5

ESTIMACIÓN

5.1 INTRODUCCIÓN

Los métodos de inferencia sirven para determinar la probabilidad de que cualquier conclusión sobre una población que se haya derivado de la información aportada por un grupo de datos sea correcta. Los valores de los estadísticos muestrales, por muy bueno que haya sido el muestreo, siempre presentarán diferencias con respecto al respectivo valor poblacional o parámetro, debido fundamentalmente a que se está tratando con variables aleatorias que asumen valores distintos y que ocurren en la población con frecuencias diferentes. De modo que al ser imposible eliminar la aleatoriedad y si se quieren hacer generalizaciones a partir de la información obtenida de una muestra se debe establecer la confianza que se tiene en la muestra. Es decir se debe determinar que tan buena es la aproximación entre valor del estadístico y el valor del parámetro respectivo. En éste punto la estadística inferencial es de gran ayuda al ofrecer métodos que cuantifican el grado de confianza requerido para hacer las generalizaciones mencionadas anteriormente. Son dos los métodos de inferencia, en unos se usa la información proporcionada por los estadísticos muestrales para estimar con cierta probabilidad el valor de un parámetro poblacional; el otro tipo de método, usa esa misma información para decidir con una probabilidad conocida si el parámetro poblacional es igual a algún valor preconcebido. El primero de estos procedimientos se conoce como Estimación y el segundo como Prueba de Hipótesis. En éste capítulo nos ocuparemos de los métodos de estimación y dejaremos para el próximo las pruebas de hipótesis.

La estimación de un parámetro se puede hacer en forma puntual o construyendo un intervalo. A continuación revisaremos estas dos técnicas de inferencia.

5.2 ESTIMACIÓN PUNTUAL

Una estimación puntual consiste en calcular en una muestra el valor de un estadístico y considerar que el mismo es la mejor aproximación que se tiene a la magnitud del parámetro poblacional correspondiente. Por ejemplo, un valor cualquiera de una media muestral ($\bar{x}$) es una estimación puntual de la media poblacional (μ). Igualmente un valor determinado de la varianza muestral (S^2) es una estimación puntual de la varianza poblacional (σ^2).
Un mismo parámetro puede tener varios estimadores. Así tenemos que la media poblacional (μ) además de poder ser estimada por la media muestral ($\bar{x}$), también es estimada por la mediana ($\tilde{x}$) y por la moda (Mo) para una variable que se distribuye en forma simétrica (Figura 5.1)

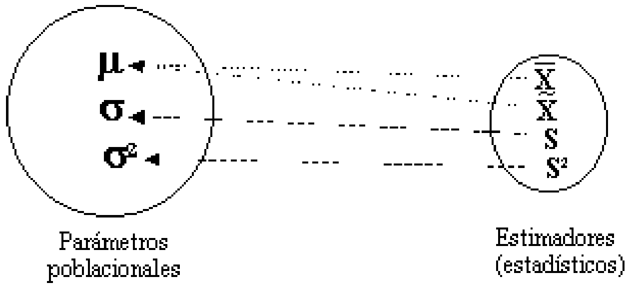


Figura 5.1

La escogencia del mejor estimador de un parámetro se dificulta, porque además de existir varios estimadores para un mismo parámetro, ellos son variables aleatorias que pueden tener una amplia distribución de valores. El mejor estimador siempre será aquel que esté más cerca del valor del parámetro que se estima. Como esto no se puede conocer, la calidad de un estimador se debe evaluar en términos de algunas de sus propiedades como son: la insesgabilidad, la consistencia y la eficiencia.
Supongamos que se está estimando un parámetro poblacional cualquiera, que será representado por la letra griega θ y su respectivo estimador será representado por la misma letra griega con un copete $\hat{\theta}$ (Figura 5.2)

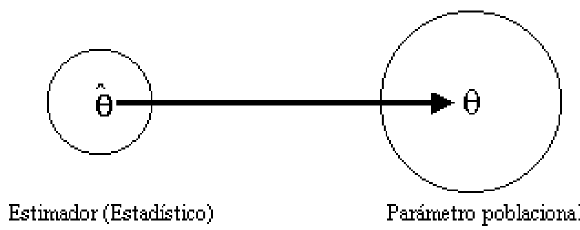


Figura 5.2

5.2.1 Estimador insesgado

Se dice que un estimador $\hat{\theta}$ del parámetro θ es insesgado cuando el valor esperado o promedio de la distribución de $\hat{\theta}$ coincide con el valor del parámetro θ:

$$E(\hat{\theta}) = \theta$$

Por ejemplo, la media muestral $\overline{x}$ es un estimador insesgado de μ, debido a que la media de las medias muestrales $\mu_{\overline{X}}$ es igual a la media poblacional μ_X

$$E(\overline{x}) = \mu_{\overline{X}} = \mu_X$$

Igualmente, la mediana de una muestra ($\tilde{x}$) es un estimador insesgado de μ, porque la media de las medianas muestrales es igual a la media poblacional, cuando la distribución de la variable estudiada es simétrica.

$$E(\tilde{x}) = \mu_X$$

En cambio la varianza muestral puede ser un estimador sesgado si para su cálculo se usan n grados de libertad.

$$E(S^2) \neq \sigma^2 \quad si \quad S^2 = \frac{\sum (x_i - \overline{x})^2}{n}$$

Para hacer insesgada la varianza muestral, la misma debe calcularse usando n-1 grados de libertad, de modo que:

$$E(S^2) = \sigma^2 \quad si \quad S^2 = \frac{\sum (x_i - \overline{x})^2}{n-1}$$

5.2.2 Estimador consistente

Se dice que un estimador $\hat{\theta}$ del parámetro θ es consistente si el valor absoluto de la diferencia entre los valores del estimador y del parámetro es menor a medida que aumenta el tamaño de la muestra (n). En términos más formales:

$$\lim_{n \to \infty} P\left(\left|\hat{\theta} - \theta\right| \leq \varepsilon \right) = 1 \qquad para\ todo \quad \varepsilon > 0$$

Ya sabemos que la media y la mediana muestrales son estimadores insesgados de μ, pero ¿son igualmente consistentes?. La respuesta es afirmativa si la distribución de la variable estudiada es simétrica. Pero si la variable se distribuye asimétricamente la mediana muestral se aproximará mas al valor de la mediana poblacional cuando n aumenta y la media muestral se acercará mas a la media poblacional (μ). Recordemos que la media poblacional y la mediana poblacional son dos parámetros diferentes. De lo dicho anteriormente se puede concluir que la media muestral es más consistente que la mediana muestral como estimador de la media poblacional (μ). Si comparamos la media y la mediana en términos de la insesgabilidad y consistencia, ambas son buenos estimadores de μ, siempre y cuando las muestras provengan de poblaciones con una distribución simétrica. En cambio, si la población se distribuye en forma asimétrica, la media es mejor estimador de μ, porque además de ser insesgada es más consistente.

5.2.3 Estimador eficiente

Se dice que un estimador $\hat{\theta}_1$ del parámetro θ es el más eficiente si no existe otro estimador $\hat{\theta}_2$ cuya varianza sea menor a la de $\hat{\theta}$.

$$E(\hat{\theta}_1 - \theta)^2 \ < \ E(\hat{\theta}_2 - \theta)^2$$

Si continuamos con la comparación entre la media y la mediana muestral como estimadores de μ, es necesario determinar para el caso de poblaciones con una distribución simétrica, cual de los dos estadísticos es mejor estimador de la media poblacional. Por lo tanto es necesario usar otras propiedades diferentes a la insesgabilidad y la consistencia. Cuando se examina la eficiencia de los dos estimadores, se encuentra que la varianza de la media muestra es menor que la varianza de la mediana muestral: $\sigma^2_{\tilde{x}} = (1{,}57)\ \sigma^2_{\overline{x}}$

Por lo tanto, en función de la insesgabilidad, consistencia y eficiencia, la media muestral ($\overline{x}$) es un mejor estimador de μ que la mediana muestral ($\tilde{x}$) para variables con distribuciones tanto simétricas como asimétricas.

5.3 ESTIMACIÓN POR INTERVALO

Aunque un estimador como la media muestral sea insesgado, consistente y eficiente, lo más probable es que, aún en muestras grandes, el valor del estimador ($\hat{\theta}$) no coincida con el valor del parámetro (θ). Por lo tanto se utiliza otro procedimiento más seguro para inferir el valor del parámetro, como es la estimación por intervalo.

Con éste método se construye un intervalo a partir del valor de un estimador puntual ($\hat{\theta}$), mediante la definición de dos límites: uno superior (LS) y otro inferior (LI). Se supone que el intervalo contiene el parámetro poblacional (θ) con cierta probabilidad:

5.3.1 Intervalo de confianza para una media poblacional

La deducción de un intervalo de confianza para la media poblacional depende de varios aspectos que combinados de cierta manera conforman una situación particular que determina la forma del intervalo. Los aspectos a considerar en la construcción de un intervalo de confianza son: i) el tipo de distribución de la variable estudiada, ii) el conocimiento de la varianza poblacional, y iii) el tamaño de la muestra. A continuación estudiaremos las distintas situaciones o casos que se pueden presentar en el desarrollo de un intervalo de confianza.

