INTRODUCCIÓN

En esta unidad iniciamos el análisis confirmatorio de datos, generalmente ligado a la Estadística lnferencial. La inferencia estadística comprende una serie de técnicas de uso imprescindible para tomar decisiones con respecto a la cuestión planteada por el investigador al comienzo de su tarea de análisis de datos.

Es necesario destacar que las decisiones que debe tomar el investigador, ante la situación de incertidumbre que implica inferir sobre casos particulares a la generalidad, deben estar respaldadas por la objetividad que garantiza la aplicación del método científico.

Aquí trataremos las dos grandes ramas en que se divide la inferencia estadística:

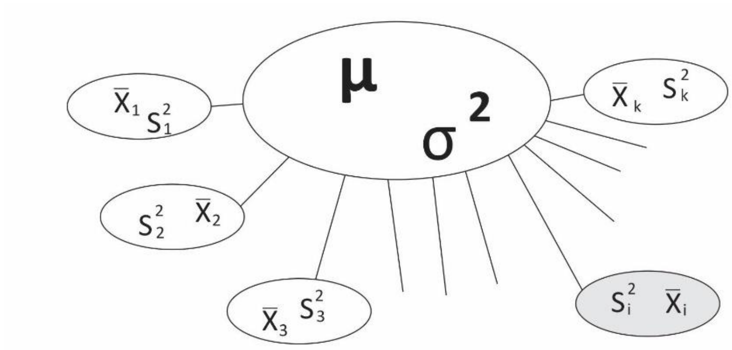
Antes de comenzar el estudio de estos dos temas, veremos algunos conceptos referidos a parámetros poblacionales, estadísticos muestrales, distribución de estos estadísticos y condiciones que deben reunir para ser buenos estimadores.

En esta unidad calcularemos estimaciones puntuales de los parámetros poblaciones.

1. ESTIMADORES PUNTUALES

El objetivo de la estadística consiste en hacer inferencias acerca de los parámetros de una población teniendo en cuenta la información contenida en la muestra.

Ahora bien, como en general los parámetros poblacionales son desconocidos, existe una amplia gama de técnicas estadísticas que tienen como objetivo la estimación de estos parámetros a través de estadísticos muestrales adecuados a cada caso en particular.



La base teórica que sustenta la metodología que aplicamos a los resultados de una muestra, se fundamenta en la distribución de probabilidad del estimador calculado en cada una de las muestras posibles.

Un **estimador** es una regla que nos dice cómo calcular la estimación basándose en la información contenida en la muestra. Generalmente se expresa mediante una fórmula.

Por ejemplo, la media muestral se obtiene sumando todas las observaciones de la muestra y dividiendo esta suma por el tamaño de la muestra.

Es decir,

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

y este estadístico es un estimador de la media de la población μ.

Cualquier persona podría definir otra combinación de las observaciones muestrales como estimador del parámetro μ y entonces cabría la pregunta:

¿Cuál es el "mejor" estimador de un parámetro?

Un problema importante que debió resolver la teoría estadística, fue el de determinar el mejor estimador de cada parámetro en particular.

Es fácil pensar que, para cada parámetro a estimar, existen "buenos" y "malos" estimadores. Por este motivo, pasaremos a considerar ciertas propiedades que deben reunir los estimadores para ser considerados "buenos" estimadores.

1.1 ESTIMADOR PARA LA MEDIA

Por ejemplo, supongamos que se desea conocer la altura promedio de niños de una determinada edad, en la ciudad de Aguascalientes, México. Lógicamente sería casi imposible por problemas de costo, de tiempo, etc., determinar la altura de todos los niños que reúnan las características consideradas.

Difícilmente se podrá conocer el verdadero valor del parámetro μ, altura promedio de todos los niños de cierta edad en la ciudad de Aguascalientes. La única solución a este problema será recurrir a una muestra de niños seleccionados aleatoriamente de esta población, registrar su altura y calcular el promedio.

Se obtendrá así el valor de un estadístico muestral mediante el cual se estimará el verdadero promedio de altura de niños de la población objetivo.

