CAPÍTULO 6

DISTRIBUCIONES MUESTRALES

Uno de los objetivos de la estadística es saber acerca del comportamiento de parámetros poblacionales tales como: la media (μ), la varianza (σ^2) o la proporción (p). Para ello se extrae una muestra aleatoria de la población y se calcula el valor de un estadístico correspondiente, por ejemplo, la media muestral ($\overline{X}$), la varianza muestral (s^2) o la proporción muestral ($\hat{p}$). El valor del estadístico es aleatorio porque depende de los elementos elegidos en la muestra seleccionada y, por lo tanto, el estadístico tiene una distribución de probabilidad la cual es llamada la Distribución Muestral del Estadístico.

6.1 Distribución de la Media Muestral cuando la población es normal

Si se estraen muestras aleatorias de tamaño n de una población infinita que tiene media poblacional μ y varianza σ^2, entonces sea cual sea la distribución de la población se tiene que:

i) La media de las medias muestrales es igual a la media poblacional. Es decir $\mu_{\bar{x}} = \mu$.

ii) La varianza de las medias muestrales es igual a la varianza poblacional dividida por n. En consecuencia la desviación estándar de las medias muestrales (llamada también el **error estándar** de la media muestral), es igual a la deviación estándar poblacional dividida por la raíz cuadrada de n. Es decir $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$.

Si la población fuera finita de tamaño N, entonces se aplica el factor de correción $\sqrt{\dfrac{N-n}{N-1}}$ al error estándar de la media muestral. Pero en la práctica este factor es omitido a menos que la muestra sea lo suficientemente grande comparada con la población.

Si además la población se distribuye normalmente, entonces la media muestral también tiene una distribución normal con la media y varianza anteriormente indicadas.

6.2 El Teorema del Límite Central

Un importante resultado en Probabilidades y Estadística es el llamado ***Teorema del Límite Central*** que dice que si de una población infinita con media μ y varianza σ^2 se extraen muestras aleatorias de tamaño n, entonces la media muestral se comporta aproximadamente como una variable aleatoria normal con media igual a la media poblacional y con varianza igual a la varianza poblacional dividida por el tamaño de la muestra, siempre que n sea grande. Lo importante de este resultado es que es independiente de la forma de la distribución de la población. Es decir,

$$\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$$

Estandarizando, esto es equivalente a:

$$Z = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

Si la población es bastante simétrica entonces, un tamaño de muestra n mayor que 30 es suficiente para una buena aproximación a la normal. Si la población es bastante asimétrica, entonces el tamaño de muestra debe ser mucho más grande.

En **MINITAB** se puede tratar de corroborar el Teorema del Límite Central a través de un proceso de simulación.

Ejemplo 6.1 Considerar una población que consiste de 3, 4, 6, 8, 10, 11, 12, 15, 20. Primero calculamos la media y desviación estándar de dicha población.

Descriptive Statistics

Variable	N	Mean	Median	Tr Mean	StDev	SE Mean
C1	9	9.89	10.00	9.89	5.42	1.81

Variable	Min	Max	Q1	Q3
C1	3.00	20.00	5.00	13.50

Notar que $\mu = 9.89$ y $\sigma = 5.42$.

Segundo, extraemos 30 muestras de tamaño 4 de dicha población, ejecutando 4 veces la siguiente secuencia **Calc** ▸ **Random Data** ▸ **Sample from columns.** Guardar cada una de las 4 observaciones de las muestras en 4 columnas distintas: *Obs1, Obs2, Obs3, y Obs4*.

Tercero, calculamos las medias de todas esas muestras usando la opción **Row Statistics** del menú **Calc** y tratamos de ver gráficamente al menos si hay acercamiento a Normalidad. Asimismo se debe observar que la media de todas estas medias debería estar cerca de μ y la varianza cerca de σ^2/n.

Las 30 muestras elegidas y sus respectivas medias son:

Muestra obs1 obs2 obs3 obs4 media

1	6	4	3	8	5.25
2	11	8	4	3	6.50
3	3	3	15	3	6.00
4	10	8	10	6	8.50
5	15	12	11	8	11.50
6	4	12	6	6	7.00
7	12	11	20	10	13.25
8	12	8	20	12	13.00
9	8	10	12	11	10.25
10	8	20	11	20	14.75
11	20	10	6	8	11.00
12	11	10	12	12	11.25
13	11	3	8	11	8.25
14	3	10	11	4	7.00
15	20	12	20	3	13.75
16	20	3	15	11	12.25
17	12	20	20	15	16.75
18	3	3	11	20	9.25
19	20	11	10	15	14.00
20	11	3	11	15	10.00
21	6	8	6	15	8.75
22	11	3	12	6	8.00
23	10	8	3	20	10.25
24	6	20	12	6	11.00
25	15	6	4	12	9.25
26	11	10	3	4	7.00
27	11	11	11	11	11.00
28	10	10	6	10	9.00
29	4	20	20	3	11.75
30	11	6	6	8	7.75