5.3.1.1 Caso 1: Muestreo en una población distribuida normalmente y con varianza conocida. Este primer caso servirá para analizar el proceso de generación de un intervalo de confianza. Supóngase que se desea estimar el valor de la media poblacional de una variable que se distribuye normalmente con varianza conocida (σ_X^2), para lo cual se extrae una muestra de tamaño n y se calcula la media de la muestra ($\bar{x}$). El valor de $\bar{x}$ es uno del total que conforman la población de valores de la variable aleatoria $\bar{X}$ que como se sabe se distribuye normalmente alrededor de una media $\mu_{\bar{X}}$ con varianza σ_X^2/n tal y como se ilustra en la Figura 5.3

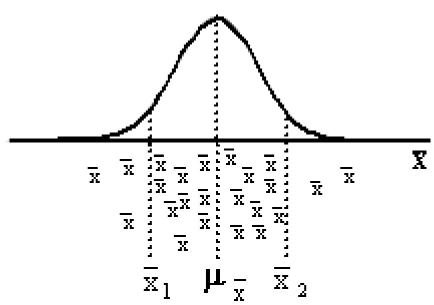


Figura 5.3. Ubicación figurativa de diferentes medias obtenidas de la población de valores de $\bar{X}$ (Explicación en el texto)

En esta población se pueden encontrar dos valores $\bar{x}_1$ y $\bar{x}_2$ separados simétricamente de $\mu_{\bar{X}}$ que definen un intervalo dentro del cual queda incluído una proporción $(1-\alpha)$ del total de valores de $\bar{X}$. Los valores $\bar{x}_1$ y $\bar{x}_2$ se encuentran transformando la variable $\bar{X}$ en la variable ***Z*** (Figura 5.4).

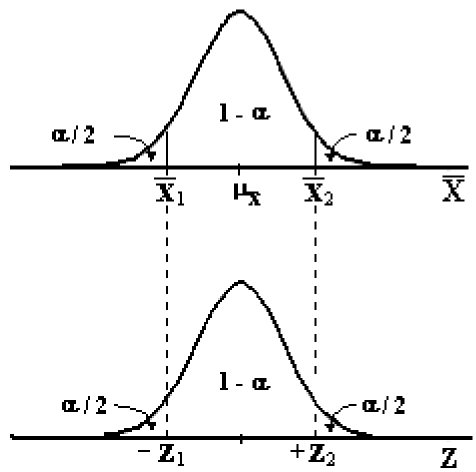


Figura 5.4. Relación entre la distribución de $\overline{X}$ y de **Z.** La expresión $1-\alpha$ representa la proporción de la población incluída entre dos valores de $\overline{X}$ o de **Z.**

La transformación de $\overline{X}$ a **Z** se efectúa a través de las expresiones siguientes:

$$-z_1 = \frac{\overline{x}_1 - \mu_{\overline{x}}}{\sigma_x / \sqrt{n}} \qquad\qquad + \; z_2 = \frac{\overline{x}_2 - \mu_{\overline{x}}}{\sigma_x / \sqrt{n}}$$

Para hacer más general la situación, las dos ecuaciones anteriores pueden expresarse de la manera siguiente:

$$- \; z_{(1-\alpha/2)} = \frac{\overline{x}_1 - \mu_{\overline{x}}}{\sigma_x / \sqrt{n}} \qquad\qquad + \; z_{(1-\alpha/2)} = \frac{\overline{x}_2 - \mu_{\overline{x}}}{\sigma_x / \sqrt{n}}$$

donde $+z_{(1-\alpha/2)}$ es el valor de Z a la izquierda del cual se encuentra una fraccíon del área igual a $1-\alpha/2$. Por simetría el valor de $-z_1$ es igual al valor de $+z_2$. Estos valores de Z se encuentran en la tabla de áreas de la distribución de Z. Por lo tanto es posible conocer los valores $\overline{x}_1$ y $\overline{x}_2$ mediante un despeje en las dos expresiones anteriores.

$$\overline{x}_1 = \mu_{\overline{x}} - \; z_{(1-\alpha/2)}\,\sigma_x / \sqrt{n} \qquad \text{y} \qquad \overline{x}_2 = \mu_{\overline{x}} + \; z_{(1-\alpha/2)}\,\sigma_x / \sqrt{n}$$

Los valores de $\overline{x}_1$ y $\overline{x}_2$ representan el límite inferior y superior del intervalo que contiene el $(1-\alpha)$100% de los valores de $\overline{X}$. Este intervalo puede expresarse de la manera siguiente: $\left[\; \mu_{\overline{x}} \pm \; z_{(1-\alpha/2)}\,\sigma_x / \sqrt{n} \;\right]$. La proporción de medias muestrales que se espera queden dentro del intervalo depende del valor de $z_{(1-\alpha/2)}$. A continuación se presentan algunos intervalos y la proporción de valores de $\overline{X}$, que se espera esté contenida dentro de ellos.

El intervalo $\left[\ \mu_{\overline{x}} \pm\ \ 1{,}29\ \ \sigma_x \big/ \sqrt{n}\ \right]$ contendrá cerca del 80% de los valores de $\overline{X}$.

El intervalo $\left[\ \mu_{\overline{x}} \pm\ \ 1{,}44\ \ \sigma_x \big/ \sqrt{n}\ \right]$ contendrá cerca del 85% de los valores de $\overline{X}$.

El intervalo $\left[\ \mu_{\overline{x}} \pm\ \ 1{,}65\ \ \sigma_x \big/ \sqrt{n}\ \right]$ contendrá cerca del 90% de los valores de $\overline{X}$.

El intervalo $\left[\ \mu_{\overline{x}} \pm\ \ 1{,}96\ \ \sigma_x \big/ \sqrt{n}\ \right]$ contendrá cerca del 95% de los valores de $\overline{X}$.

El intervalo $\left[\ \mu_{\overline{x}} \pm\ \ 2{,}58\ \ \sigma_x \big/ \sqrt{n}\ \right]$ contendrá cerca del 99% de los valores de $\overline{X}$.

La construcción de un intervalo como los anteriores no resuelve el problema de estimar $\mu_{\overline{x}}$, porque precisamente desconocemos su valor y no hay forma de encontrar los límites que definan un intervalo. Pero supongáse que se construye a partir de una media muestral cualquiera, un intervalo similar al siguiente: $[\ \overline{x} \pm\ z_{(1-\alpha/2)}\sigma_x \big/ \sqrt{n}\ ]$. Este intervalo contendrá a $\mu_{\overline{x}}$ siempre y cuando el valor de la $\overline{x}$ se encuentre entre los límites del intervalo $[\ \mu_{\overline{x}} \pm\ z_{(1-\alpha/2)}\sigma_x \big/ \sqrt{n}\ ]$. Por ejemplo, si construímos un intervalo para cada una de las medias muestrales de la Figura 5.3 y cada uno tiene el mismo tamaño que el intervalo construído a partir de $\mu_{\overline{x}}$. La mayoría de estos intervalos incluirán al valor de $\mu_{\overline{x}}$. Solamente aquellos intervalos generados a partir de aquellas pocas medias muestrales que se encuentran muy alejados de la media poblacional no incluyen a ésta última (Figura 5.5).

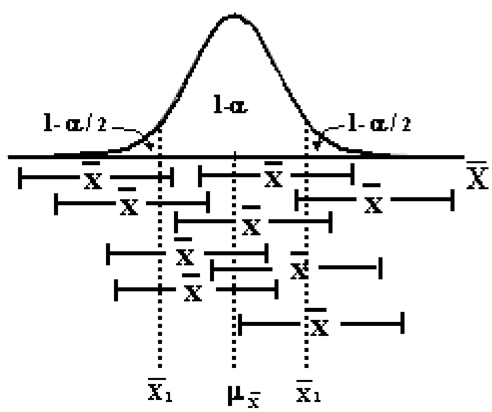


Figura 5.5. Intervalos del mismo tamaño construídos con algunas medias muestrales.

Se puede ver que solo aquellos intervalos construídos de medias cuya probabilidad de ocurrencia es muy baja, es decir con valores menores a $\overline{X}_1$ ó mayores a $\overline{X}_2$, no incluyen a $\mu_{\overline{x}}$. De modo que un intervalo de la forma $[\ \overline{x} \pm z_{(1-\alpha/2)}\sigma_x \big/ \sqrt{n}\ ]$ recibe el nombre de intervalo de confianza del (1-α)100%. Los valores extremos se denominan límites de confianza**,** existiendo un límite superior ($LS\ =\ \overline{x}\ +\ \ z_{(1-\alpha/2)}\sigma_x \big/ \sqrt{n}$) y un límite inferior ($LI\ =\ \overline{x}\ -\ z_{(1-\alpha/2)}\sigma_x \big/ \sqrt{n}$). El término $z_{(1-\alpha/2)}$ recibe el nombre de coeficiente de confiabilidad. La fracción 1-α se denomina nivel de confianza y representa la probabilidad de que el intervalo contenga el parámetro poblacional. Consecuentemente α representa la probabilidad de que el intervalo no contenga el parámetro poblacional

El tamaño del intervalo cambia en forma inversa al valor de α. Por ejemplo si $\alpha = 0{,}05$, se tiene un intervalo de confianza del 95% cuya amplitud es menor que un intervalo de confianza del 99% que resulta de haber definido un $\alpha = 0{,}01$. Por supuesto que a mayor amplitud del intervalo aumenta la probabilidad de que el parámetro esté incluido dentro del intervalo dado, pero también es mayor la incertidumbre sobre el valor del parámetro. Lo ideal sería construir intervalos estrechos con un alto nivel de confianza.

