Tema 11: Intervalos de confianza.

Presentación y Objetivos.

En este tema se trata la estimación de parámetros por intervalos de confianza. Consiste en aproximar el valor de un parámetro desconocido por un intervalo en la recta real calculado a partir de la muestra aleatoria simple obtenida de la población.

Los objetivos de este tema son:

1. Comprender la finalidad de la estimación por intervalos como complemento a la estimación puntual.
2. Entender la relación que existe entre la expresión de un intervalo de confianza y la variable pivote de la que procede.
3. Interpretar correctamente el concepto de confianza.
4. Obtener intervalos de confianza cuando la distribución de la(s) población(es) sea Normal.
5. Obtener intervalos de confianza para proporciones.

Esquema Inicial.

1. Intervalos de confianza. Método de la variable pivote.
2. Intervalos de confianza en poblaciones normales.
3. Intervalos de confianza para proporciones.

Desarrollo del Tema.

1. Intervalos de confianza. Método de la variable pivote.

En la estimación por intervalos de confianza, en lugar de dar un valor concreto aproximado para el parámetro, se da una región o intervalo de la recta real en la que éste puede encontrarse con cierto grado de confianza. Se introduce el método con un ejemplo para posteriormente desarrollarlo teóricamente.

1.1. Ejemplo de motivación.

Sea X variable aleatoria con distribución $\mathcal{N}(\mu, 1)$. Se toma una muestra aleatoria simple $X_1, \dots, X_n$ de X. Se sabe que el estimador puntual de μ es la media muestral, $\hat{\mu} = \bar{X}$, y, además, $\bar{X} \sim \mathcal{N}(\mu, \frac{1}{\sqrt{n}})$. Se trata de encontrar dos estadísticos $T_1(X_1, \dots, X_n)$ y $T_2(X_1, \dots, X_n)$ de forma que, por ejemplo:

$$P(T_1(X_1, \dots, X_n) < \mu < T_2(X_1, \dots, X_n)) = 0{,}95.$$

En este caso concreto se sabe que:

$$Z = \frac{\bar{X} - \mu}{\frac{1}{\sqrt{n}}} \sim \mathcal{N}(0,1)$$

y, buscando en la tabla correspondiente, el intervalo que para la distribución $\mathcal{N}(0,1)$ encierra una probabilidad de 0,95 es (1,96;1,96):

$$P(-1{,}96 < Z < 1{,}96) = 0{,}95.$$

Sustituyendo Z se tiene:

$$P\left(-1{,}96 < \frac{\bar{X} - \mu}{\frac{1}{\sqrt{n}}} < 1{,}96\right) = 0{,}95,$$

y despejando µ para que quede en el centro de la desigualdad se obtiene:

$$P\left(\bar{X} - 1{,}96\frac{1}{\sqrt{n}} < \mu < \bar{X} + 1{,}96\frac{1}{\sqrt{n}}\right) = 0{,}95,$$

con lo que:

$$S_1(X_1, \dots, X_n) = \bar{X} - 1{,}96\frac{1}{\sqrt{n}}$$

$$S_2(X_1, \dots, X_n) = \bar{X} + 1{,}96\frac{1}{\sqrt{n}}.$$

Si se observa la muestra de tamaño $n = 10$:

$$4{,}31;\ \ 5{,}33;\ \ 5{,}14;\ \ 5{,}68;\ \ 6{,}27;\ \ 4{,}9;\ \ 3{,}32;\ \ 4{,}11;\ \ 4{,}47;\ \ 5{,}71$$

se tiene que $\bar{x} = 4{,}924\ y\ \ \bar{x} - 1{,}96\frac{1}{\sqrt{10}} = 4{,}3\ \ y\ \ \bar{x} + 1{,}96\frac{1}{\sqrt{10}} =\ 5{,}54.$ Se dirá que $(4{,}3; 5{,}54)$ es un intervalo al 95% de confianza para el parámetro µ.