Las medidas estadísticas de la media muestral son:

```
Variable          N      Mean    Median  Tr Mean     StDev  SE Mean
media            30    10.108    10.125   10.019     2.806    0.512

Variable        Min       Max        Q1       Q3
media         5.250    16.750     7.938   11.875
```

En la Figura 6.1 se muestra el histograma de la distribución de las medias muestrales y la **curva normal** que más se aproxima al histograma.

***Interpretación**: Notar que la media de las medias muestrales es* $\mu_{\bar{x}} = 10.108$ *que está bien cerca de la media poblacional* $\mu = 9.89$. *Además la desviación estándar de la media muestral es 2.806 mientras que* $\sigma/\sqrt{n}$ *es igual a 5.42/2=2.71 ambos valores también están relativamente cerca. El histograma si está un poco alejado de la normalidad.*

Si se incrementa el tamaño de las muestras se puede notar una mejor aproximación a la Normal.

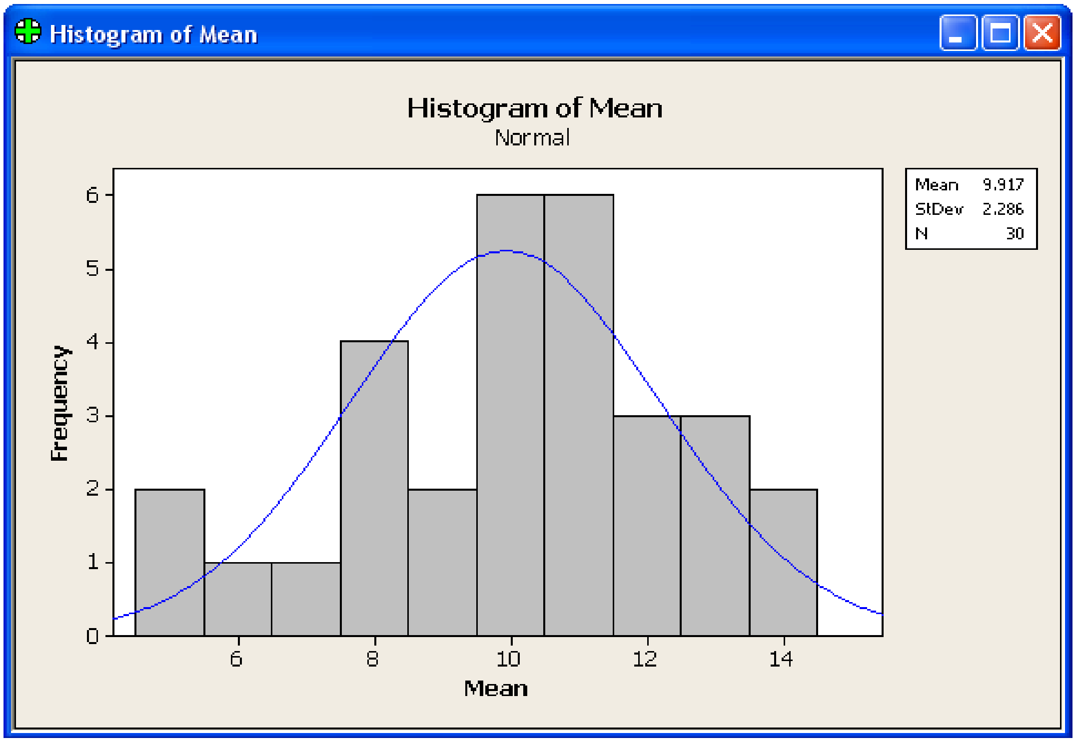

Figura 6.1 Histograma de la distibución de las medias maestrales del Ejemplo 6.1

Luego de aplicar estandarización, las siguientes fórmulas se cumplen, aproximadamente si la población no es normal y exactamente si lo es.

i) $$P(\overline{X} < a) = P\left(Z < \frac{a-\mu}{\frac{\sigma}{\sqrt{n}}}\right)$$

ii) $$P(a < \overline{X} < b) = P\left(\frac{a-\mu}{\frac{\sigma}{\sqrt{n}}} < Z < \frac{b-\mu}{\frac{\sigma}{\sqrt{n}}}\right)$$

iii) $$P(\overline{X} > b) = P\left(Z > \frac{b-\mu}{\frac{\sigma}{\sqrt{n}}}\right)$$

Las probabilidades pueden ser calculadas usando la tabla de la normal estándar que aparece en al apéndice del texto. Sin embargo, éstas pueden ser halladas directamente en **MINITAB** sin necesidad de estandarización.