Cuando en una situación real se construye un intervalo de confianza, la media poblacional puede o no estar incluída dentro del intervalo. Sin embargo existe una probabilidad igual a 1-α de que el parámetro quedará incluído. Si el intervalo se construyó con un nivel de confianza del 95%, se dice que existe una probabilidad de 0,95 de que la media poblacional esté contenida dentro del intervalo. Otra forma de decirlo es, que si se construyen infinidad de intervalos similares, el 95% de los mismos contendrán a la media poblacional. Es importante adevertir que es un error generalizado el de señalar que la media poblacional se encuentra entre los valores de los límites con un $(1-\alpha)100\%$ de confianza, porque la media poblacional como cualquier otro parámetro es un valor fijo, y la afirmación anterior sugiere que el parámetro puede asumir cualquier valor entre los dos límites con cierta probabilidad.

Si se analiza con un poco más de detalle la relación entre los intervalos construídos a partir de las medias muestrales y la media poblacional, se observa que ambas cantidades se encuentran alejadas cierta distancia ε (Figura 5.6)

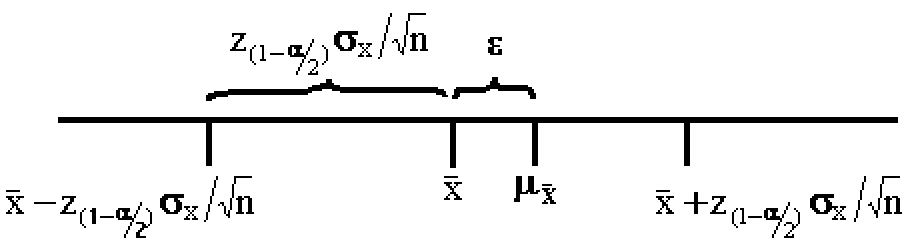


Figura 5.6. Error de estimación

La distancia ε se denomina error de estimación. Para que un intervalo contenga a la media poblacional con una probabilidad igual a 1 - α, ese error de estimación debe ser menor a la distancia $z_{(1-\alpha/2)}\sigma_x/\sqrt{n}$. Por extensión se puede decir que el error máximo ε_m permitido para que el intervalo contenga $(1-\alpha)100\%$ de las veces la media poblacional es igual a:

$$\varepsilon_m = \left| z_{(1-\alpha/2)}\sigma_x/\sqrt{n} \right|$$

Una consecuencia directa de conocer el valor de ε_m es que permite determinar cuál debe ser el tamaño muestral adecuado para cometer ese error máximo un $(1-\alpha)100\%$ de las veces, dado que:

$$n = \left(\frac{z_{(1-\alpha/2)}\sigma_x}{\varepsilon_m} \right)^2$$

Ejemplo 5.1
Al examinar 9 porciones de agua se encontró una concentración promedio de ión nitrato igual a 0,5 μg/ml. Se desea estimar mediante un intervalo de confianza del 95% la concentración promedio del nitrato en el agua, si se sabe que la desviación del método para éste análisis es de 0,15 μg/ml.

El intervalo que se requiere es de la forma $\left[\, \bar{x} \pm \; z_{(1-\alpha/2)}\,\sigma_x/\sqrt{n} \,\right]$ teniendo como límites los valores siguientes:

$$Li = \bar{x} - z_{(1-\alpha/2)}\,\sigma_x\Big/\sqrt{n} = \bar{x} - z_{(0{,}975)}\,\sigma_x\Big/\sqrt{n} = 0{,}5 - 1{,}96\left(0{,}15\Big/\sqrt{9}\right) = 0{,}4020\ \mu_g/ml$$

$$Ls = \bar{x} + z_{(1-\alpha/2)}\,\sigma_x\Big/\sqrt{n} = \bar{x} + z_{(0{,}975)}\,\sigma_x\Big/\sqrt{n} = 0{,}5 + 1{,}96\left(0{,}15\Big/\sqrt{9}\right) = 0{,}5980\ \mu_g/ml$$

El intervalo buscado es $\left[0{,}4020\ \mu_g\ /\ ml;\ 0{,}5980\ \mu_g\,/\ ml\ \right]$. Se concluye que se tiene un 95% de confianza de que la concentración promedio del ión nitrato en el agua se encuentra incluída dentro de éste intervalo.

También se puede decir que el error máximo de estimación con un 95% de confianza es:

$$\varepsilon_m = \left| z_{(1-\alpha/2)}\,\sigma_x\Big/\sqrt{n} \right| = \left| 1{,}96\left(0{,}15\Big/\sqrt{9}\right) \right| = 0{,}098\ \mu_g/ml$$

Por consiguiente, el tamaño de muestra necesario para cometer éste error un 95% de las veces será igual a:

$$n = \left(\frac{z_{(1-\alpha/2)}\sigma_x}{\varepsilon_m}\right)^2 = \left(\frac{z_{(0{,}975)}\sigma_x}{\varepsilon_m}\right)^2 = \left(\frac{1{,}96\ (015)}{0{,}098}\right)^2 = 9$$

Ahora bien, si se desea aumentar el nivel de confianza, por ejemplo a un 99%, sin aumentar el error de estimación, el tamaño de la muestra debe ser igual a:

$$n = \left(\frac{z_{(1-\alpha/2)}\sigma_x}{\varepsilon_m}\right)^2 = \left(\frac{z_{(0{,}995)}\sigma_x}{\varepsilon_m}\right)^2 = \left(\frac{2{,}58\ (0{,}15)}{0{,}098}\right)^2 \cong 16$$

Por otra parte, si se quiere reducir el error de estimación a unos 0,05 $\mu_g\ /\ ml$, manteniendo el nivel de confianza del 95%, entonces el tamaño muestral debe ser:

$$n = \left(\frac{z_{(1-\alpha/2)}\sigma_x}{\varepsilon_m} \right)^2 = \left(\frac{z_{(0,975)}\sigma_x}{\varepsilon_m} \right)^2 = \left(\frac{1,96\ (0,15)}{0,05} \right)^2 \cong 35$$

5.3.1.2 Caso 2: Muestreo a partir de una población distribuida normalmente, con varianza desconocida y tamaño de muestra grande ($n \geq 30$ **).**

La situación más común cuando se trata de estimar el valor de una media poblacional mediante un intervalo de confianza es que no solo se desconoce el valor de μ sino también el de la varianza poblacional σ_x^2 . Cuando se presenta una situación como la descrita, se puede utilizar la varianza de la muestra (S_x^2) como una estimación puntual de la varianza poblacional (σ_x^2). Si el tamaño de la muestra es grande $(n \geq 30)$, el estadístico $(\bar{x} - \mu_{\bar{x}})/(S_x/\sqrt{n})$ se distribuye normalmente, quedando el intervalo de confianza de la forma $\left[\bar{x} \pm z_{(1-\alpha/2)} S_x/\sqrt{n} \right]$

Ejemplo 5.2

Al determinar el valor de pH de una solución *buffer*, se encontró que 36 mediciones produjeron un valor promedio de pH igual a 5,2 con una desviación de 1,3 unidades. Estime mediante un intervalo el verdadero valor de pH de la solución con una confianza del 90%.

El intervalo que se requiere es de la forma $\left[\bar{x} \pm z_{(1-\alpha/2)} S_x/\sqrt{n} \right]$ teniendo como límites los valores siguientes:

$$LI = \bar{x} - z_{(1-\alpha/2)} S_x/\sqrt{n} = \bar{x} - z_{(0,95)} S_x/\sqrt{n} = 5,2 - 1,65 \left(1,3/\sqrt{36} \right) = 4,84$$

$$LS = \bar{x} + z_{(1-\alpha/2)} S_x/\sqrt{n} = \bar{x} + z_{(0.95)} S_x/\sqrt{n} = 5,2 + 1,65 \left(1,3/\sqrt{36} \right) = 5,56$$

El intervalo buscado es $[\ 4,84;\ 5,56\ ]$. Se concluye que se tiene un 90% de confianza de que el valor promedio de pH de la solución se encuentra incluído dentro de éste intervalo.

