Probabilidad y Estadística – MEDIDAS MUESTRALES Y ESTIMADORES PUNTUALES – UTN

***MEDIDAS MUESTRALES**

En esta instancia trabajaremos situaciones en las que se nos presente una variable aleatoria de la cual no tengamos ninguna información disponible sobre su distribución, pero a cambio de eso, podemos obtener lo que se llama una muestra aleatoria (m.a.). Veamos la definición formal de m.a. y luego interpretémoslo a través de un ejemplo).

Definición: Dada una v.a. X de define **muestra aleatoria de tamaño n** al conjunto de v.a. independientes $X_1, X_2, \dots, X_n$ obtenidas a partir de X es decir, con la misma distribución que X.

Ejemplo: En un experimento se hicieron 16 mediciones independientes de la temperatura de sublimación del iridio obteniendo los siguientes valores, en grados centígrados

136,6	145,2	151,5	162,7
159,1	159,8	160,8	173,9
160,1	160,4	161,1	160,6
160,2	159,5	160,3	159,2

a) Definir las variables aleatorias.
En este experimento interfieren 17 v.a.:

$$\begin{aligned} X &= \text{temperatura de sublimación del iridio} \\ X_i &= \text{medición i-ésima de la temperatura de sublimación del iridio,} \end{aligned}$$

con $i = 1, 2, \dots, 16$. De esta manera $X_1, X_2, \dots, X_{16}$ son una muestra aleatoria de X.

Nos podemos preguntar, qué podemos hacer con estos valores? Como sabemos estos valores son observaciones de la variable aleatoria X, de la cual no conocemos su distribución, pero seguro sabemos que tiene una esperanza y una varianza. Por lo tanto podemos calcular la ESPERANZA y la VARIANZA MUESTRALES, como así también la MEDIANA MUESTRAL:

Definición: Sea $X_1, X_2, \dots, X_n$ una muestra aleatoria con distribución X, podemos definir

1. Media muestral o promedio muestral a la v.a. $\overline{X} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$
2. Varianza muestral a la v.a. $S^2 = \dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n}(X_i - \overline{X})^2$
3. Desvío estándar muestral a la v.a. $S = \sqrt{\dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n}(X_i - \overline{X})^2}$
4. Mediana muestral: Si previamente definimos $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ a la muestra ordenada de menor a mayor, entonces se obtiene como:

a) Si n es una cantidad par: $\tilde{X} = \dfrac{X_{\left(\frac{n}{2}\right)} + X_{\left(\frac{n+2}{2}\right)}}{2}$

b) Si n es una cantidad impar: $\tilde{X} = X_{\left(\frac{n+1}{2}\right)}$

5. Rango muestral a la v.a. $Rg = X_{(n)} - X_{(1)}$.

Volvamos al **Ejemplo:**
b) Determinar la media, la varianza, el desvío estándar y la mediana muestrales.

- Media muestral:
  $\overline{x} = \frac{136{,}6+145{,}2+151{,}5+162{,}7+159{,}1+159{,}8+160{,}8+173{,}9+160{,}1+160{,}4+161{,}1+160{,}6+160{,}2+159{,}5+160{,}3+159{,}2}{16} = 158{,}1875$

- Varianza muestral:
  $s^2 = \frac{1}{15}\Big((136{,}6 - 158{,}1875)^2 + (145{,}2 - 158{,}1875)^2 + \ldots + (159{,}2 - 158{,}1875)^2\Big) = 66.06917$

- Desvío estándar muestral:
  $s = \sqrt{66{,}06917} = 8{,}128294$

- Mediana muestral: Para poder obtenerla, primero hay que ordenar los datos de menor a mayor

  136,6 145,2 151,5 159,1 159,2 159,5 159,8 160,1 160,2 160,3 160,4 160,8 160,6 161,1 162,7 173,9

  Como n es par, resulta encontes que la mediana se obtiene de la siguiente manera:
  $\tilde{x} = \frac{160{,}1+160{,}2}{2} = 160{,}15$

c) Obtener el mínimo, el máximo y el rango muestral.

- Mínimo: $x_{(1)} = 136{,}6$

- Máximo: $x_{(16)} = 173{,}9$

- Rango muestral: $rg = x_{(16)} - x_{(1)} = 173{,}9 - 136{,}6 = 37{,}3$

---

*ESTIMADORES PUNTUALES

Supongamos que ahora se da el caso en que sí conocemos la distribución de la v.a. X pero no tenemos los valores de sus parámetros, entonces, para poder calcular probabilidades necesitaremos estimarlos, es decir, aproximar su valor, y para ello es que utilizaremos las medidas muestrales definidas en la sección anterior.

Definición: Un estimador es una variable aleatoria que se utiliza para aproximar el valor de un parámetro. (Pues varía según la muestra extraída)

Cómo hacemos para obtener estos estimadores?

Estimador de momentos: Sean X una v.a. con una distribución de parámetro θ (donde θ representa cualquier parámetro de todos los que vimos según la distribución especial, ya sea discreta o continua) y $X_1, X_2, \ldots, X_n$ una m.a. de X, entonces el estimador $\hat{\theta}$ se obtiene de resolver la ecuación:

$$\frac{1}{n}\sum_{i=1}^{n} X_i = E(X)$$

Estimadores usuales: Sea X una v.a. cualquiera con $E(X) = \mu$ y $V(X) = \sigma^2$ y $X_1, X_2, \ldots, X_n$ una m.a. entonces los estimadores se calculan como:

$$\widehat{\mu} = \overline{X} \qquad \widehat{\sigma^2} = S^2$$

Consistencia de los estimadores usuales:

***Ley de los Grandes Números:** Sea $X_1, X_2, X_3, \ldots$ una sucesión infinita de v.a.i.i.d tales que $E(X_i) = \mu$, entonces

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{p}_{n\to\infty} \mu$$

Se dice que $\overline{X}$ converge en probabilidad (p) a μ.