Ejemplo 6.2. El tiempo de atención por cliente de un cajero de un Banco es normal con media 6 minutos y desviación estándar 2.5 minutos.

a) ¿Cuál es la probabilidad de que el tiempo promedio de atención para una muestra de 15 clientes sea menor de 7 minutos?
b) ¿Cuál es la probabilidad de que el tiempo de atención a un grupo de 15 clientes sea más de una hora y 15 minutos?
c) Si el tiempo en que el cajero atiende a un grupo de 15 clientes excede las dos horas entonces éste es despedido. ¿Cuál es la probabilidad de que esto ocurra?

Solución:
Usando el hecho que el tiempo promedio de atención para una muestra de tamaño 15 es normal con media 6, y desviación estándar $\frac{2.5}{\sqrt{15}} = 0.645$, con la ayuda de **MINITAB** se obtiene:

a) $P(\overline{X} < 7) = 0.9395$
b) Un tiempo de atención de 75 minutos a 15 clientes equivale a un tiempo promedio de atención de 75/15 = 5 minutos. Luego, hay que hallar $P(\overline{X} > 5) = 1 - 0.0605 = 0.9395$
c) Un tiempo de atención de 120 minutos a 15 clientes equivale a un tiempo promedio de atención de 120/15 = 8 minutos por cliente. Luego, hay que hallar $P(\overline{X} > 8) =$ 1 - 0.9990 = .001.

Ejemplo 6.3. Los pesos de las personas que suben a un ascensor se distribuyen normalmente con media igual a 125 libras y desviación estándar de 30 libras. Un grupo de 9 personas sube al ascensor:
a) ¿Cuál es la probabilidad de que el peso promedio del grupo sea menor de 100 libras?
b) El ascensor tiene una capacidad máxima de 1400 libras. ¿Cuál es la probabilidad de que se exceda ésta capacidad con un grupo de 9 personas?

Solución:
a) El peso promedio de un grupo de 9 personas se distribuye normalmente con media 125 y desviación estándar igual a $\frac{30}{\sqrt{9}} = 10$. Luego usando la secuencia **Calc** ▸ **Probability Distributions** ▸ **Normal** en **MINITAB** se obtiene que $P(\overline{X} < 100) = 0.0062$.
b) Decir que la suma de los pesos del grupo sea mayor que 1400, equivale a que el peso promedio del grupo de 9 personas sea mayor que 1400/9 = 166.66 libras. Luego, la probabilidad pedida será $P(\overline{X} > 166.66) = 1 - P(\overline{X} < 166.66) = 1 - .9989 = 0.0011$.

6.3 Distribución de la Proporción Muestral

Si de una población distribuida Binomialmente con probabilidad de éxito p, se extrae una muestra aleatoria de tamaño n, entonces se puede mostrar que la media de *X*:

número de éxitos en la muestra, es $\mu = np$ y que su varianza es $\sigma^2 = npq$. En consecuencia la proporción muestral $\hat{p} = \frac{X}{n}$ tiene media p, y varianza $\frac{pq}{n}$. Así, por el Teorema del Limite Central, cuando el tamaño de muestra es grande, entonces:

$$z = \frac{X - np}{\sqrt{npq}} = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$

Se distribuye aproximadamente como una normal estándar. La aproximación es bastante confiable si tanto $n\hat{p}$ como $n\hat{q}$ son mayores que 5. Cuando $\hat{p}$ es cercano a 0 ó 1 se debe tomar un tamaño de muestra más grande para mejorar la aproximación.

Asímismo, como se están aproximando probabilidades de una distribución discreta por probabilidades de una distribución contínua, se debe aplicar un ***Factor de Corrección por Continuidad*** de 1/2, antes de calcular las probabilidades. Este 1/2 se explica porque un valor entero k de la variable discreta representa a todos los valores de la variable continua que caen en el intervalo $(k - 1/2, k + 1/2)$. Cuando el tamaño de muestra es bien grande entonces el efecto de considerar el factor de corrección por continuidad es insignificante.

Fórmulas de aproximación Normal a la Binomial.