5.3.1.3 Caso 3: Muestreo a partir de una población distribuida normalmente, con varianza desconocida y tamaño de muestra pequeño ($n < 30$ **).**

Una nueva situación se presenta si de una población que se distribuye normalmente con varianza desconocida se toma una muestra pequeña ($n < 30$). En éste caso S_x ya nos es un buen estimador de σ_x y el estadístico $(\bar{x} - \mu_{\bar{x}})/(S_x/\sqrt{n})$ no se distribuye normalmente. Afortunadamente, existe otro modelo que describe su distribución de probabilidades, conocido como distribución de T o de ¨Student¨. En éste caso se dice que la variable $(\bar{x} - \mu_{\bar{x}})/(S_x/\sqrt{n})$ se distribuye como T con *n-1* grados de libertad. El intervalo de confianza vendrá dado por la expresión siguiente:

$$\left[\ \bar{x} \pm\ t_{(1-\alpha/2;\ n-1)}\, S_x \big/ \sqrt{n}\ \right]$$

Donde: $t_{(1-\alpha/2;\ n-1)}$ es el valor de T a la izquierda del cual se encuentra el $(1-\alpha/2)\ 100\%$ de los valores de T.

5.3.1.3.1 Distribución de T

La variable aleatoria T tiene la función de probabilidad siguiente:

$$f(t) = \frac{\int_0^\infty y^{\left((\nu-1)/2\right)} e^{-y}\, dy}{\sqrt{\pi\nu}\int_0^\infty y^{\left(\frac{\nu^2}{2}-1\right)} e^{-y}\, dy} \left[1+\left(t^2/\nu\right)\right]^{-\left((\nu+1)/2\right)}$$

Donde ν es un parámetro de la distribución, conocido con el nombre de grados de libertad y se obtiene a partir del tamaño de la muestra menos uno (n-1). La función de probabilidad f(t), se caracteriza por lo siguiente: 1) la variable T toma valores entre $-\infty\ y\ +\infty$ ($-\infty \leq T \leq +\infty$); 2) los valores de T se distribuyen simétricamente alrededor de la media $\mu = 0$. Su forma es parecida a la distribución normal pero más prominente y con las colas más levantadas. En la medida que ν se hace más grande la forma de la distribución de T se asemeja más a la distribución de Z, y 3) Para cada valor de ν existe una distribución de T.

La función acumulada de la variable T se ha tabulado de una manera diferente a la tabla de Z. Las tablas de la distribución acumulativa de T tienen dos entradas: i) los grados de libertad (n-1) y ii) 1-α = la probabilidad de tener un valor menor a t.

Tabla de porcentuales de la Distribución de Student (T)
(ν = grados de libertad; ***1-α*** = proporción del área a la izquierda de t)

ν \ 1-α	$t_{(0,90)}$	$t_{(0,95)}$	$t_{(0,975)}$	$t_{(0,99)}$	$t_{(0,995)}$	$t_{(0,9975)}$	$t_{(0,999)}$	$t_{(0,9995)}$
1	3,078	6,314	12,706	31,821	63,656	127,321	318,289	636,578
2	1,886	2,920	4,303	6,965	9,925	14,089	22,328	31,600
3	1,638	2,353	3,182	4,541	5,841	7,453	10,214	12,924
4	1,533	2,132	2,776	3,747	4,604	5,598	7,173	8,610
5	1,476	2,015	2,571	3,365	4,032	4,773	5,894	6,869
6	1,440	1,943	2,447	3,143	3,707	4,317	5,208	5,959
7	1,415	1,895	2,365	2,998	3,499	4,029	4,785	5,408

El cuerpo de la tabla lo conforman los valores de T a la izquierda del cual se encuentra una proporción 1- α del área. También se puede decir que 1-α es la probabilidad de encontrar un valor de la variable T a la izquierda del t tabulado. Cualquier valor de t se identifica de la manera siguiente: $t_{(1-\alpha;\ n-1)}$. Por ejemplo $t_{(0975;\ 6)} = 2{,}447$ es el valor de t a la izquierda del cual se encuentra una proporción del área igual a 0,975 con 6 grados de libertad, o de otra manera: existe una probabilidad igual a 0,975 de encontrar un valor igual o menor a $t = 2{,}447$ para 6 grados de libertad.

Ejemplo 5.3
Se capturaron 25 murciélagos en una selva nublada y se encontró que esta muestra proporcionó un peso promedio de 100 g y una varianza de 400 g². Si se sabe que la variable peso se distribuye normalmente, estime el peso promedio de la población con la seguridad de no equivocarse en más de un 10% de las veces.

Se desea estimar μ_X con una probabilidad del 90%, a partir de una muestra pequeña (n < 30) que proviene de una población distribuida normalmente pero con varianza desconocida. En éste caso el estadístico $(\bar{x} - \mu_{\bar{x}}) / (S_x / \sqrt{n})$ sigue la distribución de T, y el intervalo de confianza es de la forma siguiente:

$$\left[\bar{x} \pm t_{[1-(\alpha/2);\ n-1]}\, S_x / \sqrt{n} \right]$$

El valor de sus límites son:

$$Li = \bar{x} - t_{(1-\alpha/2;\ n-1)} \frac{S_x}{\sqrt{n}} = \bar{x} - t_{(0,95;\ 24)} \frac{S_x}{\sqrt{n}} = 100 - 1{,}714 \left(\frac{20}{\sqrt{25}} \right) = 93{,}144$$

$$Li = \bar{x} + t_{(1-\alpha/2;\ n-1)} \frac{S_x}{\sqrt{n}} = \bar{x} + t_{(0,95;\ 24)} \frac{S_x}{\sqrt{n}} = 100 + 1{,}714 \left(\frac{20}{\sqrt{25}} \right) = 106{,}856$$

El intervalo buscado es *[93,14 ; 106,86]*. Se concluye que se tiene un 90% de confianza que el intervalo incluya el valor del peso promedio de la población de murciélagos.

5.3.1.4 Caso 4: Muestreo a partir de una población con distribución desconocida, varianza conocida y tamaño de muestra grande ($n \geq 30$).
Cuando se desconoce la forma de la distribución de valores de una variable no se puede predecir como será la distribución de la media muestral, a menos que el tamaño de la muestra sea grande. Si este es el caso, es decir si $n \geq 30$, entonces es aplicable el Teorema del Límite Central y la variable $\bar{X}$ tenderá a distribuirse normalmente con media $\mu_{\bar{X}} = \mu_X$ y varianza $\sigma^2_{\bar{X}} = \sigma^2_X / n$, de modo que el intervalo de confianza será de la forma [$\bar{x} \pm z_{(1-\alpha/2)}\, \sigma_X / \sqrt{n}$]

Ejemplo 5.4
Con el propósito de conocer el valor promedio de la concentración de cierto metabolito en la sangre de una población determinada, se analizó la sangre de 30 adultos. La concentración promedio del metabolito en esta muestra fue de 92 μg/l. Estudios anteriores habían

determinado para la misma población una varianza de 100 μg/l. Usando ésta información construya un intervalo de confianza del 99% para la media poblacional del metabolito.

En este problema aunque se conoce la varianza de la población se desconoce la forma como esta distribuida la concentración del metabolito. Como se tiene una muestra grande, se puede aplicar el Teorema del Límite Central y el intervalo a construir será de la forma [$\bar{x} \pm z_{(1-\alpha/2)}\sigma_x/\sqrt{n}$], por lo tanto sus límites serán:

$$LI = \bar{x} - z_{(1-\alpha/2)} \sigma_x/\sqrt{n} = \bar{x} - z_{(0,995)} \sigma_x/\sqrt{n} = 92 - 2,58 \left(10/\sqrt{30}\right) = 87,29$$

$$LS = \bar{x} + z_{(1-\alpha/2)} \sigma_x/\sqrt{n} = \bar{x} + z_{(0,995)} \sigma_x/\sqrt{n} = 92 + 2,58 \left(10/\sqrt{30}\right) = 96,71$$

El intervalo buscado es *[87,29; 96,71]*. Se concluye que se tiene un 99% de confianza que el valor promedio de la concentración del metabolito en la sangre de la población estudiada esté incluido en ese intervalo.

5.3.1.5 Caso 5: Muestreo a partir de una población con distribución y varianza desconocidas y tamaño de muestra grande ($n \geq 30$ **).**

Como en el caso anterior al ser $n \geq 30$, es aplicable el Teorema del Límite Central por lo que la media muestral se distribuye normalmente. La varianza de la muestra S_X^2 se usa como estimador de σ_X^2 y el intervalo de confianza será de la forma: [$\bar{x} \pm z_{(1-\alpha/2)} S_X/\sqrt{n}$]

Ejemplo 5.5

Los niveles de glucosa en la sangre de 40 estudiantes de nuevo ingreso en la Facultad de Ciencias dieron un valor promedio de 4,05 mmol/l y una desviación igual a 0,3 mmol/l. Construya un intervalo de confianza para la media poblacional. Use un $\alpha = 0,08$.

