Intervalos de Confianza

1) Se ha comprobado que la concentración promedio de zinc que se saca del agua de un río a partir de una muestra de mediciones de zinc en 36 sitios diferentes es de 2.6 gramos por mililitro.

   a. Encontrar los intervalos de confianza del 95% y 99% para la concentración media de zinc en el río, suponiendo que la desviación típica de la población es 0.3.
   b. ¿Qué diferencia hay entre ambos intervalos? ¿Qué se puede decir que le pasa a la amplitud del IC cuando aumenta el Nivel de Confianza?

2) Determinar un intervalo de confianza al nivel $\alpha = 0.5$ para la probabilidad de que un recién nacido sea niño, si en una muestra de tamaño 123 se han contabilizado 67 niños.

3) El encargado del departamento de producción de una fábrica recibe un lote de 2000 piezas necesarias para el montaje de un artículo.
   a. Si se toma una muestra de 100 piezas elegidas al azar y se encuentran 4 defectuosas, determinar un intervalo de confianza para la proporción de piezas con defectos, considerando un nivel de confianza del 95%.
   a. Si el fabricante de las piezas asegura que en este lote no hay más de 100 piezas defectuosas, según nuestros resultados ¿deberíamos confiar en lo que dice y aceptar el lote?

4) Dos compañías A y B fabrican el mismo tipo de cable. Un distribuidor desea conocer la diferencia promedio de la resistencia a la rotura de los mismos, para lo cual toma muestras de 100 cables de A y 50 cables de B. La muestra de los cables de la compañía A arroja una resistencia promedio a la rotura de 4500 kilogramos, mientras que los cables de la compañía B arrojan una resistencia promedio a la rotura de 4000 kilogramos. Se sabe, por experiencia, que la desviación típica de la resistencia a la rotura es de 300 kilogramos para la compañía A y de 200 kilogramos para la compañía B.
   a. Estimar con un nivel de confianza del 95% el intervalo de confianza de la diferencia de medias de la resistencia a la rotura entre los dos cables, si la resistencia a la rotura se distribuye normalmente para ambas compañías.

5) Calcular un intervalo de confianza al 95% para la varianza poblacional mediante la muestra siguiente, que proviene de una $Normal(0, \sigma)$:

1.2, -2.2, -3.1, -0.2, 0.5, 0.6, -2.1, 2.2, 1.3

6) Para estudiar la efectividad de un medicamento contra la diabetes se mide la cantidad de glucemia en sangre antes y después de la administración de dicho medicamento, obteniéndose los resultados siguientes:

Antes	**7.2**	**7.3**	**6.5**	**4.2**	**3.1**	**5.3**	**5.6**
Después	**5.2**	**5.4**	**5.3**	**4.7**	**4.1**	**5.4**	**4.9**

a) Estimar un IC para la reducción de glucemia en sangre producida por el medicamento.

7) El peso de los adultos de una población numerosa se distribuye normalmente con varianza 9. ¿Cuántos individuos debemos seleccionar en la muestra si queremos que la media muestral no difiera en más de 1kg de la media de la población, con probabilidad 0.99?

TAREA

1) La estatura de los habitantes mayores de edad de una determinada ciudad sigue una distribución normal de media conocida y varianza 36 cm^2. En una muestra aleatoria de 80 individuos de esta ciudad, hemos obtenido una estatura media de 172cm.
   a. Determina un intervalo de confianza del 90% para la estatura media de los habitantes mayores de edad de dicha ciudad.
   b. Si aumentamos el nivel de confianza al 95%, ¿el intervalo de confianza será más amplio (más valores posibles) o más estrecho?
   c. ¿Qué necesitamos cambiar en nuestro problema para obtener un mejor intervalo cuando incrementamos la confianza?

Resumen de teoría y fórmulas

<u>Parámetros:</u> Los parámetros de la población se desconocen generalmente. Siendo uno de los primordiales principios de la estadística el estimarlos.

<u>Estimador:</u> El método empleado para estimarlo se denomina "estimador" del parámetro poblacional. Los estimadores dependen de los elementos de la muestra.

<u>Estimación puntual:</u> Hablamos de "estimación puntual" cuando sustituimos los valores de la muestra en el estimador, lo que da como resultado un valor puntual concreto del parámetro poblacional, que no tiene por qué ser exactamente su valor real, pero si el estimador tiene buenas propiedades, la estimación puntual no debería alejarse mucho del valor real del parámetro poblacional.

