TEMA 12 – INFERENCIA ESTADÍSTICA. ESTIMACIÓN DE LA MEDIA

12.1 – DISTRIBUCIÓN NORMAL. REPASO DE TÉCNICAS BÁSICAS

UTILIZACIÓN DE LA TABLA DE LA NORMAL N(0,1)

En la distribución N(0,1), a la variable se le suele representar por la letra z. La tabla nos da las probabilidades $P[z \leq k]$ para valores de k de 0 a 4, de centésima en centésima. A estas probabilidades se las llama $\phi(k)$: $\phi(k) = P[z \leq k]$ z se distribuye N(0,1)
$\phi(k)$ es la función de distribución de esta variable aleatoria.

El valor de k se busca así:
- Unidades y décimas en la columna de la izquierda
- Centésimas en la fila de arriba
- El número que nos da la tabla es el valor de : $\phi(k) = P[z \leq k]$

CÁLCULO DE PROBABILIDADES EN UNA DISTRIBUCIÓN N(0,1)

- Si $k \geq 0$, las probabilidades $\phi(k) = P[z \leq k] = P[z < k]$ se encuentran directamente en la tabla.
- $P[z \geq k] = 1 - P[z < k] = 1 - \phi(k)$
- Para abscisas negativas: $P[z \leq -k] = P[z \geq k] = 1 - \phi(k)$
- $P[a \leq z \leq b] = P[z \leq b] - P[z \leq a]$

CALCULO DE PROBABILIDADES EN UNA DISTRIBUCIÓN N(μ,σ)

Como ya sabemos, las probabilidades en dos distribuciones normales cualesquiera se reparten de forma análoga. Por tanto, para calcular probabilidades en una distribución N(μ,σ), la relacionaremos con la N(0,1) para la cual disponemos del recurso de las tablas.

Si x es N(μ,σ), para calcular la probabilidad $P[b < x < k]$ se procede del siguiente modo: $P[b < x < k] = P\left[\frac{b-\mu}{\sigma} < z < \frac{k-\mu}{\sigma}\right]$. El cambio $z \Rightarrow \frac{x-\mu}{\sigma}$ se llama tipificación de la variable. La variable ya tipificada sigue una distribución N(0,1)

12.2 – INTERVALOS CARACTERÍSTICOS

Si la variable x tiene una distribución de media μ, se llama **intervalo característico** correspondiente a una probabilidad p a un intervalo centrado en la media, $(\mu - k, \mu + k)$ tal que la probabilidad de que x pertenezca a dicho intervalo es p:

$$P[\mu - k < x < \mu + k] = p$$

INTERVALOS CARACTERÍSTICOS EN DISTRIBUCIONES N(0,1)

En una distribución normal N(0,1), si (-k,k) es el intervalo característico correspondiente a una probabilidad p, es decir, si $P[-k < z < k] = p = 1 - \alpha$

$\Rightarrow P[z > z_{\alpha/2}] = \alpha/2 \Rightarrow P[z \leq z_{\alpha/2}] = 1 - \alpha/2 \Rightarrow$ Intervalo característico = $(-z_{\alpha/2}, z_{\alpha/2})$

INTERVALOS CARACTERÍSTICOS EN DISTRIBUCIONES N(μ,σ)

En una distribución normal $N(\mu,\sigma)$, si $(\mu - k, \mu + k)$ es el intervalo característico correspondiente a una probabilidad p, es decir, si $P[\mu - k < X < \mu + k] = p = 1 - \alpha$

$\Rightarrow P[z > z_{\alpha/2}] = \alpha/2 \Rightarrow$

$P[z \leq z_{\alpha/2}] = 1 - \alpha/2 \Rightarrow$ Intervalo característico = $(\mu - z_{\alpha/2}.\sigma, \mu + z_{\alpha/2}.\sigma)$

12.3 – DISTRIBUCIÓN DE LAS MEDIAS MUESTRALES

DISTRIBUCIÓN DE LAS MEDIAS MUESTRALES

Dada una población de media μ y desviación típica σ, no necesariamente normal, la distribución de las medias de las muestras de tamaño n:

- Tiene la misma media, μ, que la población.
- Su desviación típica es $\sigma/\sqrt{n}$ y, por consiguiente, desminuye al aumentar n.
- Cuando $n \geq 30$ es prácticamente normal.