El intervalo a construir tiene la forma $\bar{x} \pm z_{(1-\alpha/2)} S_X/\sqrt{n}$. Si $\alpha = 0,08$, entonces $\alpha/2 = 0,04$ y consecuentemente $1-\alpha/2 = 1-0,04 = 0,96$, así que los límites del intervalos serán los siguientes:

$$LI = \bar{x} - z_{(1-\alpha/2)} S_X/\sqrt{n} = \bar{x} - z_{(0,96)} S_X/\sqrt{n} = 4,05 - 1,75 \left(0,3/\sqrt{40}\right) = 3,96$$

$$LS = \bar{x} + z_{(1-\alpha/2)} S_X/\sqrt{n} = \bar{x} + z_{(0,96)} S_X/\sqrt{n} = 4,05 + 1,75 \left(0,3/\sqrt{40}\right) = 4,13$$

El intervalo buscado es *[3,96 ; 4,13]*. Se concluye que se tiene un 92% de confianza que el valor promedio de los niveles de glucosa en la sangre de la población de estudiantes de nuevo ingreso esté incluido en ese intervalo.

5.3.1.6 Caso 6: Muestreo a partir de una población con distribución desconocida y tamaño de muestra pequeño ($n \geq 30$).

Cuando no se conoce la distribución de la variable y el tamaño de la muestra es pequeño $(n < 30)$, no es posible predecir la distribución que asume la media muestral. Por lo tanto no se puede construir un intervalo desconfianza, a menos que los datos sean transformados y se logren aproximar a una distribución normal.

En la Figura 5.7 se presentan un esquema con la combinación de los diferentes aspectos que determinan la construcción de un intervalo de confianza

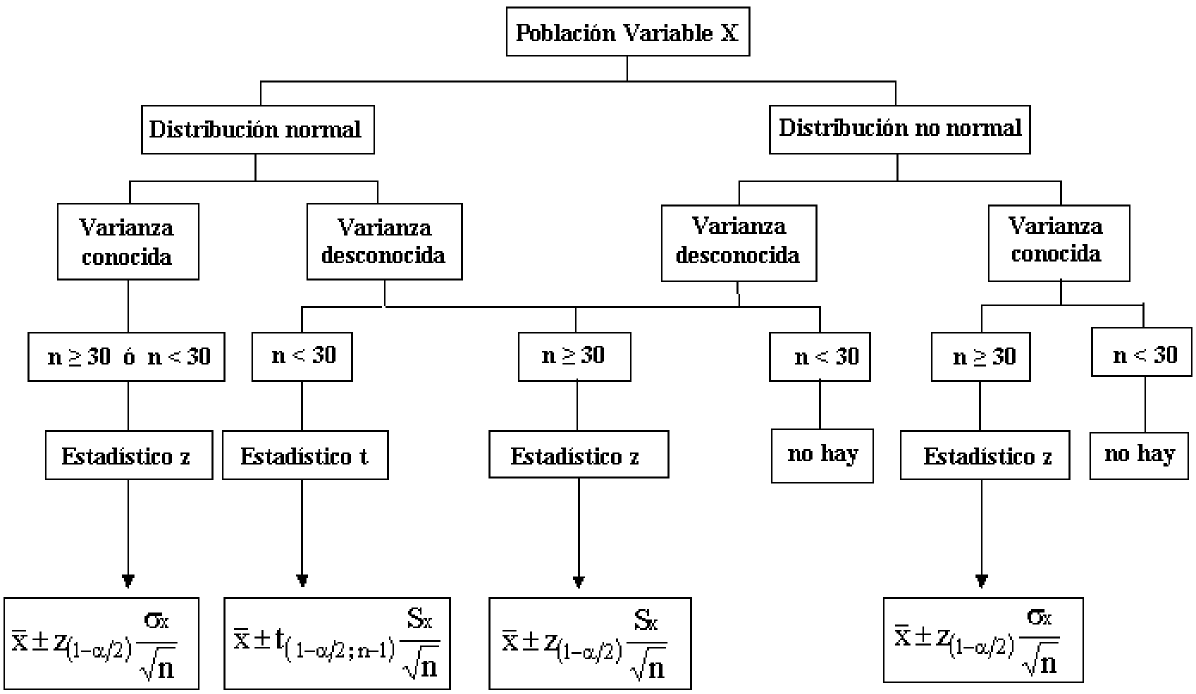


Figura 5.7

5.3.2 Intervalo de confianza para la diferencia de dos medias poblacionales

Igual que en la estimación de una media poblacional en la construcción de un intervalo para una diferencia de medias poblacionales es necesario considerar el tipo de distribución de la variable, el conocimiento de las varianzas poblacionales y el tamaño de las muestras.

5.3.2.1 Caso 1: Muestreo a partir de poblaciones distribuidas normalmente y con varianzas conocidas.

Recordemos que cuando se hace un muestreo de dos poblaciones distribuidas normalmente, se puede generar una nueva variable conocida como diferencia de medias muestrales, cuya distribución de valores se caracteriza por tener también una distribución normal, siendo su media y varianza las siguientes:

$$\mu_{(\bar{x}_2-\bar{x}_1)} = \mu_{x_2} - \mu_{x_1} \qquad \sigma^2_{(\bar{x}_2-\bar{x}_1)} = \frac{\sigma^2_{x_1}}{n_1} + \frac{\sigma^2_{x_2}}{n_2}$$

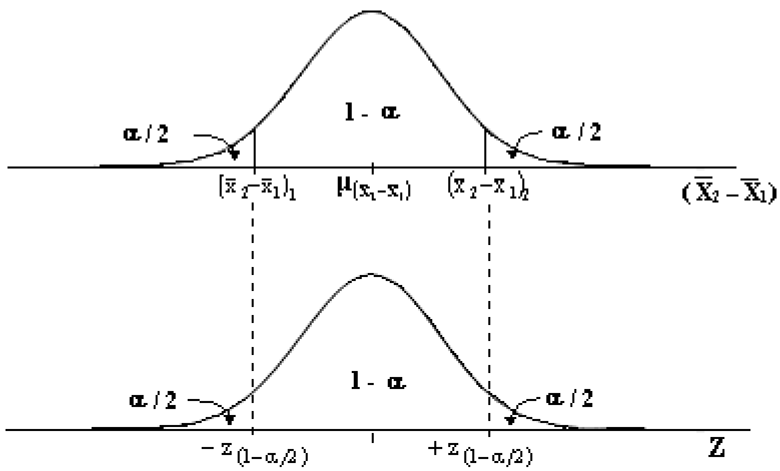


Figura 5.8. Relación entre la distribuciones de la variable $\bar{X}_2 - \bar{X}_1$ y la variable Z**.** La expresión $1-\alpha$ representa la proporción de la población incluída entre dos valores de $\bar{X}_2 - \bar{X}_1$ o de Z**.**

La deducción del intervalo de confianza para la diferencia de media poblacionales se puede comenzar estableciendo que la probabilidad de que la variable $(\bar{X}_2 - \bar{X}_1)$ se encuentre entre dos valores cualquiera es igual a $1-\alpha$ (Figura 5.8)

$$P\left\{(\bar{x}_2-\bar{x}_1)_1 \le \bar{X}_2 - \bar{X}_1 \le (\bar{x}_2-\bar{x}_1)_2\right\} = 1-\alpha$$

Esta es la misma probabilidad de que la variable Z se encuentre entre dos valores:

$$P\left\{z_1 \le Z \le z_2\right\} = 1\text{-}\alpha$$

Sabiendo que $Z = \dfrac{(\bar{x}_2-\bar{x}_1)-\left(\mu_{\bar{x}_2-\bar{x}_1}\right)}{\sigma_{\bar{x}_2-\bar{x}_1}}$

$$P\left\{z_1 \le Z \le z_2\right\} = P\left\{-z_{(1-\alpha/2)} \le \frac{(\bar{x}_2-\bar{x}_1)-\left(\mu_{\bar{x}_2-\bar{x}_1}\right)}{\sigma_{\bar{x}_2-\bar{x}_1}} \le +z_{(1-\alpha/2)}\right\} = 1-\alpha$$

El desarrollo algebraico de la ecuación anterior conduce a la expresión siguiente:

$$P\left\{(\bar{x}_2-\bar{x}_1) - z_{(1-\alpha/2)}\sigma_{\bar{x}_2-\bar{x}_1} \le \mu_{\bar{x}_2-\bar{x}_1} \le (\bar{x}_2-\bar{x}_1) + z_{(1-\alpha/2)}\sigma_{\bar{x}_2-\bar{x}_1}\right\} = 1-\alpha$$

Sustituyendo en la expresión anterior $\sigma_{(\overline{x}_2-\overline{x}_1)}$ por $\sqrt{\frac{\sigma^2_{x_1}}{n_1}+\frac{\sigma^2_{x_2}}{n_2}}$ se tiene,

$$P\left\{(\overline{x}_2-\overline{x}_1)-z_{(1-\alpha/2)}\sqrt{\frac{\sigma^2_{x_2}}{n_2}+\frac{\sigma^2_{x_1}}{n_1}} \leq \mu_{\overline{x}_2-\overline{x}_1} \leq (\overline{x}_2-\overline{x}_1)+z_{(1-\alpha/2)}\sqrt{\frac{\sigma^2_{x_2}}{n_2}+\frac{\sigma^2_{x_1}}{n_1}}\right\}=1-\alpha$$

De manera que el intervalo de confianza para estimar la diferencia entre dos medias poblacionales tiene la forma general,

$$\left[(\overline{x}_2-\overline{x}_1)\pm z_{(1-\alpha/2)}\sqrt{\frac{\sigma^2_{x_2}}{n_2}+\frac{\sigma^2_{x_1}}{n_1}}\right]$$

Ejemplo 5.6
En un trabajo de investigación se encontró que el contenido promedio de ácido úrico en 12 niños con el Síndrome de Down fue de 4,75 mg/100ml, mientras que en 18 niños normales el valor promedio encontrado fue de 3,95 mg/100 ml.. Mediante trabajos previos se había determinado que las varianzas de ambos grupos son 1,02 y 0,98 respectivamente. Suponiendo que la concentración de ácido úrico es una variable que se distribuye normalmente construya un intervalo de confianza del 98% para la diferencia de medias poblacionales.

