_{.11} = \frac{0.11 - 0.07}{0.0136} = 2.94$$

El área que nos interesa se encuentra entre estos dos puntos por lo tanto:

$$P(0.09 \leq p \leq 0.11) = 0.49836 - 0.42922 = 0.06464$$

Entonces la probabilidad de que el porcentaje de botellas parcialmente llenas se encuentre entre el 9% y el 11% es del 6.46%.

b) La probabilidad solicitada la podemos representar como: $P(p \leq 0.08)$, estandarizando el valor de 0.08



$$z_{.08} = \frac{0.08 - 0.07}{0.0136} = 0.73$$

El área que nos interesa se encuentra a la izquierda de este valor, por lo que:

$$P(p \leq 0.08) = 0.5 + 0.26730 = 0.76730$$

Es decir la probabilidad de que el porcentaje de botellas sea menor del 8% es

igual al 76.73%

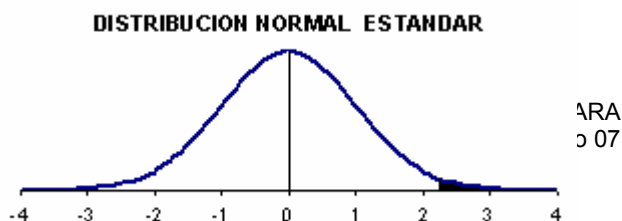
Ejemplo 2.- El departamento de compras de una gran compañía rechaza, por sistema, remesas de refacciones si en una muestra aleatoria de 100 de un lote de 10 000 partes presenta 10 o más productos defectuosos. Encuentre la probabilidad de que se rechace el lote si se tiene un porcentaje de producción del:

a) 5%,

b) 12%

Solución: Como la población es finita

$\pi = 0.04$
 $N = 10000$
 $n = 100$



calculamos la desviación estándar de la distribución de proporciones como:

$$\sigma_p = \sqrt{\frac{0.05(1-0.05)}{100}} \sqrt{\frac{10000-100}{10000-1}} = 0.0216$$

La probabilidad solicitada la podemos representar como: $P(p \geq 0.10)$, estandarizando el valor de 0.10 tenemos:

$$z_{.10} = \frac{0.10-0.05}{0.0216} = 2.31$$

El área bajo la curva normal a calcular esta la derecha de 2.31,

$$P(p \geq 0.10) = 0.5 - 0.48956 = 0.01044$$

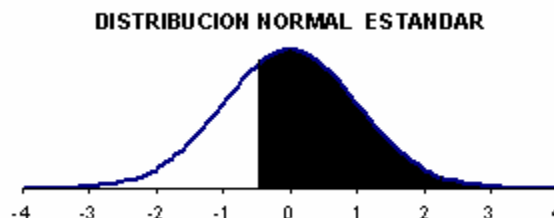
Es decir se tienen una probabilidad del 1.04% de rechazar este lote con las normas establecidas.

b) Debido a que solo cambia el porcentaje de defectuosos en la producción, tendremos que:

$$\sigma_p = \sqrt{\frac{0.12(1-0.12)}{100}} \sqrt{\frac{10000-100}{10000-1}} = 0.0269$$

Ahora el valor de estandarizado de 0.10

$$\text{es: } z_{.10} = \frac{0.10-0.12}{0.0323} = -0.52$$



Como nos interesa el área a la derecha tendremos:

$$P(p \geq 0.10) = 0.5 + 0.19847 = 0.69847$$

Es decir se tienen una probabilidad del 69.84%% de rechazar este lote cuando el porcentaje de defectuosos en la producción es del 12%.

Ejemplo 3.- Un fabricante de autos compra las tuercas necesarias para el armado a un proveedor que afirma entregar remesas con aproximadamente un 6% de tuercas defectuosas. Calcular la probabilidad de que en una remesa de 400 tuercas se obtenga:

- más de 15 tuercas defectuosas
- entre 20 y 30 defectuosas.

Solución: Debido a que se mencionan número de piezas defectuosas, es necesario expresar estas piezas defectuosas como un porcentaje respecto al total de la remesa

mediante la expresión $p = \frac{x}{n}$

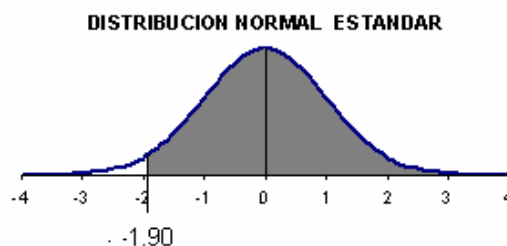
Debemos calcular la desviación estándar de la distribución muestral de proporciones:

$$\pi = 0.06 \quad n = 400 \quad \sigma_p = \sqrt{\frac{0.06(1-0.06)}{400}} = 0.0118$$

a) 15 piezas es igual al $p = \frac{15}{400} = 0.0375$, por lo que

la probabilidad que buscamos es: $P(p \geq 0.0375)$, estandarizando el valor 0.375 tenemos:

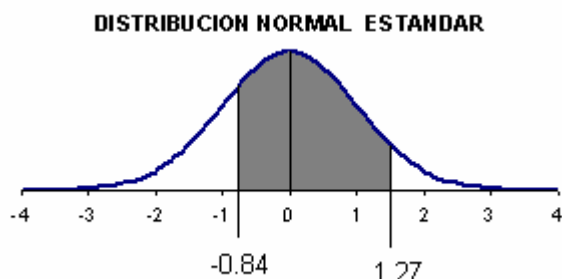
$$z_{.0375} = \frac{0.0375-0.06}{0.0118} = -1.90$$