Supongamos que, en una muestra de 50 niños, el promedio de altura encontrado es de 1.26 mts. En este caso se está haciendo una estimación puntual del parámetro altura promedio en la población.

$$\mu \rightarrow \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

1.2 ESTIMADOR PARA UNA PROPORCIÓN

Por medio de un ejemplo sencillo, veremos cómo se distribuye este estimador muestral.

Supongamos que se considera una población de $N = 4$ familias a las que se les pregunta si tienen o no televisión por cable. Las respuestas dadas por estas familias se observan en la siguiente tabla:

Familia	Posee televisión por cable
A	NO
B	SI
C	NO
D	NO

La variable a considerar es la cantidad de familias con cable, lo que implica que se trata de una distribución binomial. La proporción de familias con cable en la población es $P = 0.35$. Sin embargo, en esta muestra tenemos que solamente una de cada cuatro familias tiene cable:

$$p = \frac{x}{n} = \frac{1}{4} = 0.25$$

El mejor estimador puntual de P será, intuitivamente, la proporción muestral p.

$$p \rightarrow P$$

Esta proporción se obtiene:

$$p = \frac{x}{n}$$

donde la cantidad x representa el número total de éxitos en n pruebas binomiales y n el total de pruebas.

Se conoce que la proporción muestral p es un estimador insesgado de P y además posee varianza mínima cuando se lo compara con otros estimadores.

1.3 DISTRIBUCIÓN DE PROBABILIDAD

Supongamos que extraemos todas las muestras posibles de tamaño *n* = 2 con reposición y los resultados se dan en la siguiente tabla:

Familias seleccionadas en la muestra	Variable cantidad de familias con cable		
	Familia 1	Familia 2	Proporción
AA	NO	NO	0
AB	NO	SI	0.5
AC	NO	NO	0
AD	NO	NO	0
BA	SI	NO	0.5
BB	SI	SI	1
BC	SI	NO	0.5
BD	SI	NO	0.5
CA	NO	NO	0
CB	NO	SI	0.5
CC	NO	NO	0
CD	NO	NO	0
DA	NO	NO	0
DB	NO	SI	0.5
DC	NO	NO	0
DD	NO	NO	0

Vemos que, la variable aleatoria "proporción muestral", toma solamente 3 valores: 0, 0.5 y 1. La siguiente tabla muestra la distribución de probabilidad de esta variable:

Proporción muestral (p)	Probabilidad (f(p))
0	9/16
0.5	6/16
1	1/16

Calculamos ahora la esperanza y la varianza de la proporción muestral.

$$E(p) = \sum_{i=1}^{n} pf(p) = 0 \times \frac{9}{16} + 0.5 \times \frac{6}{16} + 1 \times \frac{1}{16} = \frac{4}{16} = 0.25 = P$$

$$V(p) = \sum_{i=1}^{n} (p - 0.25)^2 f(p) = \frac{1.5}{16} = \frac{0.1875}{2} = \frac{PQ}{n}$$

La distribución de *p* es bastante asimétrica, pero se va asemejando a la distribución normal cuando *nP* y *nQ* son ambos mayores a 5. En este ejemplo, nP = 2 x 0.25 = 0.5 y nQ = 2 x 0.75 = 1.5 por lo cual se puede pensar que la distribución de probabilidad es bastante asimétrica.

¿Quieres graficar la distribución de probabilidad y verificarlo?

El teorema central del límite también se cumple para la variable proporción muestral ya que una variable binomial se puede considerar como suma de variables bipuntuales. Entonces, se puede escribir lo siguiente:

$$z = \frac{p - P}{\sqrt{\frac{PQ}{n}}}$$

Realizar esta operación se conoce como **estandarización de la variable.**

Ejemplo

Se conoce por experiencias pasadas que 5 de cada 10 ventas de un comercio de ropa deportiva fueron realizadas con tarjetas de crédito. En el día de hoy se han hecho 100 ventas.

La proporción de ventas con tarjetas de crédito en la población es P = 0.50.

Como n = 100, podemos suponer que la variable aleatoria proporción muestral se distribuye normalmente. Además, nP = 100 x 0.5 = 5 y nQ = 100 x 0.5 = 5.