1.2. Interpretación y metodología.

Se quieren encontrar dos estadísticos $T_1(X_1, \dots, X_n)$ y $T_2(X_1, \dots, X_n)$ tales que:

$$P(T_1(X_1, \dots, X_n) < \theta < T_2(X_1, \dots, X_n)) = 1 - \alpha,\ \ con\ \ \alpha \in (0,1)$$

y α pequeño. Si después de observar la muestra los valores de $T_1(X_1, \dots, X_n)$ y $T_2(X_1, \dots, X_n)$ son, respectivamente, los números a y b, se dirá que (a, b) es un intervalo de confianza al $(1-\alpha)100\%$ para el parámetro θ.

La expresión anterior no significa que la probabilidad de que el parámetro esté en el intervalo (a, b) sea de $1-\alpha$. El parámetro es una constante y, como tal, estará o no en un intervalo determinado. Antes de observar la muestra, se considera $1-\alpha$ como la probabilidad de que θ esté en el intervalo aleatorio $(T_1(\cdot), T_2(\cdot))$. Después de observarla, θ o está o no está en el intervalo calculado, no tiene sentido hablar de probabilidad y lo que se dice es que existe una *confianza del* $(1-\alpha)100\%$ *de que θ esté en* (a, b), interpretándose esta frase de la siguiente forma:

> Si se construyen muchos intervalos al $(1-\alpha)100\%$ de confianza, con muchas muestras, el $(1-\alpha)100\%$ de los mismos contendrá el verdadero valor del parámetro y el resto no.

En la práctica solamente se dispondrá de una muestra con la que se podrá construir un solo intervalo. En este intervalo, no tiene sentido hablar de la probabilidad de que el parámetro esté contenido en él, ya que dicho parámetro o está (probabilidad 1) o no está (probabilidad 0). Por ello, para expresar la incertidumbre que se tiene sobre si el intervalo calculado contiene o no el valor del parámetro desconocido, se utilizará el nivel de confianza.

1.3. Construcción.

Este método de construcción de intervalos que ya ha sido ilustrado con un ejemplo se llama método de la **variable pivote**. Se trata de encontrar una variable aleatoria que sea función de la muestra y del parámetro desconocido, de la que se conozca su distribución y, además, ésta no dependa del parámetro. En el ejemplo de la sección 6.1. esta variable era:

$$\frac{\bar{X}-\mu}{\frac{1}{\sqrt{n}}} \sim \mathcal{N}(0,1)$$

siendo μ el parámetro desconocido que se quiere estimar. A esta variable se le llama **variable pivote** ya que nos permite pivotar o pasar de una expresión del tipo:

$$P(\dots \leq Variable\ Pivote \leq \cdots)$$

a otra del tipo:

$$P(\dots \leq \theta \leq \cdots)$$

que es lo que se necesita para construir el intervalo de confianza.

Los valores de α, que determinan el nivel de confianza $(1-\alpha)100\%$, se escogen pequeños, típicamente 0,05; 0,01; 0,025, etc.

2. Intervalos de confianza en poblaciones normales.

A continuación se obtienen intervalos de confianza para los parámetros μ y σ^2 de una población normal bajo diversos supuestos.

2.1. Intervalos de confianza para la media de una población normal con varianza conocida.

Es el caso del ejemplo de la sección 6.1. En este caso se tiene:

Sea $(X_1, \dots, X_n)$ una muestra aleatoria simple de una variable aleatoria $X \sim N(\mu, \sigma)$, con μ desconocida y σ conocida. Se toma como variable pivote:

$$\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim \mathcal{N}(0,1)$$

En la variable pivote lo único desconocido es el parámetro μ. Por tanto y siendo $z_{\alpha/2}$ el punto de la distribución $\mathcal{N}(0,1)$ que deja a la derecha un área de $\frac{\alpha}{2}$:

$$P\left(-z_{\alpha/2} \leq \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \leq z_{\alpha/2}\right) = 1-\alpha.$$

Pivotando en la expresión anterior obtenemos el intervalo al $(1-\alpha)100\%$ de confianza para μ:

