Estimadores muestrales y teorema del límite central

Universidad de Guayaquil

noviembre 20, 2018

Distribuciones y estimadores muestrales

- La distribución muestral de un estadístico (como una proporción muestral o una media muestral) es la distribución de todos los valores del estadístico cuando se obtienen todas las muestras posibles del mismo tamaño n de la misma población. (La distribución muestral de un estadístico generalmente se representa como la distribución de probabilidad en el formato de tabla, histograma de probabilidad o fórmula).
- Algunos estadísticos (como la proporción y la media) son útiles para estimar valores de parámetros poblacionales, mientras que otros estadísticos (como la mediana) no son buenos estimadores de los parámetros poblacionales.

Distribución muestral de proporciones (I)

La **distribución muestral de la proporción** es la distribución de probabilidad de proporciones muestrales, donde todas las muestras tienen el mismo tamaño muestral n y provienen de la misma población.

Propiedades de la distribución de proporciones muestrales

- Las proporciones muestrales tienden a coincidir con el valor de la proporción poblacional. (Es decir, todas las proporciones muestrales posibles tienen una media igual a la proporción poblacional).
- En ciertas condiciones, la distribución de la proporción muestral puede aproximarse por medio de una distribución normal.

Distribución muestral de la media

La distribución muestral de la media es la distribución de medias muestrales, donde todas las medias tienen el mismo tamaño muestral n y se obtienen de la misma población. (La distribución muestral de la media generalmente se representa como una distribución de probabilidad en formato de tabla, histograma de probabilidad o fórmula).

¿Cuáles estadísticos son buenos estimadores de parámetros?

- El valor de un estadístico, como la media muestral depende de los valores particulares incluidos en la muestra y generalmente varía de una muestra a otra. Esta variabilidad se conoce como **Variabilidad de muestreo**
- Estadísticos que coinciden con los parámetros poblacionales: media, varianza, proporción
- Estadísticos que no coinciden con los parámetros poblacionales: mediana, rango, desviación estándar

El teorema del límite central (I)

El teorema del límite central nos dice que si el tamaño de una muestra es lo suficientemente grande, la distribución de las medias muestrales se puede aproximar por medio de una distribución normal, aún cuando la población original no esté distribuida de forma normal. Cuando se selecciona una muestra aleatoria simple de una población con media μ y desviación estándar σ, es esencial conocer los siguientes principios:

- Si $n > 30$, entonces las medias muestrales tienen una distribución que se puede aproximar por medio de una distribución normal, con una media μ y una desviación estándar.
- Si $n \leq 30$ y la población original tiene una distribución normal, entonces las medias muestrales tienen una distribución normal con una media μ y una desviación estándar $\sigma/\sqrt{n}$

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

El teorema del límite central (II)

Para la mayoría de las poblaciones, si el tamaño muestral es mayor a 30, la aproximación del teorema del límite central es buena.

Ejemplo

Sea X el número de imperfecciones en una pulgada de un alambre de cobre. La función de masa de probabilidad de X se muestra en la siguiente tabla.

x	P(X=x)
0	0.48
1	0.39
2	0.12
3	0.01

Se toma una muestra de 100 alambres de esta población. ¿Cuál es la probabilidad de que el número promedio de imperfecciones por alambre en esta muestra sea menor a 0.5?

El teorema del límite central (III)

```
x=c(0,1,2,3)
p=c(0.48,0.39,0.12,0.01)
sum(p*x) # valor esperado
```

```
## [1] 0.66
```

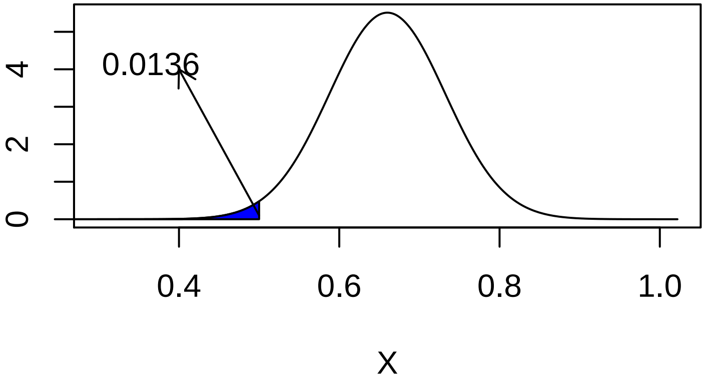
```
round(sqrt(sum((x^2)*p)-(sum(p*x))^2),digits=3) #desviacion estandar
```

```
## [1] 0.724
```