Se pide:

a) ¿Cuál es la probabilidad de que las ventas con tarjetas de crédito estén entre el 60% y el 70%?

b) ¿Cuál es la probabilidad de que las ventas con tarjetas de crédito superen al 55%?

c) ¿Cuál es la probabilidad de que las ventas con tarjetas de crédito sean inferiores al 40%?

a)

$$P(0.60 < p < 0.70) = P\left(\frac{0.60 - 0.50}{\sqrt{\frac{0.5\times 0.5}{100}}} < z < \frac{0.70 - 0.50}{\sqrt{\frac{0.5\times 0.5}{100}}}\right)$$

$$= P(2 < z < 4) = F(4) - F(2) = 1 - 0.9772 = 0.0228$$

b)

$$P(p > 0.55) = P\left(z > \frac{0.55 - 0.50}{\sqrt{\frac{0.5\times0.5}{100}}}\right)$$

$$= P\left(z > \frac{0.05}{0.05}\right) = P(z > 1) = 1 - F(1) = 1 - 0.8413 = 0.1587$$

c)

$$P(p < 0.40) = P\left(z > \frac{0.40 - 0.50}{\sqrt{\frac{0.5\times0.5}{100}}}\right)$$

$$= P\left(z < \frac{-0.10}{0.05}\right) = P(z < -2) = 1 - F(2) = 1 - 0.9772 = 0.0228$$

Actividades optativas

Una empresa de auditoría médica, sabe que el 15% de los pacientes ambulatorios de los hospitales se queja del no cumplimiento de los turnos dados.
Para controlar este servicio, decide tomar una muestra aleatoria de 120 pacientes y desea conocer la probabilidad de que:

a) Se queje más del 20%.
b) Se queje menos del 10%.

Actividades optativas

De acuerdo a registros del Departamento de Calidad de una empresa, el 10% de las piezas fabricadas en un turno van a scrap (desperdicio). A fin de solucionar este problema, se decide tomar una muestra de 200 piezas fabricadas durante 5 días consecutivos. Cuál es la probabilidad de que en la muestra se encuentre:

a) ¿entre un 8% y un 10% de scrap?
b) ¿más de un 13% de scrap?
c) ¿menos de 11% de scrap?

1.4 ESTIMADORES PARA LA VARIANZA

Existen situaciones prácticas, generalmente ligadas a la investigación industrial, en las que interesa estimar al parámetro varianza poblacional.

Así, los instrumentos de medición, tan utilizados por los ingenieros que monitorean sistemas productivos, deben medir con la mayor exactitud posible. El sesgo de un instrumento de medición puede ser corregido pero su precisión se mide por la desviación estándar de las mediciones que efectúa. La precisión generalmente está ligada al diseño del instrumento y no puede ser aumentada por el operador de la máquina. Por ello, resulta de gran importancia conocer la variabilidad de las mediciones antes de decidir, por ejemplo, la compra de un instrumento de medición.

Los profesionales ligados al área de control de calidad de una empresa reconocen la importancia que tiene, en la calidad de un producto, el estudio de la capacidad de un proceso productivo para garantizar la producción de piezas homogéneas. Cuanto más homogéneas son las mediciones de las piezas, mayor es su calidad. No es lo mismo vender, por ejemplo, un producto alimenticio empaquetado en bolsitas de 1 Kg. Que han sido llenadas por máquinas que operan con una precisión de ± 0.050 Kg. que con máquinas que empaquetan con una precisión de ± 0.350 Kg.

Siempre que tenemos que estimar variabilidad, debemos hacer una estimación del parámetro poblacional σ^2.

El estimador lógico del parámetro σ^2 será, evidentemente, la varianza muestral S^2.

La varianza se calcula de la siguiente manera:

$$S^2 = \frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n}$$

Ahora bien, si tomáramos hipotéticamente todas las muestras posibles de un cierto tamaño n, calculáremos en cada una, la varianza muestral y obtuviéramos la esperanza matemática y la varianza muestral de la variable aleatoria, veríamos que ésta no es igual a σ^2, el correspondiente parámetro poblacional.

En símbolos:

$$E(S^2) \neq \sigma^2$$

Podemos obtener la varianza muestral correspondiente a cada una de las muestras de tamaño $n = 2$ utilizando nuevamente la información proporcionada por la Tabla 1.