***Corolario:** Sea $X_1, X_2, X_3, \ldots$ una sucesión infinita de v.a.i.i.d tales que $E(X_i) = \mu$ y $V(X_i) = \sigma^2$, entonces

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 \xrightarrow{p}_{n\to\infty} \sigma^2$$

es decir que S^2 converge en probabilidad a σ^2.

Asimismo se pueden plantear diferentes estimadores para un parámetro a partir de la m.a., entonces definiremos distintas maneras de ver si un estimador es bueno o no.

Definiciones:

1. Sea $\widehat{\theta}$ un estimador del parámetro θ, se define sesgo del estimador a la diferencia:
$$B(\widehat{\theta}) = E(\widehat{\theta}) - \theta$$
2. Diremos que un estimador es insesgado si $B(\widehat{\theta}) = 0$, es decir
$$E(\widehat{\theta}) = \theta$$
, por lo tanto entre varios estimadores elegiremos aquellos que cumplan con esta propiedad
3. Dada la varianza $V(\widehat{\theta})$, elegiremos el estimador que tenga mínima varianza, es decir, el de menor varianza posible.
4. Definimos error cuadrático medio del estimador a
$$\begin{aligned} ECM(\widehat{\theta}) &= E\Big[(\widehat{\theta} - \theta)^2\Big] \\ &= V(\widehat{\theta}) + \Big[E(\widehat{\theta}) - \theta\Big]^2 \\ &= V(\widehat{\theta}) + B(\widehat{\theta})^2 \end{aligned}$$

por lo tanto, entre distintos estimadores elegiremos el de menor error cuadrático medio.

Observación: Si tenemos que elegir entre diferentes estimadores que son todos insesgados, optaremos por el de menor ECM y equivalentemente obtenendremos el de menor varianza.

Propiedad: Los estimadores $\overline{X}$ y S^2 son insesgados y de mínima varianza para μ y σ^2, respectivamente.

Distribución de los estimadores:

*En la primera parte ya estuvimos trabajando con diferentes alternativas para la distribución de $\overline{X}$, pero agregaremos una alternativa más:

Consecuencia TCL y LGN: Sean $X_1, X_2, \ldots, X_n$, v.a.i.i.d., tales que $E(X_i) = \mu$ y la $V(X_i) = \sigma^2$ para todo $i = 1, 2, \ldots, n$, entonces

$$Z_n = \frac{\overline{X} - \mu}{S/\sqrt{n}} = \sqrt{n}\frac{\overline{X} - \mu}{S} \xrightarrow{D}_{n\to\infty} N(0;1)$$

Entonces, si n es suficientemente grande ($n > 30$), diremos que

$$Z_n = \sqrt{n}\frac{\overline{X} - \mu}{S} \overset{(a)}{\sim} N(0;1) \quad Z_n \text{ tiene distribución aproximadamente } N(0;1)$$

* Ahora veamos la distribución de la varianza muestral:

Teorema: Sean $X_1, X_2, \ldots, X_n$, v.a.i.i.d. $\sim N(\mu, \sigma^2)$ (es decir que la distribución para la varianza muestral solo es válida para el caso en que la m.a. tiene distribución normal), entonces

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

Se llama distribución Ji o Chi cuadrada con parámetro $n-1$. A este valor se le suele llamar "grados de libertad". Y los valores de su función de distribución acumulada se encuentran tabulados.

Volvamos al **Ejemplo:**

d) Si se sabe que la temperatura de sublimación del iridio es una variable aleatoria con distribución $N(\mu, \sigma^2)$:

i) Estimar la esperanza con los siguientes estimadores:

$$\widehat{\mu}_1 = \overline{X} \qquad \widehat{\mu}_2 = \frac{X_3 + 2X_{10} + 3X_{12}}{6} \qquad \widehat{\mu}_3 = \frac{X_3 + 2X_{10} + 3X_{12}}{7}$$

ii) Calcular el sesgo, varianza y ECM de cada estimador. Cuál de ellos eligiría?

iii) Calcular aproximadamente la probabilidad de que la temperatura de sublimación sea superior a $150°$.

iv) Hallar la temperatura media aproximada que no es superada un 30 % de las veces.

v) Supongamos ahora que $\sigma^2 = 75$. En este contexto, obtener la probabilidad de que la varianza de la muestra sea de a lo sumo 42,735.

vi) Si llamamos p a la proporción de temperaturas que son inferiores a $160°$, estimar p. Por otra parte, suponiendo que las estimaciones obtenidas se hicieron en base a una muestra aleatoria de tamaño 36, hallar la probabilidad de que la proporción supere el 45 %.

e) Si ahora la temperatura es una v.a. de parámetro γ cuya función de densidad está dada por:

$$f(x) = \begin{cases} 4\gamma(x-140)+\gamma & \text{si } 140 \leq x \leq 180 \\ 0 & \text{caso contrario} \end{cases}$$

Hallar el estimador de momentos de γ.