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

2.2. Intervalos de confianza para la media de una población normal con varianza desconocida.

Como σ es desconocida no se puede utilizar la variable pivote anterior. Se utiliza:

$$\frac{\bar{X}-\mu}{\frac{s}{\sqrt{n}}} \sim t_{n-1}$$

cuya distribución es una T-Student con n-1 grados de libertad. En esa variable pivote lo único desconocido es μ ya que s se calcula a partir de la muestra y en la distribución t no hay ningún parámetro desconocido ya que los grados de libertad es el tamaño muestral menos uno. Pivotando se obtiene el intervalo:

$$\left(\bar{X} - t_{n-1,\alpha/2}\frac{s}{\sqrt{n}}, \qquad \bar{X} + t_{n-1,\alpha/2}\frac{s}{\sqrt{n}}\right)$$

Ejemplo 1: En una muestra aleatoria simple de $n = 6$ coches americanos se obtienen los siguientes valores para la variable $X = consumo,\ en\ Km\ por\ litro$:

18,6; 18,4; 19,2; 20,8; 19,4; 20,5

Si $X \sim \mathcal{N}(\mu, \sigma)$, construir un intervalo al 95% de confianza para el consumo medio μ.

Con los seis datos se calcula:

$$\sum_{i=1}^{6} x_i = 116{,}9 \qquad\qquad \sum_{i=1}^{6} x_i^2 = 2282{,}41$$

para obtener:

$$\bar{x} = \frac{116{,}9}{6} = 19{,}483 \qquad s^2 = \frac{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}{n-1} = \frac{2282.41 - (6)(19{,}483)^2}{5} = 0{,}961$$

y, por tanto, $s = 0{,}98$. Como el intervalo es al 95% de confianza, $1 - \alpha = 0{,}94 \Rightarrow \alpha = 0{,}05$ y se necesita el percentil $t_{n-1,\alpha/2} = t_{5;0{,}025} = 2{,}571$. Sustituyendo el la fórmula del intervalo de confianza, se obtiene:

$$\left(19{,}483 - 2{,}571\frac{0{,}98}{\sqrt{6}}, \qquad 19{,}483 + 2{,}571\frac{0{,}98}{\sqrt{6}}\right) = (18{,}445; 20{,}511)$$

Ése sería el intervalo de confianza al 95% para el consumo medio.

2.3. Intervalos de confianza para la varianza de una población normal.

Se supone ahora que se tiene una muestra aleatoria simple $(X_1, \dots, X_n)$ de una variable aleatoria $X \sim \mathcal{N}(\mu, \sigma)$ de la que no se conoce µ. Se quiere construir un intervalo de confianza para la varianza σ^2. Se utiliza como variable pivote:

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

Por tanto,

$$P\left(\chi^2_{n-1,1-\frac{\alpha}{2}} \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi^2_{n-1,\frac{\alpha}{2}}\right) = 1 - \alpha.$$

Pivotando, se obtiene el intervalo de confianza al $(1-\alpha)100\%$:

$$\left(\frac{(n-1)S^2}{\chi^2_{n-1,\alpha/2}}, \frac{(n-1)S^2}{\chi^2_{n-1,1-\alpha/2}}\right)$$

Ejemplo 2: Con la muestra aleatoria simple del consumo de los coches americanos del ejemplo 1, construir un intervalo de confianza al 90% para la varianza del consumo, suponiendo que la variable $X = consumo,\ en\ Km\ por\ litro$ sigue una distribución normal.

Como $s^2 = 0{,}961$ y $1-\alpha = 0{,}90 \Rightarrow \alpha = 0{,}1$ y $\frac{\alpha}{2} = 0{,}05$, se necesitan los percentiles $\chi^2_{5;0,95}$ y $\chi^2_{5;0,05}$. Buscando en la tabla correspondiente, se tiene $\chi^2_{5;0,95} = 1{,}145$ y $\chi^2_{5;0,05} = 11{,}07$. Sustituyendo en la fórmula del intervalo:

$$\left(\frac{5\cdot(0{,}961)}{11{,}07}, \frac{5\cdot(0{,}961)}{1{,}145}\right) = (0{,}347; 3{,}354),$$

siendo éste el intervalo al 90% de confianza para la varianza del consumo de los coches.