Si las muestras provienen de poblaciones distribuidas normalmente y con varianza conocida, y el nivel de confianza $1-\alpha=0,98$, el intervalo de confianza tiene la forma siguiente:

$$\left[(\overline{x}_2-\overline{x}_1)\pm z_{(1-\alpha/2)}\sqrt{\left(\sigma^2_{x_2}/n_2\right)+\left(\sigma^2_{x_1}/n_1\right)}\right]$$

Sus límites son los siguientes:

$$LI=\left[(\overline{x}_2-\overline{x}_1)-z_{(0,99)}\sqrt{\left(\sigma^2_{x_2}/n_2\right)+\left(\sigma^2_{x_1}/n_1\right)}\right]=\left[0,80-2,33\sqrt{(1,02/12)+(0,98/18)}=0,1099\right]$$

$$LS=\left[(\overline{x}_2-\overline{x}_1)+z_{(0,99)}\sqrt{\left(\sigma^2_{x_2}/n_2\right)+\left(\sigma^2_{x_1}/n_1\right)}\right]=\left[0,80+2,33\sqrt{(1,02/12)+(0,98/18)}=1,8501\right]$$

El intervalo buscado es *[0,1099 ; 1,85501]*. Se concluye que se tiene un 98% de confianza que el valor de la diferencia de medias poblacionales sea un punto dentro de ese intervalo.

5.3.2.2 Otros casos

Los otros tipos de intervalos de confianza para la diferencia de medias poblacionales que resultan de la combinación de varias situaciones se muestran en el esquema de la Figura 5.9.

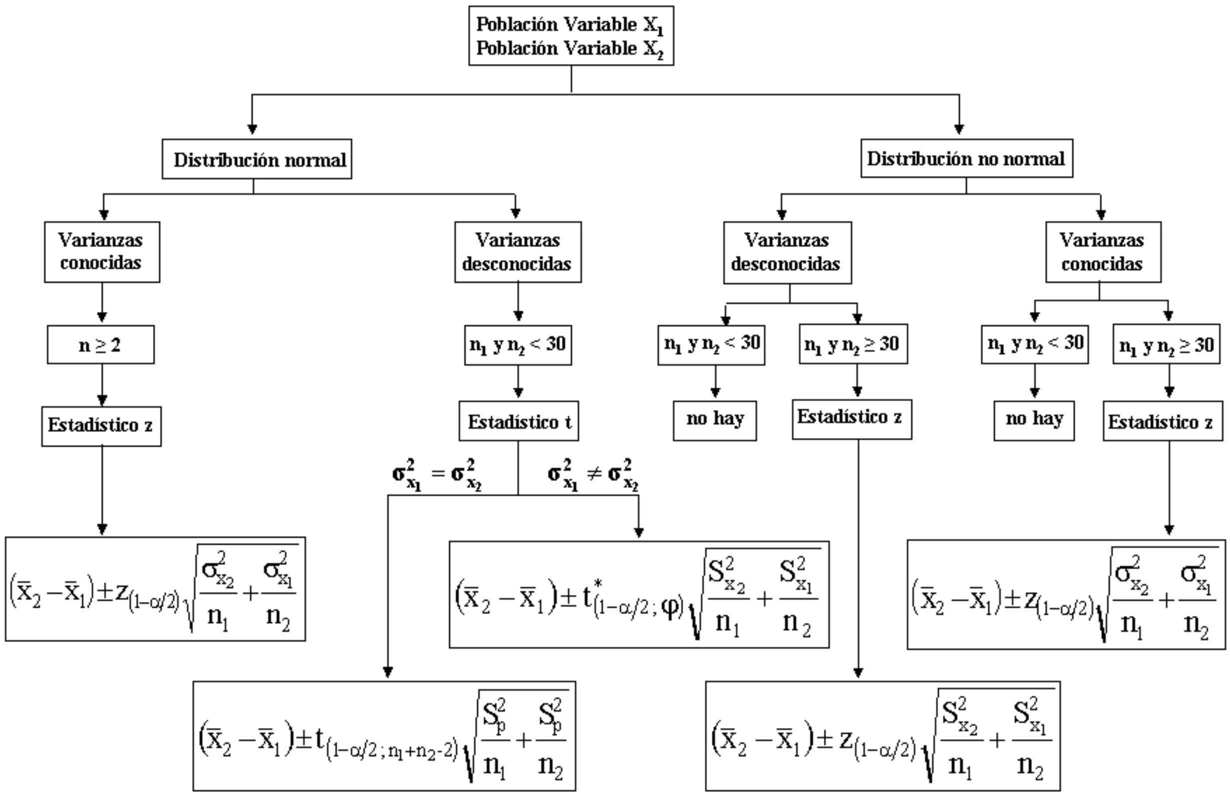


Figura 5.9

5.3.2.3 Intervalos de confianza para $\mu_{x_2} - \mu_{x_1}$ y el estadístico t.

En el esquema de la figura anterior se puede ver que el uso del estadístico t está condicionado, por la suposición de que dichas varianzas sean iguales o diferentes. De modo que la primera tarea antes de construir un intervalo de confianza es determinar cual de las dos situaciones siguientes se tiene:

i) σ_1^2 y σ_2^2 iguales y desconocidas.

ii) σ_1^2 y σ_2^2 diferentes y desconocidas.

Aunque más adelante se verá un procedimiento para comparar varianzas poblacionales y determinar si existen diferencias entre ellas, vamos a establecer una regla práctica que permitirá decidir rápidamente esta cuestión. Lo primero que se debe hacer es calcular la razón de varianzas RV, como el cociente de la varianza muestral mayor entre la varianza muestral menor:

$$RV = \frac{s_1^2}{s_2^2}\,,\ \text{siendo}\ s_1^2 > s_2^2$$

Luego se toma una decisión sobre la base de las reglas siguientes:

Si $\alpha = 0{,}10$ y RV > 2,0 las varianzas son diferentes.

Si $\alpha = 0{,}05$ y RV > 2,5 las varianzas son diferentes.

Si $\alpha = 0{,}001$ y RV > 3,5 las varianzas son diferentes.

5.3.2.3.1 Varianzas iguales

Cuando se acepta la suposición que las dos varianzas poblacionales aunque desconocidas son iguales, se pueden promediar las varianzas de las muestras para hacer una mejor estimación de la varianza poblacional. Para obtener el promedio, el valor de las varianzas muestrales debe ser ponderado por el tamaño de las muestras de acuerdo a la fórmula siguiente:

$$S_p^2 = \frac{(n_1 - 1)\ S_{x_1}^2 + (n_2 - 1)\ S_{x_2}^2}{n_1 + n_2 - 2}$$

De manera que la desviación de la distribución de medias muestrales $S_{(\bar{x}_2 - \bar{x}_1)}$ queda igual a:

$$\sqrt{\frac{S_p^2}{n_2} + \frac{S_p^2}{n_1}}$$

y el intervalo de confianza para $\mu_{x_2} - \mu_{x_1}$ se obtiene con la fórmula siguiente:

$$(\bar{x}_2 - \bar{x}_1) \pm t_{(1-\alpha/2\ ;\ n_1+n_2-2)} \sqrt{\frac{S_p^2}{n_2} + \frac{S_p^2}{n_1}}$$

Ejemplo 5.7

En una investigación sobre la calidad química del agua en un río, se determinó el fósforo total en solución en dos épocas distintas. En la primera fecha, 25 mediciones del elemento proporcionaron un valor promedio de 0,78 mg/l y una varianza de 0,063 mg/l. En la segunda fecha se efectuaron 16 mediciones y los valores encontrados para la media y la varianza fueron de 0,06 y 0,048 mg/l respectivamente. Suponiendo que la variable concentración de fósforo se distribuye normalmente construya un intervalo de confianza del 95% para la diferencia de medias poblacionales.

Las muestras de valores del contenido de fósforo son pequeñas (n_1 y $n_2 < 30$) y provienen de dos poblaciones diferentes que se distribuyen normalmente con varianzas desconocidas. Por lo tanto antes de construir el intervalo de confianza se debe comprobar si es posible suponer que las dos varianzas poblacionales son iguales, para lo cual se aplica una de las reglas prácticas para la comparación de varianzas.

