Resumen: Estadística Inferencial

- **Definición:** Llamamos **población** al conjunto en el que queremos estudiar una determinada característica. **Muestra** será una parte reducida y representativa de esta población.
- **Definición: Inferencia estadística** es el proceso que se encarga de obtener resultados sobre la población total a partir de resultados obtenidos en la muestra.

Existen diferentes técnicas para seleccionar una muestra. Las principales son:

- **Muestreo aleatorio simple:** Se eligen n elementos al azar entre los **N** que constituyen la población. La selección de los individuos puede hacerse con o sin reposición. En ambos casos todas las muestras de n elementos tienen la misma probabilidad de ser seleccionadas.
- **Muestreo aleatorio sistemático:** Se hace una lista con los elementos de la población de 1 a **N** y se elige aleatoriamente un número de salto $h \leq \frac{N}{n}$. A continuación se selecciona al azar un número $a \leq h$, que sirve de origen y, a partir de él, se seleccionan los elementos:
$$a,\ a+h,\ a+2h,\ \dots,\ a+(n-1)h$$
- **Muestreo aleatorio estratificado:** Los dos casos anteriores se usan cuando la población es homogénea. Si la población está separada es diferentes estratos es conveniente, a la hora de elegir los elementos que configuran la muestra, seleccionarlos de los diferentes estratos de acuerdo al peso que tenga dicho grupo dentro de la población total.

- **DISTRIBUCIÓN MUESTRAL DE MEDIAS**

Consideremos todas las posibles muestras de n elementos que puedan seleccionarse de una población de tamaño N en la que se ha estudiados la variable X con una media μ y una desviación típica σ. En cada una de esas muestras podemos buscar la media de la variable estadística que desee analizarse y podemos hablar, al considerarlas todas, de una distribución muestral de las medias $\overline{X}$.

Si X es normal o bien $n \geq 30 \Rightarrow \overline{X}$ es normal con una media $\mu_{\bar{x}} = \mu$ y una desviación típica $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$(muestreo con reposición), $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}$ (muestreo sin reposición).
Si la población es infinita o bien $5n \leq N \Rightarrow \sqrt{\frac{N-n}{N-1}} \approx 1$

- **DISTRIBUCIÓN MUESTRAL DE PROPORCIONES**

Sea ahora p la proporción de individuos de la población que poseen una característica determinada ($q = 1 - p$ será la proporción de los que no la cumplen). Tomadas todas la posibles muestras de tamaño n, el valor de la proporción en cada una de ellas constituye una variable que llamaremos P.

Tomaremos muestras tales que $n \geq 30 \Rightarrow P$ es normal con una media $\mu_p = p$ y una desviación típica $\sigma_p = \sqrt{\frac{p \cdot q}{n}}$ (muestreo con reposición), $\sigma_{\bar{x}} = \sqrt{\frac{p \cdot q}{n}} \cdot \sqrt{\frac{N-n}{N-1}}$ (muestreo sin reposición).
Si la población es infinita o bien $5n \leq N \Rightarrow \sqrt{\frac{N-n}{N-1}} \approx 1$