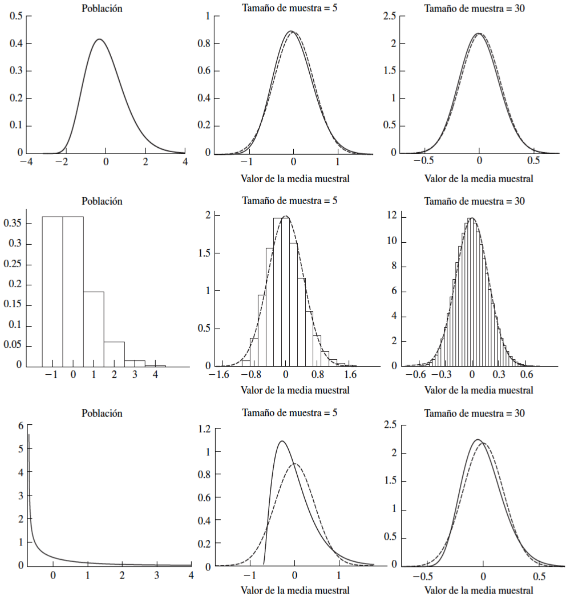
$$\bar{X} \sim N\left(0{,}66, \frac{0{,}724}{\sqrt{100}}\right)$$

$$z = \frac{0{,}5 - 0{,}66}{\frac{0{,}724}{\sqrt{100}}} = -2{,}21; \quad P(\bar{X} < 0{,}5) = 0{,}0136$$

El teorema del límite central (IV)

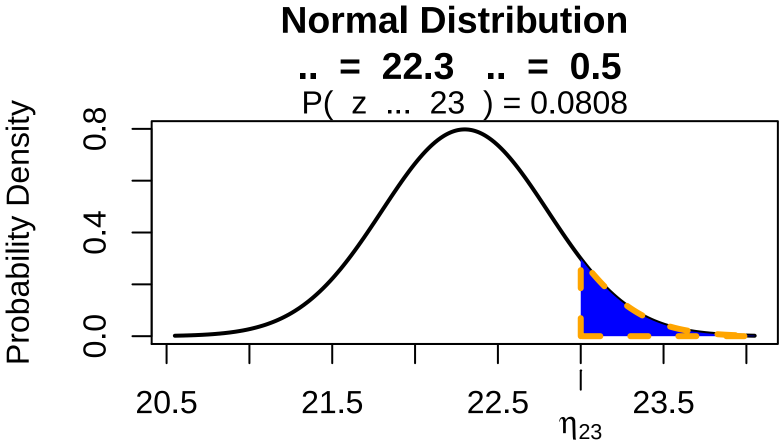


El teorema del límite central (V)



El teorema del límite central - Ejercicios Propuestos

- En una universidad grande, la media de la edad de los estudiantes es 22.3 años y la desviación estándar es de cuatro años. Se toma una muestra aleatoria de 64 estudiantes. ¿Cuál es la probabilidad de que la edad promedio de estos estudiantes sea mayor a 23 años?



Aplicación del teorema del límite central

- El tamaño de la muestra es mayor que 30, o la población original se distribuye normalmente
- Cuando trabaje con un valor individual de una población distribuida normalmente, utilice:

$$z = \frac{x - \mu}{\sigma}$$

- Cuando trabaje con una media de alguna muestra (o grupo), asegúrese de utilizar el valor de para la desviación estándar de las medias muestrales. Utilice:

$$z = \frac{x - \mu}{\frac{\sigma}{\sqrt{n}}}$$

Corrección para una población finita

Al aplicar el teorema del límite central, el uso de $\sigma_{\bar{x}} = \sigma/\sqrt{n}$supone que la población tiene un número infinito de miembros. Cuando hacemos un muestreo con reemplazo (es decir, cada elemento seleccionado se reincorpora a la muestra antes de hacer la siguiente selección), la población es efectivamente infinita. Aunque muchas aplicaciones realistas implican un muestreo sin reemplazo, estas muestras sucesivas dependen de resultados previos.

