$$T' = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_{X_1} - \mu_{X_2})}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \qquad \nu = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2 - 1}}$$

Entonces:

$$P(-t_{\alpha/2} < T' < t_{\alpha/2}) = P\left(-t_{\alpha/2} < \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_{X_1} - \mu_{X_2})}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} < t_{\alpha/2}\right) = 1 - \alpha$$

$t_{\alpha/2}$ es el valor de t con ν grados de libertad con un área igual a $\alpha/2$ a su derecha.

Muestras aleatorias pequeñas tomadas independientemente de las dos poblaciones aproximadamente normales, proporcionarán las estimaciones puntuales:

$\bar{x}_1 - \bar{x}_2$ del estadístico $\overline{X}_1 - \overline{X}_2$

s_1^2 del estadístico S_1^2 $\qquad$ s_2^2 del estadístico S_2^2

Estas estimaciones permiten obtener los grados de libertad ν para la DM(T'), donde ν se redondea al entero más próximo y construir el intervalo de confianza.

Así, se obtiene el siguiente intervalo de confianza del $(1-\alpha)100\%$ para la diferencia de medias $\mu_{X_1} - \mu_{X_2}$ en dos poblaciones normales con varianzas desconocidas y distintas $\sigma_{X_1}^2 \neq \sigma_{X_2}^2$

$$(\bar{x}_1 - \bar{x}_2) - t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} < \mu_{X_1} - \mu_{X_2} < (\bar{x}_1 - \bar{x}_2) + t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

IC(7)

El error estándar estimado de la estimación puntual es:

$$\widehat{se}(\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Ejemplo 4

En un estudio para estimar la diferencia en la cantidad de ortofósforo químico, medido en dos estaciones diferentes en el cauce de un río; se tomaron 15 muestras de la Estación 1 y 12 de la Estación 2. Las 15 primeras tuvieron un contenido de ortofósforo de 3,84 mg/lt y una desviación estándar de 3,07 mg / lt. Para las 12 de la Estación 2 se obtuvieron una media de 1,49 mg / lt y una desviación estándar de 0,80 mg/lt. Encuentra el intervalo de confianza del 95% para la diferencia entre los contenidos de ortofósforo entre las dos estaciones. Asume que las observaciones surgen de poblaciones normales con varianzas distintas.

Empleamos es estadístico T' / DM(T') es aproximadamente t de Student con ν grados de libertad, donde ν se calcula empleando la fórmula dada.

La estimación puntual de la diferencia de medias en el contenido de ortofósforo, obtenida a partir de muestras de tamaño $n_1 = 15$ y $n_2 = 12$, tomadas en las Estaciones 1 y 2, respectivamente es: $\bar{x}_1 - \bar{x}_2 = 3{,}84 - 1{,}49 = 2{,}35$

Las desviaciones estándar $s_1 = 3{,}07$ y $s_2 = 0{,}80$ obtenidas de las muestras permiten obtener la estimación puntual de los grados de libertad de la $\mathrm{DM}(T')$:

$$\nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}} = \frac{\left(\frac{3{,}07^2}{15} + \frac{0{,}8^2}{12}\right)^2}{\frac{\left(\frac{3{,}07^2}{15}\right)^2}{14} + \frac{\left(\frac{0{,}8^2}{12}\right)^2}{11}} = 16{,}3 \cong 16$$

$(1 - \alpha)\,100\% = 95\ \% \rightarrow \textit{grado de confianza} \quad \alpha = 0{,}05 \rightarrow \frac{\alpha}{2} = 0{,}025$

$t_{0,025} = 2{,}120$ es el valor de t con $\nu = 16$ grados de libertad con un área igual a $0{,}025$ a su derecha.