Si X es una Binomial con parámetros n y p, entonces

i) $P(X = k) \cong P(k - .5 < X < k + .5) = P(\frac{k - .5 - np}{\sqrt{npq}} < Z < \frac{k + .5 - np}{\sqrt{npq}})$

ii) $P(a < X < b) = P(a + .5 < X < b - .5) = P(\frac{a + .5 - np}{\sqrt{npq}} < Z < \frac{b - .5 - np}{\sqrt{npq}})$

iii) $P(a \le X \le b) = P(a - .5 < X < b + .5) = P(\frac{a - .5 - np}{\sqrt{npq}} < Z < \frac{b + .5 - np}{\sqrt{npq}})$

Similarmente se pueden definir fórmulas para aproximar probabilidades para proporciones muestrales.

Ejemplo 6.4. Según reportes del centro nacional para estadísticas de salud, alrededor del 20 % de la población masculina adulta de los Estados Unidos es obesa. Se elige al azar una muestra de 150 hombres adultos en los Estados Unidos. ¿Cuál es la probabilidad de que:

a) Haya a lo más 25 personas obesas?
b) Haya más de 22 pero menos de 35 obesos?
c) Haya por lo menos un 25% de obesos en la muestra?

Solución:
Usando aproximación normal a la Binomial se tiene que:

a) $P(X \le 25) \cong P(X < 25.5) = P\left(Z < \dfrac{25.5-30}{\sqrt{24}}\right) = P(Z < -0.91) = 0.1814$

b) $P(22 < X < 35) \cong P(22.5 < x < 34.5) = P\left(\dfrac{22.5-30}{\sqrt{24}} < Z < \dfrac{34.5-30}{\sqrt{24}}\right) =$

$P(-1.53 < Z < 0.91) = 0.8186 - 0.0063 = 0.8123.$

c) $P(\hat{p} \ge .25) = P(X \ge 37.5) = P(Z > \dfrac{37.5-30}{\sqrt{24}}) =$ P(Z>1.53) = 1-P(Z<1.53) =

1-.9730 = .0630.

EJERCICIOS

1. Los tiempos de espera en la fila de un proceso de matrícula de una universidad se distribuyen normalmente con media 45 minutos y desviación estándar de 20 minutos. Se elige al azar una muestra de 16 estudiantes que se van a matricular.
   a) ¿Cuál es la probabilidad de que el tiempo de espera promedio de la muestra sea mayor de 60 minutos?
   b) ¿Cuál es la probabilidad de que el tiempo de espera promedio de la muestra sea mayor de 35 minutos pero menor de 55 minutos?

2. Los tiempos que se demoran los empleados de una fábrica en realizar una tarea de ensamblaje se distribuyen normalmente con media de 12 minutos y desviación estándar de 6. Se toma una muestra de 10 empleados:
   a) ¿Cuál es la probabilidad de que el tiempo promedio que usan los empleados para terminar la tarea de ensamblaje sea mayor de 15, pero menor de 17 minutos?
   b) Si los 10 empleados tardan menos de hora y media en terminar la tarea de ensamblaje entonces la fábrica recibe un premio. ¿Cuál es la probabilidad de que esto ocurra?

3. El contenido promedio de cereal en un paquete es de 450 gramos con una desviación estándar de 13 gramos. Si se tomó una muestra de 35 paquetes
   a) ¿Cuál es la probabilidad de que el promedio de esta muestra sea mayor a 455 gramos?
   b) ¿Cuál es la probabilidad de que el promedio de ésta muestra se encuentre entre 445 y 458 gramos?

4. Haga uso del programa **MINITAB** para:
   a) Generar 60 muestras aleatorias de tamaño 25 de una población normal con media 60 y deviación estandar 13.
   b) Calcule la media para cada muestra generada en la parte a).
   c) Calcule la desviación estándar de los promedios calculados en la parte a)
   d) Compare los resultados obtenidos en la parte b) y c) , con lo propuesto en la parte a)

5. Un restaurant determinó que en 1 de cada 5 almuerzos vendidos el cliente pide un postre. Si en un día el restaurant realiza 600 ventas:
   a) Calcular la probabilidad de más de 150 clientes acompañe su almuerzo con un postre.
   b) Calcular la probabilidad de que a lo más 450 clientes acompañen su almuerzo con un postre.

6. En la época de invierno en los Estados Unidos se estima que el 90% de la población contrae enfermedades respiratorias, para una muestra de 350 cuál es la probabilidad de que más de 315 podrian eventualmente sufrir algún tipo de enfermedades respiratorias?.