Como α = 0,05 y $RV = \frac{s_1^2}{s_2^2} = \frac{0{,}063}{0{,}048} = 1{,}31$ es menor a 2 se acepta que las dos varianzas son iguales. Por lo tanto, de acuerdo al diagrama de la Figura 5.9 el intervalo a usar es el que se muestra a continuación:

$$(\bar{x}_2 - \bar{x}_1) \pm t_{(1-\alpha/2\ ;\ n_1+n_2-2)} \sqrt{\frac{S_p^2}{n_2} + \frac{S_p^2}{n_1}}$$

El primer paso es calcular la varianza ponderada

$$S_p^2 = \frac{(n_1-1)\ S_{x_1}^2 + (n_2-1)\ S_{x_2}^2}{n_1+n_2-2} = \frac{(24)(0,063)+(15)(0,048)}{39} = 0,0572$$

Luego se calculan los límites:

$$LS = (\bar{x}_2-\bar{x}_1) + t_{(0,975;\ 39)}\sqrt{\frac{S_p^2}{n_2}+\frac{S_p^2}{n_1}} = 0,15 + 2,3313\sqrt{\frac{0,0572}{25}+\frac{0,0572}{16}} = 0,3285$$

$$LI = (\bar{x}_2-\bar{x}_1) - t_{(0,975;\ 39)}\sqrt{\frac{S_p^2}{n_2}+\frac{S_p^2}{n_1}} = 0,15 - 2,3313\sqrt{\frac{0,0572}{25}+\frac{0,0572}{16}} = -\ 0,0285$$

El intervalo buscado es *[-0,0285; 0,3285]*. Se concluye que se tiene un 95% de confianza que éste intervalo contenga el valor de la diferencia de las medias poblacionales.

5.3.2.3.2 Varianzas diferentes

Si se asume que las varianzas de dos poblaciones de una variable que se distribuye normalmente, son diferentes aunque desconocidas, no se puede usar el estadístico:

$$\frac{(\bar{x}_2-\bar{x}_1)-(\mu_{x_2}-\mu_{x_1})}{\sqrt{\frac{S_{x_2}^2}{n_2}+\frac{S_{x_1}^2}{n_1}}}$$

Para calcular el coeficiente de confiabilidad $t_{(1-\alpha/2)}$ porque su distribución no sigue el modelo de la distribución de T. Sin embargo es posible calcular un nuevo coeficiente de confiabilidad $t^*_{(1-\alpha/2)}$, usando la fórmula siguiente:

$$t^*_{(1-\alpha/2)} = \frac{w_1\ t_{(1\text{-}\alpha/2;n_1-1)} + w_2\ t_{(1\text{-}\alpha/2;n_2-1)}}{w_1+w_2}$$

Donde $w_1 = \left(\frac{s_1^2}{n_1}\right)$ y $w_2 = \left(\frac{s_2^2}{n_2}\right)$

El intervalo de confianza se obtiene mediante la fórmula: $(\bar{x}_2-\bar{x}_1) \pm t^*_{(1-\alpha/2\ )}\sqrt{\frac{S_{x_2}^2}{n_2}+\frac{S_{x_1}^2}{n_1}}$

Ejemplo 5.8

Al comparar dos métodos para determinar la concentración de Boro en un material vegetal se efectuaron varias mediciones obteniéndose los resultados siguientes para los dos métodos.

	Concentración de Boro	
	Espectrofotometría	Fluorimetría
n	10	16
Media	26,00 μg/l	28,00 μg/l
Desviación	0,23 μg/l	1,30 μg/l

Construya un intervalo de confianza del 99% para $\mu_{x_2} - \mu_{x_1}$. Suponga que la variable concentración se distribuye normalmente.

Las condiciones del problema indican que las muestras son pequeñas y provienen de dos poblaciones que se distribuyen normalmente y con varianzas desconocidas. Al seguir el esquema de la Figura 5.9 para escoger el intervalo adecuado, es necesario decidir si las desconocidas varianzas poblacionales son iguales o diferentes.

Como α = 0,01 y $RV = \frac{s_1^2}{s_2^2} = \frac{1,3}{0,23} = 5,65$ es mayor a 3,5 se acepta que las dos varianzas son diferentes. Por lo tanto de acuerdo al diagrama de la Figura 5.9 el intervalo a construir debe ser el siguiente:

$$(\bar{x}_2 - \bar{x}_1) \pm t^*_{(1-\alpha/2)} \sqrt{\frac{S^2_{x_2}}{n_2} + \frac{S^2_{x_1}}{n_1}}$$

El primer paso es calcular el coeficiente de confiabilidad $t^*_{(1-\alpha/2)}$, usando la fórmula siguiente:

$$t^*_{(1-\alpha/2)} = \frac{w_1\ t_{(1\text{-}\alpha/2;n_1-1)} + w_2\ t_{(1\text{-}\alpha/2;n_2-1)}}{w_1 + w_2},$$

Se encuentran los valores para cada término de la fórmula anterior.

$$t_{(1-\alpha/2\ ;\ n_1-1)} = t_{(0,995;\ 9)} = 3,25 \qquad \text{y} \qquad t_{(1-\alpha/2\ ;\ n_2-1)} = t_{(0,995;\ 15)} = 2,947$$

$$w_1 = \left(\frac{S^2_{x_1}}{n_1}\right) = \frac{(0,23)^2}{10} = 0,0053 \qquad w_2 = \left(\frac{S^2_{x_2}}{n_2}\right) = \frac{(1,30)^2}{16} = 0,1056$$

El valor del coeficiente de confiabilidad será:

$$t^*_{(1-\alpha/2)} = \frac{w_1\ t_{(1\text{-}\alpha/2;n_1-1)} + w_2\ t_{(1\text{-}\alpha/2;n_2-1)}}{w_1 + w_2} = \frac{(0,0053)\ 3,25 + (0,1056)\ 2,947}{0,0053 + 0,1056} = 2,96$$

Se calculan los límites de confianza:

$$LI = (\bar{x}_2 - \bar{x}_1) - t^*_{(1-\alpha/2)} \sqrt{\frac{S^2_{x_2}}{n_2} + \frac{S^2_{x_1}}{n_1}} = 2 - 2,96\sqrt{\frac{1,30^2}{16} + \frac{0,23^2}{10}} = 2 - 0,9858 = 1,0142$$

$$LS = (\overline{x}_2 - \overline{x}_1) + t^*_{(1-\alpha/2\ )}\sqrt{\frac{S^2_{x_2}}{n_2} + \frac{S^2_{x_1}}{n_1}} = 2 + 2{,}96\sqrt{\frac{1{,}30^2}{16} + \frac{0{,}23^2}{10}} = 2 + \ 0{,}9858 = 2{,}9858$$

El intervalo buscado es *[1,0142; 2,9858]*. Se concluye que se tiene un 99% de confianza que el intervalo anterior incluya el valor de $\mu_{x_2} - \mu_{x_1}$.

5.4 EJERCICIOS

1. La concentración de oxígeno disuelto en el agua de un río es una variable aleatoria que se distribuye normalmente con varianza igual a 0,5 ppm. Si 25 mediciones del oxígeno con el método de Winkler produjeron un valor promedio igual 8,2 ppm, determine:
   a. Los límites de confianza para el verdadero valor promedio del oxígeno disuelto con una confianza del 98%.
   b. Estime el error máximo
   c. Determine el tamaño que debe tener la muestra para cometer el error máximo un 95% de las veces.
   d. Determine el tamaño que debe tener la muestra para cometer el error máximo un 99% de las veces.

2. El peso de las partículas sólidas en el aire es una variable aleatoria que se distribuye normalmente y su valor promedio se usa como indicador de la contaminación atmosférica. Si en una investigación de la calidad del aire se determinó en 30 ocasiones que el peso promedio de las partículas suspendidas fue de 75 mg/m^3 con una desviación igual a 12 mg/m^3, construya con un 95% de confianza un intervalo que contenga el verdadero peso promedio de las partículas suspendidas en el aire.

3. Ciertas especies de plantas para poder producir flores requieren permanecer diariamente en oscuridad un número mínimo de horas en forma ininterrumpida. Para una determinada especie se determinó que hubo floración cuando los lapsos de oscuridad medidos en horas presentaron los valores siguientes:

   16,0; 14,0; 14,8; 13,9; 15,7; 16,2; 13,2; 15,3, 14,7

4. Bajo el supuesto que el tiempo de oscuridad requerido para la floración es una variable aleatoria que se distribuye normalmente, estime el tiempo medio de oscuridad ininterrumpida que requieren los individuos de ésta especie de planta para producir flores. Escoja el nivel de confianza que le parezca más adecuado.