El intervalo de confianza del 95 % para $\mu_{X_1} - \mu_{X_2}$ en dos poblaciones normales con varianzas desconocidas y distintas $\sigma_{X_1}^2 \neq \sigma_{X_2}^2$ es :

$$2{,}35 - 2{,}12\sqrt{\frac{3{,}07^2}{15} + \frac{0{,}8^2}{12}} < \mu_{X_1} - \mu_{X_2} < 2{,}35 + 2{,}12\sqrt{\frac{3{,}07^2}{15} + \frac{0{,}8^2}{12}}$$

$$0{,}6 < \mu_{X_1} - \mu_{X_2} < 4{,}10$$

Se tiene una confianza del 95 % que el intervalo (0,6 , 4,10) incluye la diferencia de los contenidos promedios de ortofósforo para las dos estaciones.

Observaciones pareadas

Consideremos los procedimientos de estimación para la diferencia de medias cuando *las muestras no son independientes* y las varianzas de las dos poblaciones no son necesariamente iguales. Este es el caso de las observaciones pareadas.

Las condiciones de las dos poblaciones no se asignan aleatoriamente a las unidades experimentales. Cada unidad experimental, constituida por un par de elementos homogéneos o por un elemento recibe ambas condiciones poblacionales, en consecuencia, cada unidad experimental tiene un par de observaciones, una para cada condición (población).

Por ejemplo, si se llevara a cabo una prueba sobre una nueva dieta con 15 sujetos, las condiciones a observar son el peso antes y después de la dieta y la unidad experimental es el sujeto sometido al tratamiento. Las variable aleatorias que participan de la prueba son X_1: peso de los individuos antes de someterse a la dieta y X_2 : peso de los individuos después de hacer dieta

Otro ejemplo de apareamiento consiste en la selección de n pares de objetos, teniendo cada elemento del par características similares. Entonces para cada par se selecciona un elemento al azar para obtener un valor de la variable X_1, mientras que el otro miembro proporciona el valor de X_2. Por ejemplo si cada par de elementos son estudiantes con igual coeficiente intelectual, edad, crianza, sexo, formación, etc. ; X_1 y X_2 podrían representar las calificaciones que obtienen dos individuos con características muy similares cuando se asigna al azar a uno de los estudiantes a una clase que usa el sistema de lectura convencional y al otro a una clase que usa materiales programados.

En ambos casos se puede definir la variable aleatoria diferencia $D = X_1 - X_2$
Para determinar si la dieta es eficaz, o si los materiales programados mejoran el aprendizaje de la lectura, se consideran las diferencias $d_1, d_2, \ldots, d_n$ en las observaciones pareadas. En el primer ejemplo son las diferencias para una muestra de n sujetos sometidos a la dieta y en el segundo, las diferencias en las calificaciones obtenidas por los pares de estudiantes con iguales características. Estas diferencias son los valores para una muestra de la población de diferencias $D_1, D_2, \ldots, D_n$

Las diferencias se asumen con distribución normal con media $\mu_D = \mu_1 - \mu_2$ y varianza σ_D^2 desconocida. Se estima por S_d^2 varianza muestral de las diferencias D. Para el valor promedio de la diferencias $\overline{D}$, la distribución muestral es t de Student con $\nu = n - 1$ grados de libertad

$$T = \frac{\overline{D} - \mu_D}{\frac{S_d}{\sqrt{n}}}$$

Emplearemos las DM (T) para determinar el intervalo de confianza de μ_D. Entonces:

$$P(-t_{\alpha/2} < T < t_{\alpha/2}) = 1 - \alpha$$

$t_{\alpha/2}$ es el valor de t con $\nu = n-1$ grados de libertad que deja un área de $\alpha/2$ a su derecha y

$-t_{\alpha/2}$ es el valor de t con $\nu = n-1$ grados de libertad que deja un área de $\alpha/2$ a su izquierda.

