Apuntes Unidad 3

Distribución de los promedios muestrales de distintas distribuciones

FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS UNIVERSIDAD DE CHILE

CMM Centro de Modelamiento Matemático

DISTRIBUCIÓN DE LA POBLACIÓN Y DISTRIBUCIÓN MUESTRAL

La **distribución de la población** describe cómo cambian los valores que toma una variable aleatoria dentro de una población. En la mayoría de los contextos de la vida real, se desconoce la distribución de la población. Sin embargo, es posible hacerse una idea de la forma de esta distribución analizando los valores que toma la variable para un gran número de individuos de la población.

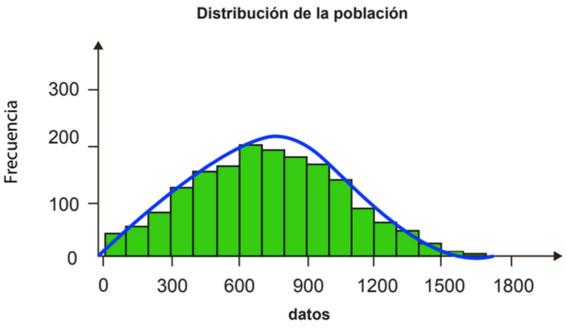

Figura 1: Gráfico ejemplo de distribución poblacional o de la población.

Por otro lado, la **distribución muestral** describe cómo varían los valores que toma la misma variable aleatoria, pero esta vez dentro de una muestra. Es importante notar que, si bien la distribución de la población es única, la distribución muestral puede cambiar de una muestra a otra. Los siguientes gráficos muestran la distribución de la misma variable en dos muestras distintas de la población:

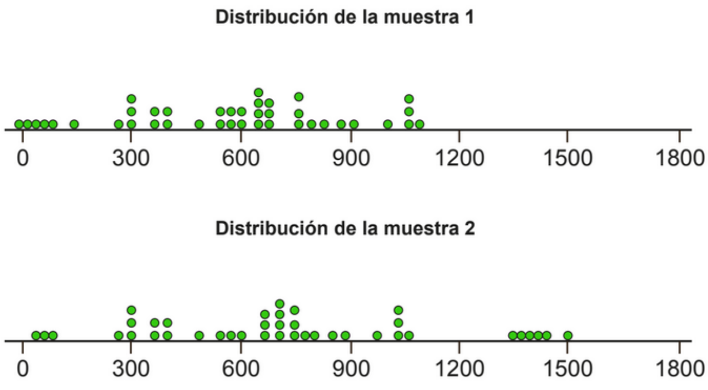

Figura 2: Distribuciones muestrales.

Curso: Probabilidad y estadística descriptiva
Unidad 3 : Situaciones o fenómenos que se modelan por medio de las distribuciones binomial y normal.
Tema: Teorema Central del Límite
Contenido: Distribución de los promedios muestrales de distintas distribuciones

Para conocer la distribución de una población, se necesita conocer sus **parámetros**, que describen características específicas de esta. Estos parámetros son, por lo general, desconocidos y deben ser estimados por medio de **estadísticos**, los cuales se determinan a partir de una muestra.

Así mismo, **cuando no se conoce la distribución de una población, comúnmente se selecciona una muestra aleatoria de esta**, y se utiliza la distribución muestral como una aproximación de la distribución de la población. La precisión de esta aproximación va a depender del proceso de selección de la muestra y el tamaño de la misma. Aun así, no existe garantía de que la distribución muestral sea lo "suficientemente" similar a la de la distribución para que la aproximación sea adecuada. Sin embargo, **la probabilidad de que esta aproximación sea adecuada aumenta a medida que el tamaño de la muestra crece**.

DISTRIBUCIÓN DE LOS PROMEDIOS MUESTRALES DE UNA DISTRIBUCIÓN NORMAL

Consideremos una variable aleatoria X que sigue una distribución normal en una cierta población, con parámetros $E(X) = \mu$ y $DE(X) = \sigma$, es decir:

$$X \sim \text{Normal}(\mu, \sigma)$$

Una muestra aleatoria de esta población se define como una colección de n variables aleatorias $X_1, \dots, X_n$ que siguen la misma distribución (normal) y que no se afectan entre sí (son independientes).

Por ejemplo, la elección de cuatro bebés de la misma población cuyos pesos fueron 3523, 2750, 3789 y 4109 gramos, corresponde a la realización de la muestra aleatoria $X_1 = 3523$, $X_2 = 2750$, $X_3 = 3789$, $X_4 = 4109$.

A partir de esta muestra aleatoria, $X_1, \dots, X_n$, podemos definir una nueva variable aleatoria, que denominaremos **promedio muestral** y la denotamos de la siguiente manera:

$$\bar{X} = \frac{X_1 + \cdots X_n}{n}$$

En particular, si cada una de las variables $X_1, \dots, X_n$ sigue una distribución normal con parámetros $E(X_i) = \mu$ y $DE(X_i) = \sigma$, entonces la **distribución del promedio muestral** es tal que:

$$\bar{X} \sim \text{Normal}\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

El promedio de una muestra de tamaño n de la población es una variable aleatoria cuya distribución estamos interesados en describir. Para esta variable, el espacio muestral corresponde a todas las muestras posibles de tamaño n. Notemos que cada muestra corresponde a una combinación distinta de n individuos. Entonces, si tenemos una población conformada de 1400 individuos, el número de muestras aleatorias distintas de tamaño $n = 100$ es igual a:

$$\binom{1400}{n} = \binom{1400}{100}$$

Por lo tanto, recordando que previamente concluímos que si la distribución de la población sigue una distribución normal, el promedio muestral sigue una distribución normal. Tras realizar muestreos aleatorios de una distribución asimétrica, podemos ver que la distribución del promedio muestral tiene forma normal en la medida de que el tamaño muestral sea grande.