Cuando realice un muestreo sin reemplazo y el tamaño de muestra n sea mayor que el 5 % de la población finita de tamaño N (es decir, n $>$ 0.05N), ajuste la desviación estándar de medias muestrales multiplicándola por el factor de corrección de población finita:

$$\sqrt{\frac{N-n}{N-1}}$$

Ejercicio propuesto

- Presión sanguínea. La presión sanguínea sistólica (en mm de Hg) de mujeres entre 18 y 24 años se distribuye normalmente, con una media de 114.8 y una desviación estándar de 13.1 (según datos de la National Health Survey). La hipertensión suele definirse como una presión sistólica mayor que 140.

1. Si se selecciona al azar a una mujer entre 18 y 24 años, calcule la probabilidad de que su presión sistólica sea mayor que 140.
2. Si se seleccionan al azar 4 mujeres del mismo rango de edad, calcule la probabilidad de que su presión sistólica media sea mayor que 140.
3. Dado que el inciso b) incluye un tamaño de muestra no mayor que 30, ¿por qué se puede utilizar el teorema del límite central?
4. Si un médico recibe un reporte que afirma que 4 mujeres tienen una presión sistólica media menor que 140, ¿puede concluir que ninguna de las mujeres es hipertensa (con una presión sanguínea mayor que 140)?

La distribución normal como aproximación de la distribución binomial

Si una distribución de probabilidad binomial satisface los requisitos $np = 5$ y $nq = 5$, entonces la distribución de probabilidad binomial puede aproximarse con una distribución normal con media $\mu = np$ y desviación estándar $\sigma = \sqrt{npq}$ así como con el número entero discreto x ajustado con una corrección por continuidad, de manera que x está representada por el intervalo de $x - 0{,}5$ a $x + 0{,}5$.

Ejemplo (I)

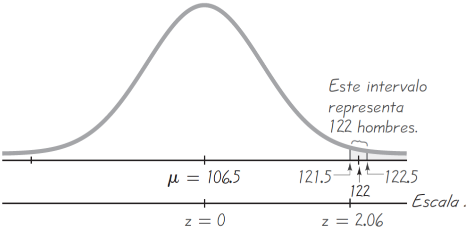
- Carga de pasajeros en un Boeing 767-300 Un avión Boeing 767-300 de American Airlines tiene 213 asientos. Cuando se llena con pasajeros, equipaje, carga y combustible, el piloto debe verificar que el peso neto no rebase el límite máximo permitido, y el peso debe distribuirse adecuadamente para que el equilibrio de la aeronave permanezca dentro de los límites de seguridad aceptables. En vez de pesar a cada pasajero, sus pesos se estiman según las reglas de la Federal Aviation Administration. En realidad, sabemos que los hombres tienen un peso medio de 172 lb y que las mujeres tienen un peso medio de 143 lb, de manera que un número desproporcionadamente mayor de hombres podría provocar una situación insegura de sobrepeso. Suponga que, si hay por lo menos 122 hombres en una lista de 213 pasajeros, la carga debe ajustarse de alguna manera. Suponiendo que los pasajeros se registran al azar, que los hombres y las mujeres son igualmente probables, y que la aeronave está llena de adultos, calcule la probabilidad de que en un Boeing 767-300 con 213 pasajeros haya al menos 122 hombres.

$$np = 213 \times 0{,}5 = 106{,}5 \quad (\text{Por lo tanto } np \geq 5)$$
$$nq = 213 \times 0{,}5 = 106{,}5 \quad (\text{Por lo tanto } nq \geq 5)$$

Ejemplo (I)

$$\mu = np = 213 \times 0{,}5 = 106{,}5$$

$$\sigma = \sqrt{npq} = \sqrt{213 \times 0{,}5 \times 0{,}5}$$



$$z = \frac{x - \mu}{\sigma}$$

$$P[z \geq 2{,}06] = 1 - P(z < 2{,}06) = 0{,}0197$$

TAREA

Leer las páginas del libro de Mario Triola que se adjuntaron en el edmodo.