Resulta:

$$P\left(-t_{\alpha/2} < \frac{\overline{D} - \mu_D}{\frac{S}{\sqrt{n}}} < t_{\alpha/2}\right) = P\left(\overline{D} - t_{\alpha/2}\frac{S}{\sqrt{n}} < \mu_X < D + t_{\alpha/2}\frac{S}{\sqrt{n}}\right) = 1 - \alpha$$

Se obtiene una estimación puntual del estadístico S_d^2, mediante el cálculo de s_d^2 a partir de la muestra pareada. La estimación puntual de $\overline{D}$ es el promedio de las diferencias d_i obtenidas de n pares aleatorios de mediciones :

$$\bar{d} = \frac{\sum_{i=1}^{n} d_i}{n}$$

Con estos valores surge el siguiente intervalo de confianza de $(1 - \alpha)$ 100% para $\mu_D = \mu_1 - \mu_2$ para observaciones pareadas

$$\bar{d} - t_{\alpha/2}\frac{s_d}{\sqrt{n}} < \mu_D < \bar{d} + t_{\alpha/2}\frac{s_d}{\sqrt{n}} \qquad \text{IC(8)}$$

El error estándar estimado de la estimación puntual es: $\widehat{se}(\bar{d}) = \frac{s_d}{\sqrt{n}}$

Ejemplo 5

Los contenidos de elementos esenciales de tomates frescos y enlatados se determinaron mediante el método de espectro-fotometría de absorción atómica. El contenido de Cobre en tomates frescos, en comparación con el que los mismos tomates registraron después de ser enlatados, se muestra en la tabla dada a continuación. Encuentre el intervalo de confianza del 98% para la diferencia real *en el contenido promedio de Cobre en tomates frescos y enlatados, suponiendo que la distribución de las diferencias es normal.*

Par	*Tomates frescos*	*Tomates enlatados*	*Diferencia* d_i
1	0,066	0,085	0,019
2	0,079	0,088	0,009
3	0,069	0,091	0,022
4	0,076	0,096	0,020
5	0,071	0,093	0,022
6	0,087	0,095	0,008
7	0,071	0,079	0,008
8	0,073	0,078	0,005
9	0,067	0,065	-0,002
10	0.062	0,068	0,006

$\mu_D = \mu_1 - \mu_2$ $\rightarrow$ diferencia entre los contenidos promedio de Cobre entre tomates enlatados y frescos

Estimación puntual de $\overline{D}$: $\bar{d} = 0{,}0117$
La desviación estándar para las diferencias muestrales es:

$$s_d = \sqrt{\frac{\sum_{i=1}^{n} d_i^2 - \frac{\left(\sum_{i=1}^{n} d_i\right)^2}{n}}{n-1}} = 0{,}0084$$

$(1-\alpha)\,100\% = 98\,\% \rightarrow \textit{grado de confianza} \quad \alpha = 0{,}02 \rightarrow \frac{\alpha}{2} = 0{,}01$

$t_{0,01} = 2{,}821$ es el valor de t con $\nu = 9$ grados de libertad con un área igual a $0{,}01$ a su derecha.

El intervalo de confianza del 98 % para μ_D es:

$$0{,}0117 - 2{,}821\frac{0{,}0084}{\sqrt{10}} < \mu_D < 0{,}0117 + 2{,}821\frac{0{,}0084}{\sqrt{10}}$$

$$0{,}0042 < \mu_D < 0{,}0192$$

Estimación de una proporción

Un estimador puntual de la proporción p en un experimento binomial está dado por el estadístico $\hat{P} = \frac{X}{n}$ donde X es el número de éxitos en n intentos. La proporción muestral $\hat{p} = \frac{x}{n}$ se utiliza como una estimación puntual del parámetro p. Siempre que la proporción desconocida p no se acerque demasiado a cero o a uno, se puede establecer un intervalo de confianza para p considerando la $\mathrm{DM}(\hat{P})$.