<u>Ejemplos de estimadores muestrales:</u> La media muestral, la varianza muestral, la cuasivarianza muestral.

<u>Estimación por Intervalos de Confianza:</u> Hablamos de "estimación por intervalos de confianza" cuando asignamos al parámetro un <u>intervalo de valores</u>, en vez de un valor puntual.

Supongamos que estimamos el parámetro θ mediante el estimador $\hat{\theta}$. Supongamos que tal estimador se distribuye normalmente con media $\mu_\theta = \theta$ (es un estimador insesgado), y desviación típica σ_θ. Por la distribución Normal (estandarizando) sabemos que:

$$P\left(-2.58 < \frac{\hat{\theta} - \mu_\theta}{\sigma_\theta} < 2.58\right) = 0.99$$

$$P\left(-2.58 < \frac{\hat{\theta} - \theta}{\sigma_\theta} < 2.58\right) = 0.99$$

$$P\left(\hat{\theta} - 2.58\,\sigma_\theta < \theta < \hat{\theta} + 2.58\,\sigma_\theta\right) = 0.99$$

<u>Nivel de confianza:</u> Es la probabilidad (por ejemplo 0.99 en el ejemplo anterior) de que el parámetro de la población (θ) se encuentre en un cierto intervalo. Se representa como $NC = 1 - \alpha$, donde α (en el ejemplo sería 0.01=1%) es el margen de error que se permite en la estimación.

<u>Nivel de significación:</u> Al margen de error α se le suele llamar nivel de significación.

<u>Intervalo final:</u> Para obtener el IC final, debemos fijar un α (por ejemplo $\alpha = 0.01$), y debemos saber qué características conocemos de la distribución poblacional.

Por ejemplo si es una Normal con varianza poblacional σ conocida, el IC final para la media poblacional μ será:

$$IC_{\alpha}(\mu) = \bar{x} \pm z_{\alpha} \frac{\sigma}{\sqrt{n}}$$

$$IC_{\alpha=0.01}(\mu) = \bar{x} \pm 2.58 \frac{\sigma}{\sqrt{n}}$$

En lo anterior, solo hemos considerado que nuestro estimador (de la media poblacional) es la media muestral $\hat{\theta} = \bar{x}$ y que al conocer la varianza poblacional podemos calcular la varianza del estimador $\sigma_{\theta} = \frac{\sigma}{\sqrt{n}}$. El z_{α} es el valor de la Normal estándar que deja a su derecha un α de probabilidad. En el caso de $\alpha = 0.01$, tenemos que $z_{\alpha} = 2.58$. Suponiendo que por ejemplo $\bar{x} = 10$, $\sigma = 1$ y $n = 16$, el intervalo quedaría:

$$IC_{\alpha=0.01}(\mu) = \left[10 - 2.58 \times \frac{1}{4}; 10 + 2.58 \times \frac{1}{4}\right] = [9.355,10.645]$$

De lo cual, tendríamos una estimación puntual de la media $\hat{\mu} = \bar{x} = 10$ y un intervalo de confianza (con un nivel de confianza del 99% = margen de error del 1%) que es un intervalo de posibles valores para la media poblacional: $[9.355,10.645]$.

Intervalos de confianza para los diferentes parámetros

Notación:

<u>Población</u>

μ: media poblacional

σ: varianza poblacional

π: proporción poblacional

<u>Muestra</u>

n: tamaño muestral

$\bar{x}$: media muestral

s^2: varianza muestral

s'^2: cuasivarianza muestral

$\hat{p}$: proporción muestral

r: coeficiente de correlación muestral

1. MEDIA
   a. Varianza de la población conocida.

   $$IC_{\alpha}(\mu) = \bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

   b. Varianza de la población desconocida y muestras grandes (n>30).

   $$IC_{\alpha}(\mu) = \bar{x} \pm z_{\alpha/2} \frac{s}{\sqrt{n-1}}$$

   c. Varianza de la población desconocida y muestras pequeñas (n<=30).

$$IC_\alpha(\mu) = \bar{x} \pm t_{n-1,\alpha/2}\frac{s}{\sqrt{n-1}}$$

2. VARIANZA

$$IC_\alpha(\sigma^2) = \left[\frac{ns^2}{\chi_1^2}; \frac{ns^2}{\chi_2^2}\right]$$