TEOREMA CENTRAL DEL LÍMITE

Según lo que hemos estudiado, observamos que a medida que el tamaño de la muestra aleatoria seleccionada es mayor, la distribución del promedio muestral toma una forma acampanada y simétrica, distinta a la de la distribución poblacional. Además, la media del promedio muestral se aproxima cada vez más a la media de la población y la desviación estándar es menor en comparación a la de la distribución de la población.

Por ello, dada una muestra aleatoria de tamaño n, extraída de la misma población, cuya media es μ y desviación estándar σ, se puede demostrar que la distribución del promedio muestral es **aproximadamente** normal con parámetros:

- $E(\bar{X}) = \mu$, es decir, su media coincide con la de la distribución de la población.

- $DE(\bar{X}) = \frac{\sigma}{\sqrt{n}}$, es decir, su desviación estándar es más pequeña que la de la población.

Recordemos que si la distribución de la población es normal, entonces la distribución del promedio muestral es exactamente normal. Si la distribución de la población se desconoce o bien se sabe que no sigue una distribución normal, se requiere que el tamaño n de la muestra sea grande para observar que el promedio muestral tiene una distribución con forma acampanada y simétrica como una distribución normal. Suele considerarse "grande"

$$N(\mu, \frac{\sigma}{\sqrt{n}})$$

un tamaño $n \geq 30$, y mientras más grande sea n, mejor será la aproximación a la distribución normal:

Conocer la distribución del promedio muestral es de mucha utilidad para el cálculo de probabilidades. Por ejemplo, si estamos interesados en saber la probabilidad de que el promedio muestral, $\bar{X}$, se encuentre entre un rango de valores a y b, este se puede aproximar por el cálculo de la probabilidad de una variable aleatoria Y que sigue una distribución normal de valor esperado μ y desviación estándar $\frac{\sigma}{\sqrt{n}}$. Luego, si n es grande, entonces:

$$\mathrm{P}(a \leq \bar{X} \leq b) \approx \mathrm{P}(a \leq Y \leq b)$$

con

$$Y \sim \mathrm{N}\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Curso: Probabilidad y estadística descriptiva
Unidad 3 : Situaciones o fenómenos que se modelan por medio de las distribuciones binomial y normal.
Tema: Teorema Central del Límite
Contenido: Distribución de los promedios muestrales de distintas distribuciones

SÍNTESIS

- La **distribución de la población** describe la variabilidad de los valores que toma una variable aleatoria asociada a la población. En la mayoría de los contextos de la vida real, se desconoce la distribución de la población. Sin embargo, es posible hacerse una idea de esta distribución por medio del estudio de una muestra. La **distribución muestral** describe la variabilidad de los valores que toma la misma variable aleatoria, pero esta vez respecto a la muestra.

- Para conocer la distribución de una población, se necesita conocer sus **parámetros**, que corresponden a valores de características específicas de la población. Estos parámetros son, por lo general, desconocidos y deben ser estimados por medio de **estadísticos**, los cuales son funciones de los valores asociados a las características en una muestra.

- Consideremos una variable aleatoria X que sigue una distribución normal en una cierta población, con parámetros $E(X) = \mu$ y $DE(X) = \sigma$. A partir de esta muestra aleatoria de dicha población, $X_1, \dots, X_n$, podemos definir la variable aleatoria promedio muestral de la siguiente manera:

$$\bar{X} = \frac{X_1 + \cdots X_n}{n}$$

  la **distribución del promedio muestral** es tal que:

$$\bar{X} \sim \mathrm{Normal}\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Consideremos la distribución de una población la cual es desconocida o no sigue una distribución normal, pero que si conocemos su media μ y desviación estándar σ.

Si se extrae de esta población una muestra aleatoria de tamaño $n \geq 30$, se puede demostrar que la distribución del promedio muestral $\bar{X}$ es aproximadamente normal con parámetros:

- $E(\bar{X}) = \mu$, es decir, su media coincide con la de la distribución de la población.

- $DE(\bar{X}) = \frac{\sigma}{\sqrt{n}}$, es decir, su desviación estándar es más pequeña que la de la población.

Suele considerarse "grande" un tamaño $n \geq 30$, y mientras más grande sea n, mejor será la aproximación a la distribución normal.

La probabilidad de que el promedio muestral, $\bar{X}$, se encuentre entre un rango de valores a y b, se puede aproximar por el cálculo de la probabilidad de una variable aleatoria Y que sigue una distribución normal de valor esperado μ y desviación estándar $\frac{\sigma}{\sqrt{n}}$ cuando n es grande, entonces:

$$\mathrm{P}(a \leq \bar{X} \leq b) \approx \mathrm{P}(a \leq Y \leq b)$$

con

$$Y \sim \mathrm{N}\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$