$\hat{P}$ está distribuido aproximadamente en forma normal con media $\mu_{\hat{P}} = p$ y varianza

$$\sigma^2_{\hat{P}} = \frac{p\,q}{n} \quad ; \quad q = 1 - p$$

$$\hat{P} \approx N\left(p, \frac{p\,q}{n}\right) \quad y \quad Z = \frac{\bar{P} - p}{\sqrt{\frac{\hat{P}\,\hat{Q}}{n}}}$$

Donde Z tiene una distribución normal estándar. Entonces si $P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha$, con $z_{\alpha/2}$ como el valor de z por encima del cual se encuentra un área de $\alpha/2$, resulta

$$P\left(\bar{P} - z_{\alpha/2}\sqrt{\frac{\hat{P}\,\hat{Q}}{n}} < p < \bar{x} + z_{\alpha/2}\sqrt{\frac{\hat{P}\,\hat{Q}}{n}}\right) = 1 - \alpha$$

Cuando n es grande se emplean las estimaciones puntuales de p y q en la determinación de $\sigma^2_{\hat{P}}$; $\hat{p} = \frac{x}{n}$ y $\hat{q} = 1 - \hat{p}$ y se obtiene el intervalo de confianza de $(1 - \alpha)\,100\%$

$$\bar{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}\,\hat{q}}{n}} < p < \bar{x} + z_{\alpha/2}\sqrt{\frac{\hat{p}\,\hat{q}}{n}}$$

IC(9)

Para usar este intervalo se requiere que ; $n\,\hat{p} > 5$ y $n\,\hat{q} > 5$

El **error estándar estimado** de la estimación puntual es:

$$\widehat{se}(\hat{p}) = \sqrt{\frac{\hat{p}\,\hat{q}}{n}}$$

Se puede tener una confianza del $(1 - \alpha)\,100\%$ de que el error que se comete al estimar p por $\hat{p}$ no excederá de $z_{\alpha/2}\sqrt{\frac{\hat{p}\,\hat{q}}{n}}$ (Ver Figura 7)

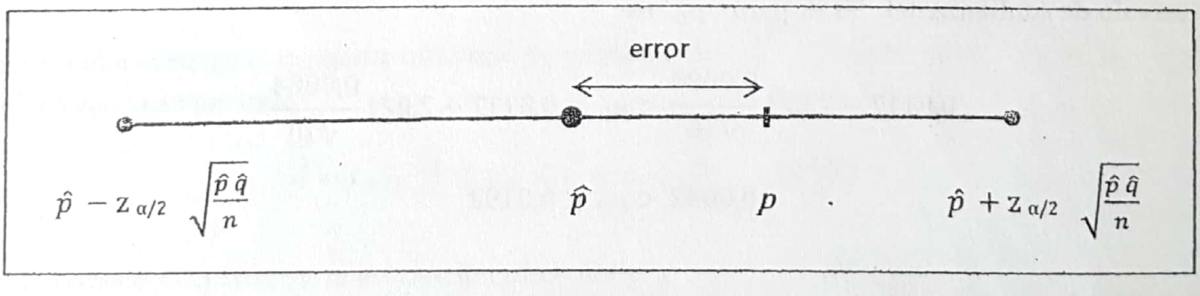

Figura 7: Error al estimar p por $\hat{p}$

Ejemplo 6

En una muestra aleatoria de 500 familias que poseen internet en una cierta ciudad se encontró que 340 se habían suscrito al servicio de banda ancha. Encuentra un intervalo de confianza del 95% para la proporción de familias de la ciudad que se suscribieron a banda ancha.

La estimación puntual de $\hat{P}$ es $\hat{p} = \frac{340}{500} = 0{,}68$

$(1-\alpha)\,100\% = 95\,\% \rightarrow \textit{grado de confianza} \quad \alpha = 0{,}05 \rightarrow \frac{\alpha}{2} = 0{,}025 \qquad z_{0,025} = 1{,}96$

El intervalo de confianza para p es:

$$0{,}68 - 1{,}96\sqrt{\frac{0{,}68\ 0{,}32}{500}} < p < 0{,}68 + 1{,}96\sqrt{\frac{0{,}68\ 0{,}32}{500}}$$

$$0{,}64 < p < 0{,}72$$

En el *Ejemplo 6* se tiene una confianza del 95% que la proporción muestral $\hat{p} = 0{,}68$ difiere de la proporción poblacional desconocida p en una cantidad que no excede 0,04.