Donde $\chi_1^2 = \chi_{n,\frac{\alpha}{2}}^2$ y $\chi_2^2 = \chi_{n,1-\frac{\alpha}{2}}^2$

3. DIFERENCIA DE MEDIAS

a. Muestras independientes y varianzas de las poblaciones conocidas.

$$IC_\alpha(\mu_1 - \mu_2) = (\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

b. Muestras independientes y varianzas de las poblaciones desconocidas pero iguales.

i. Muestras grandes $(n_1 + n_2 - 2 > 50)$

$$IC_\alpha(\mu_1 - \mu_2) = (\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2}\sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

ii. Muestras pequeñas $(n_1 + n_2 - 2 \leq 50)$

$$IC_\alpha(\mu_1 - \mu_2) = (\bar{x}_1 - \bar{x}_2) \pm t_{n_1+n_2-2,\alpha/2}\sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

c. Muestras independientes y varianzas de las poblaciones desconocidas y distintas.

i. Muestras pequeñas pero de igual tamaño $n_1 = n_2 = n$

$$IC_\alpha(\mu_1 - \mu_2) = (\bar{x}_1 - \bar{x}_2) \pm t_{2n-2,\alpha/2}\sqrt{\frac{s_1^2 + s_2^2}{n-1}}$$

ii. En otro caso, calcularemos la expresión “gl=grados de libertad”:

$$gl = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1+1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2+1}} - 2$$

- Muestras grandes ($gl > 50$)

$$IC_\alpha(\mu_1 - \mu_2) = (\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2}\sqrt{\frac{s_1^2}{n_1 - 1} + \frac{s_2^2}{n_2 - 1}}$$

- Muestras pequeñas ($gl \leq 50$)

$$IC_\alpha(\mu_1 - \mu_2) = (\bar{x}_1 - \bar{x}_2) \pm t_{gl,\alpha/2}\sqrt{\frac{s_1^2}{n_1 - 1} + \frac{s_2^2}{n_2 - 1}}$$

d. Muestras relacionadas y varianza de la diferencia conocida σ_D^2

$$IC_\alpha(\mu_1 - \mu_2) = (\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2}\frac{\sigma_D}{\sqrt{n}}$$

e. Muestras relacionadas y varianza de la diferencia desconocida (muestras grandes)

- Primero calcular la varianza muestral de la diferencia:

$$s_D^2 = \frac{\sum(D_i - \overline{D})^2}{n}$$

$$IC_\alpha(\mu_1 - \mu_2) = (\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2}\frac{s_D}{\sqrt{n-1}}$$

f. Muestras relacionadas y varianza de la diferencia desconocida (muestras pequeñas)

$$IC_\alpha(\mu_1 - \mu_2) = (\bar{x}_1 - \bar{x}_2) \pm t_{n-1,\alpha/2}\frac{s_D}{\sqrt{n-1}}$$

4. COCIENTE DE VARIANZAS

a. Observaciones independientes

$$IC_\alpha\left(\frac{\sigma_1^2}{\sigma_2^2}\right) = \left[\frac{s_1'^2}{s_2'^2}F_1; \frac{s_1'^2}{s_2'^2}F_1\right]$$

5. PROPORCIÓN

a. Muestras grandes n>30

$$IC_\alpha(\pi) = \hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

b. Muestras pequeñas

$$IC_\alpha(\pi) = \hat{p} \pm t_{n-1,\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

6. DIFERENCIA DE PROPORCIONES

a. Muestras independientes con proporciones poblacionales iguales

$$IC_\alpha(\pi_1 - \pi_2) = \left(\hat{p}_1 - \hat{p}_2\right) \pm z_{\alpha/2} D_{est}$$

donde $D_{est} = \sqrt{\left(\frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1+n_2}\right)\left(1 - \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1+n_2}\right)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$

b. Muestras independientes con proporciones poblacionales desiguales

$$IC_\alpha(\pi_1 - \pi_2) = \left(\hat{p}_1 - \hat{p}_2\right) \pm z_\alpha\sqrt{\frac{\hat{p}_1\left(1-\hat{p}_1\right)}{n_1} + \frac{\hat{p}_2\left(1-\hat{p}_2\right)}{n_2}}$$