El área que nos interesa se encuentra a la derecha de este valor, por lo que:

$$P(p \geq 0.035) = 0.5 + 0.47128 = 0.97128$$

b) 20 piezas es igual al $p = \frac{20}{400} = 0.05$, y 30 tuercas son igual al porcentaje



$p = \frac{30}{400} = 0.075$ por lo que la probabilidad que buscamos esta dada por:
 $P(0.05 \leq p \leq 0.075)$, estandarizando estos valores tenemos:

$$z_{.05} = \frac{0.05 - 0.06}{0.0118} = -0.84$$

$$z_{.075} = \frac{0.075 - 0.06}{0.0118} = 1.27$$

El área que nos interesa se encuentra entre estos valores, por lo que:

$$P(0.05 \leq p \leq 0.075) = 0.29955 + 0.39796 = 0.69751$$

Es decir se tienen una probabilidad del 69.75% de que el numero de defectuosos este entre 20 y 30 tuercas defectuosas, cuando el porcentaje de defectuosos en la producción de tuercas es del 6%.

PROBLEMAS SOBRE DISTRIBUCIÓN MUESTRAL DE MEDIAS Y PROPORCIONES

- 1.- Los puntajes en facilidad en lectura de los niños de un jardín infantil están normalmente distribuidos con una media y una desviación típica de 75 y 10 respectivamente. ¿Cuál es la probabilidad de que en una muestra de 25 niños arroje un puntaje: a) entre 70 y 78, b) al menos 80, c) no más de 73?.
- 2.- El número promedio de años de experiencia en conducir de cierto grupo de camioneros es de 10 años y la desviación típica de 3 años. ¿Cuál es la probabilidad de que una muestra aleatoria de 81 de estos camioneros tenga una media: a) mayor a 10 años 8 meses, b) entre 9 y 9 años 9 meses c) a lo mas de 9 años y medio?.
- 3.- En una población de 1200 adolescentes la cantidad promedio de dinero gastado en recreación por semana es de \$125 y desviación típica de \$ 26, ¿Cuál es la probabilidad de que en una muestra de 40 de estos adolescentes tenga una media: a) de a lo más de \$120, b) entre \$130 y \$140, c) al menos de 135.
- 4.- En un área metropolitana el 18% de los adolescentes ha tenido algún contacto con la policía por motivo de delincuencia juvenil. Se selecciona una muestra aleatoria de 100 de estos adolescentes, ¿Cuál es la probabilidad de que a) entre el 12% y 20%, b) más del 25%, c) a o más el 22% haya tenido contacto con la policía por las causas anotadas?
- 5.- Se ha reportado que en una determinada población de estudiantes el 40% ha fumado marihuana alguna vez en su vida. Si se toma una muestra de 150 de estos estudiantes,

cual es la probabilidad de que: a) a lo más la mitad, b) entre el 45% y el 55% c) más del 35% haya fumado marihuana.

6.- De los 1150 profesores que se emplean en una zona escolar, el 60% acreditan una maestría. Se selecciona una muestra aleatoria de 90 de estos profesores. ¿Cuál es la probabilidad de que la proporción de profesores con grado de maestría: a) sea mayor al 65%, b) este entre el 50% y el 64%, c) sea a lo más del 70%.

7.- La duración promedio del mezclador de un cierto fabricante es de 5 años, con una desviación estándar de 1 año. Considerando que las duraciones de estos mezcladores, siguen una distribución aproximadamente normal, encuentre la probabilidad de que en una muestra de 36 de estos mezcladores; :

a) sea al menos 4.5 años. b) este entre 4 años 9 meses y 5 años 4 meses.

8.- El promedio de tiempo que un cajero de banco atiende a un cliente es una variable aleatoria con una media de $\mu = 4.2$ y una desviación estándar de 1.8 minutos. Si se observa una muestra aleatoria de 50 clientes , encuentre la probabilidad de que el tiempo promedio de los clientes con el cajero sea:

a) Cuando mucho 3.7 minutos, b) más de 4.5 minutos, c) al menos 3.3 minutos pero menos de 4 minutos.

9.- Las calificaciones promedio para estudiantes de primer año en una prueba de aptitudes, en cierta universidad, es de 640 con una desviación estándar de 60. ¿Cuál es probabilidad de que en una muestra seleccionada aleatoriamente de 70 estudiantes, tenga una calificación promedio;

a) mayor a 660 puntos, b) de al menos de 630, c) entre 635 y 655.

10.- Si el porcentaje de hogares en los que cuando menos hay un adulto desempleado, es del 12%, en cierta área metropolitana, calcule la probabilidad de que al obtener una muestra de 100 de estos hogares el porcentaje de hogares con adultos desempleados:

a) sea al menos del 15%, b) esté entre el 10% y el 20%

c) sea cuando más del 10%.

11.- Un pequeño fabricante, adquiere un lote de 200 partes electrónicas, del “exceso de inventarios” de una empresa grande. Si se sabe que el 18% de estas piezas son defectuosas, calcule la probabilidad de que el numero de piezas defectuosas en el lote;

a) sea a lo más de 40 b) este entre 30 y 50 , c) sea al menos de 25.

12.- Se sabe que el 65% de las personas de clase media prefieren el ron para las bebidas combinadas con refresco (“cubas”), calcule la probabilidad de que en una muestra de 300 adultos de clase media, a) más de 180 prefieran el ron, b) a lo más 200 prefieran el ron, c) entre 190 y 210 prefieran el ron