Determinemos ahora *¿ Qué tamaño debe tener la muestra para asegurar que el error al estimar p sea menor que una cantidad especificada e ?*

En este caso debe escogerse n de modo que

$$z_{\alpha/2}\sqrt{\frac{\hat{p}\,\hat{q}}{n}} < e$$

Si $\hat{p}$ se usa como una estimación de p, se puede tener una confianza de $(1-\alpha)\,100\%$ de que el error será menor que una cantidad e cuando el tamaño de la muestra n sea el menor natural que verifique la inecuación

$$n > \left(\frac{z_{\alpha/2}}{e}\right)^2 \hat{p}\,\hat{q}$$

De modo que se podría tomar una muestra preliminar de tamaño $n \geq 30$ para obtener una estimación de p y de q. Entonces resultaría posible determinar, en forma aproximada, cuántas observaciones se necesitan para proporcionar el grado deseado de precisión.

Ejemplo 7

*¿ Qué tan grande se requiere que sea la muestra del **Ejemplo 6** si se desea tener una confianza del 95% de que se cometerá un error inferior a 0,02 en la estimación de p ?*

Las 500 familias constituyen una muestra preliminar que proporciona la estimación $\hat{p} = 0{,}68$ que permite obtener $n \geq \dfrac{1{,}96^2\ 0{,}68\ 0{,}32}{0{,}02^2} = 2090$

Por lo tanto, si la estimación por intervalo de confianza de p se basa en una muestra aleatoria de tamaño 2090 se puede tener una confianza del 95 % que la proporción muestral diferirá de la poblacional en menos de 0,02.

En los casos en que resulta impráctico obtener una estimación de p para determinar el tamaño de la muestra con un nivel dado de confianza, el límite inferior para n puede establecerse empleando la siguiente regla empírica :

Si $\hat{p}$ se utiliza como una estimación de p, se puede tener una confianza al menos del $(1-\alpha)$ 100% de que el error no excederá una cantidad especificada e cuando el tamaño de la muestra sea $n \geq \left(\frac{z_{\alpha/2}}{2\,e}\right)^2$

Ejemplo 8

*¿ Qué tan grande se requiere que sea la muestra del **Ejemplo 6** si se desea tener una confianza del 95% de que se cometerá un error inferior a 0,02 en la estimación de p y no se dispone de una estimación puntual de p ?*

A diferencia del *Ejemplo 7*, asumimos que no se ha tomado una muestra preliminar para obtener una estimación de p. En consecuencia, se puede tener una confianza al menos del 95% de que la proporción muestral no diferirá de la poblacional en más de 0,02 si se toma una muestra de tamaño $n \geq \frac{1{,}96^2}{4\,.0{,}02^2} = 2401$

Se observa que, la información concerniente a p que se obtuvo en una muestra preliminar, o tal vez, por experiencia pasada, permite seleccionar una muestra más pequeña en tanto se conserve el grado de seguridad requerido.

Estimación de la diferencia de dos proporciones

Se desea estimar la diferencia de proporciones para las variables aleatorias binomiales; X_1 y X_2 :

$$p_{X_1} - p_{X_2}$$

Se seleccionan muestras aleatorias independientes de tamaño n_1 y n_2 una para cada población binomial de parámetros p_{X_1} y p_{X_2} .