5. Se sabe que para cierta comunidad el consumo de calorías en varones de 20 años con pesos que varían entre los 70 y 75 Kg tiene una varianza igual a 0,7225 Kcal/día. Determine mediante un intervalo de confianza del 90% el consumo medio de calorías de la población de jóvenes con las características antes mencionadas, si la medición del consumo de calorías en 42 de ellos produjo un valor medio de 2,9 Kcal/día. ¿A cuántos

jóvenes se les debería medir el consumo diario de calorías para que la amplitud del intervalo no fuese mayor a 0,5 Kcal?.

6. En una práctica de laboratorio 36 estudiantes midieron el nivel de cadmio en un tipo de aceite vegetal usando la misma técnica. Los resultados encontrados para el total de ensayos muestran que el valor promedio de cadmio fue de 7,12 ppm y la desviación 2,24 ppm. Estime mediante un intervalo de confianza el valor promedio del cadmio para todo el volumen de aceite vegetal de donde se extrajeron las porciones que fueron evaluadas. Use un $\alpha = 0{,}04$.

7. Un taxónomo midió la longitud promedio del cuerpo de dos razas de ratones de la misma especie que viven en una sabana y encontró que 10 individuos de la raza A tienen una longitud promedio de 16,25 cm y que en 14 ejemplares de la raza B el promedio de longitud es de 15,40 cm. Si se sabe que la variable se distribuye normalmente con una desviación igual 1,0 cm para la raza A y de 1,16 cm para la raza B, estime mediante un intervalo de confianza la verdadera diferencia entre las medias poblacionales con un nivel de confianza del 80%..

8. Un investigador seleccionó dos muestras de 36 granos de semillas de dos variedades. El análisis del contenido de proteínas en la variedad A dio como resultados un valor promedio igual a 37,75% con una desviación de 4,71%. En la variedad B el valor de los mismos estadísticos fue de 35% y 3,89% respectivamente. Estime el valor de la diferencia promedio del contenido de proteínas para las dos variedades de semillas, con un $\alpha = 0{,}02$.

9. En un estudio sobre la calidad de las aguas de un río se calculó la diversidad de macroinvertebrados acuáticos en dos sitios ubicados antes y después de la salida de una fuente de contaminación. Para 12 muestras de animales en la estación río arriba el valor de la diversidad promedio fue de 3,11 especies con una desviación igual a 0,771 especies, mientras que en 10 muestras de animales recogidas rió abajo, el valor promedio de diversidad fue de 2,04 con una desviación de 0,448 especies. Encuentre un intervalo de confianza del 90% para la diferencia entre las medias poblacionales, asumiendo que la distribución de la variable en las dos poblaciones es normal.

10. En los mismos sitios de muestreo del problema anterior se determinó en 10 fechas el contenido de fósforo disuelto en el agua. Los resultados encontrados fueron los siguientes:

	Concentración de fósforo (mg/L)									
Aguas arriba	1,80	1,01	1,15	1,70	1,45	1,25	1,42	1,54	1,62	2,01
Aguas abajo	5,01	3,54	4,84	3,84	2,85	3,54	2,65	3,15	4,12	5,01

11. Encuentre un intervalo de confianza del 95% para la diferencia en los contenidos promedios reales de fósforo en las dos estaciones, asumiendo que las muestras provienen de poblaciones de valores que se distribuye normalmente.

12. En un proceso químico se comparan dos catalizadores para verificar su efecto en una reacción. Se efectuaron 32 reacciones utilizando el catalizador I y 32 reacciones usando el catalizador II. En el primer caso el rendimiento promedio fue de 85% con una desviación de 6%. Para el segundo grupo de reacciones el rendimiento fue de 81% y la desviación fue de 7%. Construya un intervalo de confianza del 90% para la verdadera diferencia promedio en el rendimiento.

13. En un ensayo sobre el efecto del nitrógeno en el crecimiento de una especie de árbol se plantaron dos lotes de 1000 árboles cada uno. Uno de los lotes fue fertilizado con Na NO_3 y el otro no fue tratado. Después de cinco meses se pesaron los tallos de 50 árboles en cada lote. Los árboles no tratados produjeron un valor promedio del tallo igual a 3,93 kg y una desviación de 0,78 kg. En los árboles fertilizados el promedio fue de 4,87 kg y una desviación de 2,52 kg. Calcule un intervalo de confianza del 99% para la diferencia promedio del peso del tronco entre los dos lotes.

14. En un estudio para determinar la capacidad de absorción de bromuro por los tejidos vegetales se le añadió a dos tipos de vegetales, tomate y pepino, la misma cantidad de bromuro y luego se procedió a medir mediante cromatografía de gases la cantidad de bromuro recuperado. Si la concentración de bromuro recuperado en el tomate es una variable aleatoria que se distribuye normalmente con una desviación igual a 13,6 μg/g, mientras que en el pepino la misma variable se distribuye normalmente con una varianza igual a 10,4 μg/g. Estime mediante un intervalo de confianza del 92% la diferencia que existe entre los promedios poblacionales de recuperación de bromuro para los dos vegetales, si siete mediciones del bromuro recuperado en el tomatte produjeron un media igual 772,6 μg/g y las misma cantidad de mediciones en el pepino produjeron una media igual a 780,6 μg/g.

15. Un investigador seleccionó dos muestras de 36 granos de semillas de dos variedades. El análisis del contenido de proteínas en la variedad A dio como resultados un valor promedio igual a 37,75% con una desviación de 4,71%. En la variedad B el valor de los mismos estadísticos fue de 35% y 3,89% respectivamente. Estime el valor de la diferencia promedio del contenido de proteínas para las dos variedades de semillas, con un $\alpha = 0{,}02$.

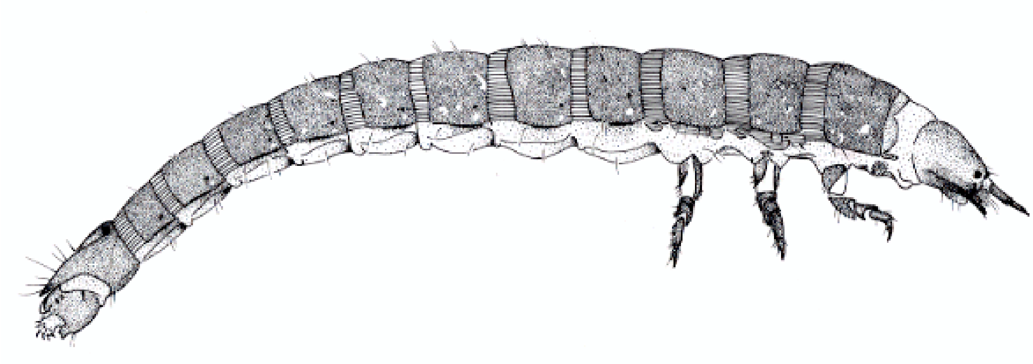
16. Para evaluar en método espectrofotométrico con el fin de determinar titanio, se aplicó el método a dos aleaciones conteniendo diferentes cantidades certificadas de titanio. Para los dos tipos de aleaciones se hicieron 8 mediciones, encontrándose que una produjo un valor medio de 0,482% de titanio con una desviación igual a 0,0257%, mientras que en la segunda aleación el promedio fue de 2,002% con una media de 0,0287%. Si la concentración de titanio se distribuye normalmente, estime mediante un intervalo de confianza del 95% la diferencia real que existe entre el contenido de titanio en las dos aleaciones.

17. En un estudio sobre la calidad de las aguas de un río se se determinó el contenido de fósforo disuelto en el agua. en dos sitios ubicados antes y después de la salida de una fuente de contaminación. En la estación aguas arriba se determinó el contenido de esta

sustancia en 15 fechas diferentes, encontrándose un valor promedio de 1,49 mg/l y una desviación de 0,8 mg/l. En la estación aguas abajo el valor promedio fósforo en 12 fechas fue de 3,84 mg/l con una desviación de 3,07 mg/l. Encuentre un intervalo de confianza del 95% para la diferencia en los contenidos promedios reales de fósforo en las dos estaciones, asumiendo que las muestras provienen de poblaciones de valores que se distribuye normalmente con varianzas distintas.

18. En un proceso químico se comparan dos catalizadores para verificar su efecto en una reacción. Se efectuaron 32 reacciones utilizando el catalizador I y 32 reacciones usando el catalizador II. En el primer caso el rendimiento promedio fue de 85% con una desviación de 6%. Para el segundo grupo de reacciones el rendimiento fue de 81% y la desviación fue de 7%. Construya un intervalo de confianza del 90% para la verdadera diferencia promedio en el rendimiento.

19. En un ensayo sobre el efecto del nitrógeno en el crecimiento de una especie de árbol se plantaron dos lotes de 1000 árboles cada uno. Uno de los lotes fue fertilizado con Na NO_3 y el otro no fue tratado. Después de cinco meses se pesaron los tallos de 50 árboles en cada lote. Los árboles no tratados produjeron un valor promedio del tallo igual a 3,93 kg y una desviación de 0,78 kg. En los árboles fertilizados el promedio fue de 4,87 kg y una desviación de 2,52 kg. Calcule un intervalo de confianza del 99% para la diferencia promedio del peso del tronco entre los dos lotes.



Larva de la Familia Elmidae (Insecta: Coleoptera)