2.4. Intervalo de confianza para la diferencia de medias en poblaciones normales.

Se supone que se tienen dos poblaciones normales independientes $X \sim \mathcal{N}(\mu_1, \sigma_1)$ e $Y \sim \mathcal{N}(\mu_2, \sigma_2)$. Se tiene una muestra aleatoria simple de cada una de ellas, de tamaños n_1 y n_2 respectivamente:

$$\left(X_1, \dots, X_{n_1}\right)\ \ m.a.s.\ de\ \ X \sim \mathcal{N}(\mu_1, \sigma_1)$$

$$\left(Y_1, \dots, Y_{n_2}\right)\ \ m.a.s.\ de\ \ Y \sim \mathcal{N}(\mu_2, \sigma_2)$$

Se quiere construir un intervalo de confianza para el parámetro diferencia de medias, es decir, $\theta = \mu_1 - \mu_2$. Es ésta una forma de comparar las dos poblaciones en base a sus medias.

2.4.1. Caso 1: Suponiendo varianzas desconocidas pero iguales.

Si se supone que la varianza de las dos poblaciones normales X e Y es la misma, i.e. $\sigma_1^2 = \sigma_2^2 = \sigma^2$, se define un estimador de esta varianza común, s_p^2:

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

No es más que una media ponderada de las cuasivarianzas de cada muestra s_1^2 y s_2^2. Con este nuevo estimador, se utiliza la variable pivote:

$$\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim t_{n_1+n_2-2}$$

Se obtiene el intervalo de confianza al $(1-\alpha)100\%$ de confianza:

$$\left((\bar{X}-\bar{Y})-t_{n_1+n_2-2,\alpha/2}s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}},\qquad (\bar{X}-\bar{Y})+t_{n_1+n_2-2,\alpha/2}s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right)$$

Ejemplo 3: Se quieren comparar dos sistemas A y B de matriculación *on line* en un curso a distancia. Para ello, se toma una muestra aleatoria simple de la variable $X = tiempo\ de\ matriculación\ con\ el\ sistema\ A$ y de la variable $Y = tiempo\ de\ matriculación\ con\ el\ sistema\ B$, que se suponen distribuidas normalmente y con la misma varianza. Calcular un intervalo de confianza al 98% para la diferencia de tiempos medios entre ambos sistemas de matriculación.

Sistema A (en minutos)	15	20	13	21	16	20
Sistema B (en minutos)	23	20	15	19	22	17

Tabla 1: Datos para el ejemplo 12.

Se realizan los cálculos de los estadísticos correspondientes que se resumen en la siguiente tabla:

	Sistema A	Sistema B
Media	$\bar{x}=17{,}5$	$\bar{y}=19{,}33$
Cuasivarianza	$s_1^2=10{,}7$	$s_2^2=9{,}06$

Tabla 2: Estadísticos del ejemplo 12.

Se calcula s_p^2:

$$s_p^2=\frac{5\cdot(10{,}7)+5\cdot(9{,}06)}{6+6-2}=9{,}88$$

con lo que $s_p=3{,}14$. Como $1-\alpha=0{,}98\ \Rightarrow \alpha=0{,}02$ y $\frac{\alpha}{2}=0{,}01$, se necesita el percentil $t_{10;0{,}01}=2{,}764$. Sustituyendo en la fórmula del intervalo se tiene:

$$\left((17{,}5-19{,}33)-(2{,}764)\cdot(3{,}14)\cdot\sqrt{\frac{1}{6}+\frac{1}{6}};\ (17{,}5-19{,}33)-(2{,}764)\cdot(3{,}14)\cdot\sqrt{\frac{1}{6}+\frac{1}{6}}\right)$$

Por tanto, el intervalo de confianza al 98% para la diferencia de tiempos medios es:

$$(-6{,}84; 3{,}18)$$

2.4.2. Caso 2: Suponiendo varianzas desconocidas.

Si las varianzas de ambas poblaciones no pueden suponerse iguales, se utiliza el siguiente intervalo aproximado:

$$\left((\bar{X}-\bar{Y}) - t_{n_1+n_2-2-\Delta,\frac{\alpha}{2}}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}, \qquad (\bar{X}-\bar{Y}) + t_{n_1+n_2-2-\Delta,\alpha/2}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}} \right)$$

siendo Δ el entero más próximo a:

$$\Delta = \frac{[(n_2-1)A-(n_1-1)B]^2}{(n_2-1)A^2+(n_1-1)B^2}$$

con $A=\frac{s_1^2}{n_1}$ y $B=\frac{s_2^2}{n_2}$. Se comprueba que $0 \leq \Delta \leq máx\,(n_1-1, n_2-1)$.