Se determinan los números x_1 y x_2 de éxitos para cada variable binomial y se obtienen las estimaciones puntuales de las proporciones muestrales:

$$\widehat{p_1} = \frac{x_1}{n_1} \quad y \quad \widehat{p_2} = \frac{x_2}{n_2}$$

La diferencia $\widehat{p_1} - \widehat{p_2}$ proporciona la estimación puntual del estadístico $\widehat{P_1} - \widehat{P_2}$ que tiene una distribución muestral aproximadamente normal con media $\mu_{\widehat{P_1}-\widehat{P_2}} = \widehat{p_1} - \widehat{p_2}$ y varianza $\sigma_{\widehat{P_1}-\widehat{P_2}} = \frac{p_1\, q_1}{n_1} + \frac{p_2\, q_2}{n_2}$

Entonces la variable aleatoria Z tiene una distribución Normal Estandard

$$Z = \frac{(\widehat{P_1} - \widehat{P_2}) - (p_{X_1} - p_{X_2})}{\sqrt{\frac{\widehat{P_1}\,\widehat{Q_1}}{n_1} + \frac{\widehat{P_2}\,\widehat{Q_2}}{n_2}}}$$

Y se puede asegurar que $P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha$, si $z_{\alpha/2}$ es el valor de z por encima del cual se encuentra un área de $\alpha/2$

Si reemplazamos Z y los estadísticos $\widehat{P_1}$, $\widehat{P_2}$, $\widehat{Q_1} = 1 - \widehat{P_1}$ y $\widehat{Q_2} = 1 - \widehat{P_2}$ por las estimaciones obtenidas a partir de una muestra única de cada población

$$\widehat{p_1} = \frac{x_1}{n_1} \;,\; \widehat{p_2} = \frac{x_2}{n_2} \;,\quad \widehat{q_1} = 1 - \widehat{p_1} \quad y \quad \widehat{q_2} = 1 - \widehat{p_2}$$

y se trata de muestras grandes independientes que cumplen las condiciones

$$n_1\widehat{p_1} \geq 5\,, n_1\widehat{q_1} \geq 5 \qquad n_2\widehat{p_2} \geq 5\,, n_2\widehat{q_2} \geq 5$$

se obtiene el siguiente intervalo de confianza del $(1-\alpha)\,100\%$ para $p_{X_1} - p_{X_2}$

$$(\widehat{p_1} - \widehat{p_2}) - z_{\alpha/2}\sqrt{\frac{\widehat{p_1}\,\widehat{q_1}}{n_1} + \frac{\widehat{p_2}\,\widehat{q_2}}{n_2}} < p_{X_1} - p_{X_2} < (\widehat{p_1} - \widehat{p_2}) + z_{\alpha/2}\sqrt{\frac{\widehat{p_1}\,\widehat{q_1}}{n_1} + \frac{\widehat{p_2}\,\widehat{q_2}}{n_2}}$$

IC(10)

El error estándar estimado de la estimación puntual es:

$$se\,(\widehat{p_1} - \widehat{p_2}) = \sqrt{\frac{\widehat{p_1}\,\widehat{q_1}}{n_1} + \frac{\widehat{p_2}\,\widehat{q_2}}{n_2}}$$

Ejemplo 9

Se está considerando cambiar el procedimiento de fabricación de partes. Se toman muestras del procedimiento actual y del nuevo para determinar si éste último es mejor. Si 75 de 1500 artículos del procedimiento actual presentan defectos y lo mismo sucedió con 80 de 2000 partes elaboradas con el nuevo procedimiento, determina el intervalo de confianza del 90% para la diferencia de las fracciones de partes defectuosas entre ambos procesos.

Se obtienen las estimaciones puntuales: $\widehat{p_1} = \frac{75}{1500} = 0{,}05$, $\widehat{p_2} = \frac{80}{2000} = 0{,}04$

$$\widehat{p_1} - \widehat{p_2} = 0{,}01$$

$(1-\alpha)\,100\% = 90\,\% \rightarrow grado\ de\ confianza \quad \alpha = 0{,}10 \rightarrow \frac{\alpha}{2} = \ 0{,}05$, $z_{0,05} = 1{,}645$