Como vemos adicionamos dos cálculos de S^2, uno dividiendo a la suma de los desvíos por n y otro dividiendo la misma suma por $n - 1$. Para diferenciarlos llamaremos al primer S_1^2 cálculo y al segundo S_2^2.

Tabla 1: Cálculo de la varianza muestral S_1^2 y de la varianza muestral corregida S_2^2

Muestras	$S_1^2=\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}$	$S_2^2=\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$
(2,2)	0,00	0,00
(2,3)	0,25	0,25
(2,6)	4,00	8,00
(2,8)	9,00	18,00
(2,11)	20,25	40,50
(3,2)	0,25	0,50
(3,3)	0,00	0,00
(3,6)	2,25	4,50
(3,8)	6,25	12,50
(3,11)	16,00	32,00
(6,2)	4,00	8,00
(6,3)	2,25	4,50
(6,6)	0,00	0,00
(6,8)	1,00	2,00
(6,11)	6,25	12,50
(8,2)	9,00	18,00
(8,3)	6,25	12,50
(8,6)	1,00	2,00
(8,8)	0,00	0,00
(8,11)	2,25	4,50
(11,2)	20,25	40,50
(11,3)	16,00	32,00
(11,6)	6,25	12,50
(11,8)	2,25	4,50
(11,11)	0,00	0,00

Podemos aún resumir más esta información efectuando las correspondientes tablas de distribución de probabilidad para cada uno de los cálculos de S^2, tal como lo presentamos en las Tablas 2 y 3.

Tabla 2: Distribución de probabilidad de la variable aleatoria varianza muestral *S_1^2*

S_1^2	n_i	$p(S_1^2)$	$S_1^2\ p(S_1^2)$
0,00	5	5/25	0,00/25
0,25	2	2/25	0,50/25
1,00	2	2/25	2,00/25
2,25	4	4/25	9,00/25
4,00	2	2/25	8,00/25
6,25	4	4/25	25,00/25
9,00	2	2/25	18,00/25
16,00	2	2/25	32,00/25
20,25	2	2/25	40,50/25
	25	25/25	135,00/25

Tabla 3: Distribución de probabilidad de la variable aleatoria varianza muestral corregida *S_2^2*

S_2^2	n_i	$p(S_2^2)$	$S_2^2\ p(S_2^2)$
0,00	5	5/25	0,00/25
0,50	2	2/25	1,00/25
2,00	2	2/25	4,00/25
4,50	4	4/25	18,00/25
8,00	2	2/25	16,00/25
12,50	4	4/25	50,00/25
18,00	2	2/25	36,00/25
32,00	2	2/25	64,00/25
40,50	2	2/25	81,00/25
	25	25/25	270,00/25

Ahora estamos en condiciones de calcular el valor esperado de estos estimadores; vemos que:

$$E(S_1^2) = \sum_{i=1}^{n} S_1^2 \cdot p(S_1^2) = \frac{135}{25} = 5.4 \neq \sigma^2$$

mientras que:

$$E(S_2^2) = \sum_{i=1}^{n} S_2^2 \cdot p(S_2^2) = \frac{270}{25} = 10.8 \neq \sigma^2$$

Podemos concluir, ante esta evidencia, que S^2 no es un estimador insesgado del parámetro que debe estimar.

Para obtener un estimador insesgado de σ^2 tuvimos que dividir la suma de los desvíos por $n - 1$ en lugar de n.

Esta distinción es bastante importante cuando se está trabajando con muestras chicas, puesto que para muestras grandes la diferencia entre n y $n - 1$ es despreciable. Esto hace que S_1^2 a pesar de ser un estimador sesgado de σ^2 sea consistente, sin embargo, se hace insesgado a medida que aumenta el tamaño de la muestra.

De acuerdo a lo establecido, de aquí en adelante siempre que trabajemos con la varianza muestral, la referiremos a la suma de desvíos dividida por $n - 1$, es decir a S_2^2. Para simplificar la llamaremos S^2.

1.5 ESTIMADORES PARA TOTALES

Si el interés es estimar totales, se puede calcular el mismo con los datos de la muestra de tal forma que:

$$\hat{T} = \sum_{i=1}^{n} x_i$$

Si x es la variable que mide ingresos familiares en una muestra de tamaño n, el total de ingresos de las n familias es $\hat{T}$.