Ejemplo 4: Si se calcula el intervalo de confianza al 98% para los datos del ejemplo 3 sin suponer varianzas iguales se tiene que Δ es el entero más próximo a $0{,}068$. Como $\Delta = 0$, los grados de libertad no se modifican y el intervalo queda:

$$\left((17{,}5-19{,}33)-(2{,}764)\cdot\sqrt{\frac{10{,}7}{6}+\frac{9{,}06}{6}};\ (17{,}5-19{,}33)-(2{,}764)\cdot\sqrt{\frac{10{,}7}{6}+\frac{9{,}06}{6}} \right)$$

es decir, (-6,8457; 3,1857), prácticamente igual al del ejemplo anterior.

2.5. Intervalo de confianza para la razón de varianzas en poblaciones normales.

De nuevo, se supone que se tienen dos poblaciones normales independientes $X \sim \mathcal{N}(\mu_1, \sigma_1)$ e $Y \sim \mathcal{N}(\mu_2, \sigma_2)$. Se tiene una muestra aleatoria simple de cada una de ellas, de tamaños n_1 y n_2 respectivamente:

$$\left(X_1, \dots, X_{n_1}\right)\ \ m.a.s.\ de\ \ X \sim \mathcal{N}(\mu_1, \sigma_1)$$

$$\left(Y_1, \dots, Y_{n_2}\right)\ \ m.a.s.\ de\ \ Y \sim \mathcal{N}(\mu_2, \sigma_2)$$

Se quiere construir un intervalo de confianza para el parámetro razón de varianzas, es decir, $\theta = \frac{\sigma_1^2}{\sigma_2^2}$. Es ésta una forma de comparar las dos poblaciones en base a sus varianzas.

Se utiliza la variable pivote:

$$\frac{s_2^2}{s_1^2}\frac{\sigma_1^2}{\sigma_2^2} \sim F_{n_2-1,n_1-1}$$

y se obtiene el intervalo al $(1-\alpha)100\%$ de confianza:

$$\left(\frac{s_1^2}{s_2^2}F_{n_2-1,n_1-1,1-\frac{\alpha}{2}}\ , \qquad \frac{s_1^2}{s_2^2}F_{n_2-1,n_1-1,\frac{\alpha}{2}}\right).$$

Ejemplo 5: Con los datos del ejemplo 3, construir un intervalo de confianza al 95% para la razón de las varianzas del tiempo de matriculación con ambos sistemas.

Como $\alpha = 0{,}05; \frac{\alpha}{2} = 0{,}025$ se necesitan los percentiles:

$$F_{5,5;0,025} = 7{,}146 \qquad y \quad F_{5,5;0,975} = \frac{1}{F_{5,5;0,025}} = \frac{1}{7{,}146} = 0{,}139$$

El intervalo queda:

$$\left(\frac{10{,}7}{9{,}06}\cdot(0{,}139), \qquad \frac{10{,}7}{9{,}06}\cdot(7{,}146)\right)$$

con lo que la razón de varianzas se encuentra en el intervalo (0,16; 8,43) con un 95% de confianza.

3. Intervalos de confianza asintóticos. Intervalos de confianza para proporciones.

3.1. Intervalo de confianza para la media en Poblaciones No Normales.

Por el Teorema Central el Límite, se sabe que si X es una variable aleatoria de la que no se conoce su distribución, con $E(X) = \mu$ y $V(X) = \sigma^2 < \infty$, y $(X_1, \dots, X_n)$ es una muestra aleatoria de X, para n suficientemente grande:

$$\bar{X}_n \sim \mathcal{N}\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Por tanto, para muestras suficientemente grandes, el intervalo de confianza para la media, al $(1-\alpha)100\%$, será:

$$\left(\bar{X} - z_{\frac{\alpha}{2}}\frac{s}{\sqrt{n}}, \bar{X} + z_{\frac{\alpha}{2}}\frac{s}{\sqrt{n}}\right)$$

Este resultado se aplicará con $n \geq 30$.