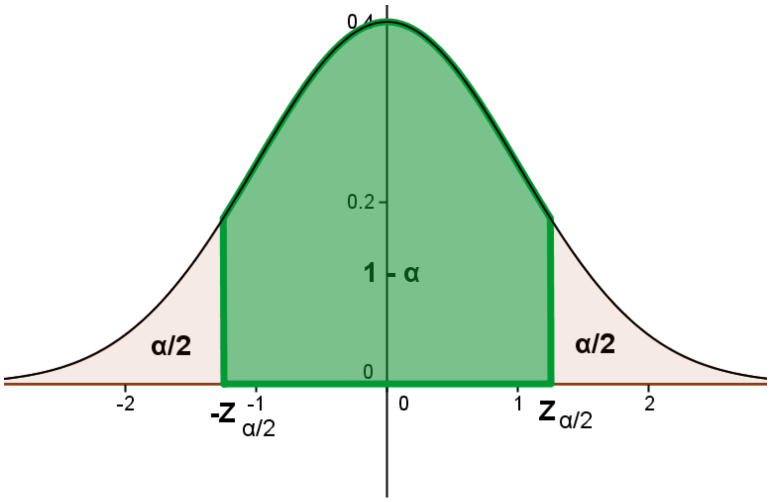
Resumen: Estadística Inferencial

❖ INTERVALO DE CONFIANZA DE LA MEDIA

Llamamos intervalo de confianza de la media a un intervalo **(a,b)** en el que se encuentra la media de la población con una probabilidad $1-\alpha$. Es decir, se cumple:

$$P(a \le \mu \le b) = 1-\alpha$$

El valor $1-\alpha$, expresado en porcentaje, recibe el nombre de nivel de confianza.
El intervalo de confianza se busca con la distribución muestral de medias.
Puesto que la variable $\overline{X}$ se distribuye según $N(\mu_{\bar{x}} = \mu, \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}})$, la variable tipificada $Z = \frac{\overline{X}-\mu_{\bar{x}}}{\sigma_{\bar{x}}}$ se distribuirá según una normal N(0,1) y se cumplirá:



$$P(-z_{\alpha/2} \le Z \le z_{\alpha/2}) = 1-\alpha$$

Si $\bar{x}_i$ es la media de la muestra elegida y para un nivel de confianza $1-\alpha$ se cumple:

$$-Z_{\frac{\alpha}{2}} < \frac{\bar{x}_i-\mu}{\sigma_{\bar{x}}} < Z_{\frac{\alpha}{2}} \quad \Rightarrow \quad -\sigma_{\bar{x}}.Z_{\frac{\alpha}{2}} < \bar{x}_i - \mu < \sigma_{\bar{x}}.Z_{\frac{\alpha}{2}} \quad \Rightarrow$$

$$\Rightarrow \quad \bar{x}_i - \sigma_{\bar{x}}.Z_{\frac{\alpha}{2}} < \mu < \bar{x}_i + \sigma_{\bar{x}}.Z_{\frac{\alpha}{2}}$$

Por lo que el intervalo buscado es:

$$\left(\bar{x}_i - \sigma_{\bar{x}}.Z_{\frac{\alpha}{2}}, \bar{x}_i + \sigma_{\bar{x}}.Z_{\frac{\alpha}{2}}\right)$$

Al valor $Z_{\alpha/2}$ se le llama valor crítico y los más usados son:

nivel de confianza	90 %	95 %	96 %	98 %	99 %
valor crítico	1'645	1'96	2'055	2'325	2'575

Nota: El intervalo de confianza anterior se elabora a partir de la desviación típica de la población. En caso de no conocerse esa desviación típica (lo más habitual), se elige la de la muestra. Esta simplificación no representa ningún problema siempre y cuando $n \ge 30$

❖ INTERVALO DE CONFIANZA DE LA PROPORCIÓN

Mediante un proceso similar al anterior, llegamos a la conclusión de que el intervalo de confianza para la proporción es:

Resumen: Estadística Inferencial

$$\left(p' - Z_{\alpha/2} \cdot \sqrt{\frac{p' \cdot q'}{n}}\,,\, p' + Z_{\alpha/2} \cdot \sqrt{\frac{p' \cdot q'}{n}}\right)$$

donde p' es la proporción de individuos con una característica determinada (la que se pretende analizar) en la muestra.

- **TAMAÑO DE LA MUESTRA**

El tamaño de la muestra seleccionada dependerá del error que estamos dispuestos a asumir, E, y del nivel de confianza que se exija.

Para la media: $E = Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \Rightarrow n = \left(Z_{\alpha/2} \cdot \frac{\sigma}{E}\right)^2$

Para la proporción: $E = Z_{\alpha/2} \cdot \sqrt{\frac{p \cdot q}{n}} \Rightarrow n = \left(Z_{\alpha/2}\right)^2 \cdot \frac{p \cdot q}{E^2}$