$$\left.\begin{array}{c} \text{Si } X \approx N(\mu,\sigma) \\ \text{ó} \\ \text{Si } n > 30 \end{array}\right\} \Rightarrow \overline{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

$P[z \leq z_{\alpha/2}] = 1 - \alpha/2 \Rightarrow$ Intervalo característico = $\left(\mu - z_{\alpha/2}.\frac{\sigma}{\sqrt{n}}, \mu + z_{\alpha/2}.\frac{\sigma}{\sqrt{n}}\right)$

DISTRIBUCIÓN DE LA SUMA DE TODOS LOS INDIVIDUOS DE LA MUESTRA

Puesto que $\sum_{i=1}^{n} x_i = n\overline{x}$, sabemos que $\sum_{i=1}^{n} x_i$ se distribuye normal de media $n\mu$ y desviación típica $n\frac{\sigma}{\sqrt{n}} = \sigma\sqrt{n}$

$$\left.\begin{array}{c} \text{Si } X \approx N(\mu,\sigma) \\ \text{ó} \\ \text{Si } n > 30 \end{array}\right\} \Rightarrow \sum_{i=1}^{n} x_i \sim N(n\mu, \sigma\sqrt{n})$$

$P[z \leq z_{\alpha/2}] = 1 - \alpha/2 \Rightarrow$ Intervalo característico = $(n\mu - z_{\alpha/2}.\sigma\sqrt{n}, \mu + z_{\alpha/2}.\sigma\sqrt{n})$

12.4 – EN QUÉ CONSISTE LA ESTADÍSTICA INFERENCIAL

ESTIMACIÓN PUNTUAL Y ESTIMACIÓN POR INTERVALOS

Los parámetros de la población se pueden estimar a partir de los de la muestra. Así:
La media muestral, $\overline{x}$, sirve para estimar la media poblacional, μ.
La desviación típica muestral, s, es una estimación de la desviación típica poblacional, σ

La estimación puntual (el valor de μ es aproximadamente $\overline{x}$), sirve de poco mientras desconozcamos cuál es el grado de aproximación de $\overline{x}$ a μ.

La estimación por intervalos: A partir de una muestra de tamaño n podemos estimar el valor de un parámetro de la población del siguiente modo:
- Dando un intervalo dentro del cual confiamos que esté el parámetro. Se llama intervalo de confianza.
- Hallando la probabilidad de que tal cosa ocurra. A dicha probabilidad se le llama nivel de confianza.

Cuanto mayor sea el tamaño de la muestra, mayor eficacia tendremos en nuestra estimación.

La eficacia de esta estimación se manifiesta de dos formas:
- En el tamaño del intervalo (cuanto más pequeño, más precisos estamos siendo)
- En el nivel de confianza (más nivel de confianza significa más seguridad en la estimación).

Tamaño de la muestra, longitud del intervalo y nivel de confianza son tres variables estrechamente relacionadas. Conocidas dos de ellas obtendremos la tercera.

12.5 – INTERVALO DE CONFIANZA PARA LA MEDIA

Se desea estimar la media, μ, de una población cuya desviación típica, σ, es conocida.

Para ello se recuerde a una muestra de tamaño n en la cual se obtiene una media muestral, $\overline{x}$.

Si la población de partida es normal, o si el tamaño de la muestra es $n \geq 30$, entonces el **intervalo de confianza** de μ con un nivel de confianza de $(1 - \alpha)$ .100% es:

$$\left(\overline{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

12.6 – RELACIÓN ENTRE NIVEL DE CONFIANZA, ERROR ADMISIBLE Y TAMAÑO DE LA MUESTRA

El valor E = $z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ se llama **error máximo admisible.**

Depende de α y de n del siguiente modo:
- Cuanto mayor sea el tamaño de la muestra, menor es E (más estrecho es el intervalo, es decir, más afinaremos en la estimación).
- Cuanto mayor sea 1 – a (es decir, cuanto más seguros queramos estar de nuestra estimación), mayor es E.

Cuanto mayor es 1 - α, mayor es $z_{\alpha/2}$ y, por tanto, mayor es E.

E, n y α son tres variables estrechamente relacionadas. Conocidas dos de ellas obtendremos la tercera despejando de la fórmula: E = $z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$