Si la muestra no es lo suficientemente grande, puede usarse la desigualdad de Tchebychev.

3.1.1. Intervalo de confianza para una proporción.

Si se quiere estimar la proporción p de elementos de una población que tienen un determinado atributo, se obtiene una muestra aleatoria simple de una distribución de Bernoulli con parámetro p, es decir:

$$(X_1, \dots, X_n)\ m.a.s.\ de\ X \sim \mathcal{B}er(p)$$

en la que cada X_i es igual a 1 ó 0 según tenga o no el atributo. En este caso, el estimador puntual para esa proporción desconocida es:

$$\hat{p} = \frac{\sum_{i=1}^{n} X_i}{n}$$

que se corresponde con la proporción muestral de elementos con el atributo. Por el Teorema de De Moivre, el intervalo de confianza al $(1-\alpha)100\%$ será:

$$\left(\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \qquad \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$$

Ejemplo 6: Se observa que un nuevo dispositivo para acelerar el acceso a una red proporciona 73 conexiones con éxito en menos de 0,022 segundos de un total de 115 intentos de conexión. Construir un intervalo de confianza al 90% para la proporción p de conexiones sin error en ese tiempo.

En este caso, la proporción muestral de éxitos observados es $\hat{p}_1 = \frac{73}{115} = 0{,}635$. El percentil de la Normal es $z_{0,05} = 1{,}645$. El intervalo será:

$$\left(0{,}635 - 1{,}645\sqrt{\frac{0{,}635(1-0{,}635)}{115}}, 0{,}635 + 1{,}645\sqrt{\frac{0{,}635(1-0{,}635)}{115}}\right).$$

Por tanto, con un 90% de confianza se puede afirmar que la proporción de conexiones con éxito se encuentra en $(0{,}561; 0{,}709)$.

3.2. Intervalo de confianza para la diferencia de medias en Poblaciones No Normales.

Si se tienen dos muestras de tamaño n_1 y n_2 respectivamente de sendas normales independientes X e Y , con $n_1\ y\ n_2\ \geq 30$, el intervalo al $(1-\alpha)100\%$ de confianza para la diferencia de medias $\mu_1 - \mu_2$ es:

$$\left((\bar{X}-\bar{Y}) - z_{\alpha/2}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}, \qquad (\bar{X}-\bar{Y}) + z_{\alpha/2}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}\right).$$

3.2.2. Intervalo de confianza para la diferencia de proporciones.

En este caso se tienen dos muestras de tamaños n_1 y n_2 de dos poblaciones independientes de Bernoulli, es decir,

$$\left(X_1, \dots, X_{n_1}\right) \ \ m.a.s.\ de\ \ X \sim \mathcal{B}er(p_1)$$

$$\left(Y_1, \dots, Y_{n_2}\right) \ \ m.a.s.\ de\ \ Y \sim \mathcal{B}er(p_2).$$

El intervalo de confianza al $(1-\alpha)100\%$ para la diferencia $p_1 - p_2$ será:

$$\left((\hat{p}_1 - \hat{p}_2) - z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}, \qquad (\hat{p}_1 - \hat{p}_2) + z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}} \right)$$

donde $\hat{p}_1 = \frac{\sum x_i}{n_1}$ y $\hat{p}_2 = \frac{\sum y_i}{n_2}$ son las proporciones en cada muestra de elementos con el atributo en estudio y $\hat{q}_i = 1 - \hat{p}_i, i = 1,2$.

Ejemplo 7: Se quiere comparar el sistema de acceso del ejemplo 6 con otro protocolo que proporciona 72 accesos con éxito de un total de 100 intentos. Construir un intervalo de confianza al 98% para la diferencia entre las proporciones de acceso con éxito mediante los dos protocolos.

Las proporciones muestrales de éxitos observadas son:

$$\hat{p}_1 = \frac{73}{115} = 0{,}635 \ \ \ para\ el\ protocolo\ I$$

$$\hat{p}_2 = \frac{72}{100} = 0{,}72 \ \ \ para\ el\ protocolo\ II$$

Como $n_1\ y\ n_2\ \geq 30$ y $z_{0{,}01} = 2{,}33$, el intervalo de confianza queda:

$$(-0{,}23; 0{,}062)$$