El intervalo de confianza para $p_{X_1} - p_{X_2}$ es:

$$0{,}01 - 1{,}645\sqrt{\frac{0{,}05\ 0{,}95}{1500} + \frac{0{,}04\ 0{,}96}{2000}} < p_{X_1} - p_{X_2} < 0{,}01 + 1{,}645\sqrt{\frac{0{,}05\ 0{,}95}{1500} + \frac{0{,}04\ 0{,}96}{2000}}$$

$$-0{,}0017 < p_{X_1} - p_{X_2} < 0{,}0217$$

Dado que el intervalo contiene el valor cero, no hay razón para creer que el nuevo procedimiento ocasionó una disminución significativa en la proporción de piezas defectuosas con respecto al método actual.

Estimación de la Varianza

Si se toma una muestra de tamaño n de una población normal con varianza σ_X^2 y se calcula s^2 , la varianza muestral para esa única muestra, se obtiene una estimación puntual del estadístico varianza muestral S^2.
El estadístico ji cuadrado : χ^2 es una variable aleatoria con distribución muestral conocida cuando las muestras de tamaño n se seleccionan de una población normal

$$\chi^2 = \frac{(n-1)S^2}{\sigma_X^2}$$

La DM (χ^2) se identifica con el nombre de la variable y los grados de libertad ν. Se dice; la distribución ji cuadrado con $\nu = n - 1$ grados de libertad

Entonces se puede escribir que $P(\chi^2_{1-(\alpha/2),\nu} < \chi^2 < \chi^2_{\alpha/2,\nu}) = 1 - \alpha$

si $\chi^2_{1-(\alpha/2),\nu}$ y $\chi^2_{\alpha/2,\nu}$ son los valores de la distribución ji cuadrado con ν= n-1 grados de libertad, con áreas $1 - \frac{\alpha}{2}$ y $\frac{\alpha}{2}$ respectivamente a su derecha

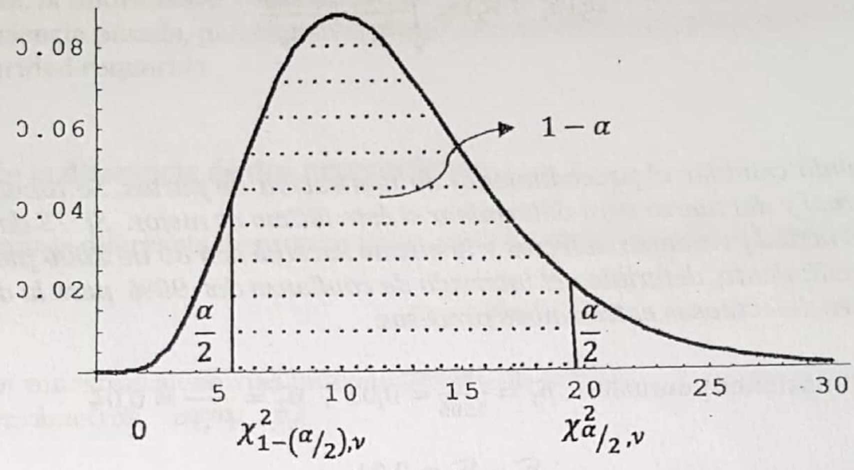

Figura 8: La distribución ji cuadrado con ν= n-1 grados de libertad
Significado de $\chi^2_{1-(\alpha/2),\nu}$ y $\chi^2_{\alpha/2,\nu}$

Al sustituir χ^2 resulta:

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

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

Si para una única muestra aleatoria de tamaño n se calcula la varianza muestral s^2; se obtiene el siguiente intervalo de confianza de $(1-\alpha)$ 100% para σ_X^2, la varianza poblacional

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2,\nu}} < \sigma_X^2 < \frac{(n-1)s^2}{\chi^2_{(1-\frac{\alpha}{2}),\nu}}$$

IC(11)

Donde la población es normal y $\chi^2_{1-(\alpha/2),\nu}$ y $\chi^2_{\alpha/2,\nu}$ son los valores de χ^2 con ν=n-1 grados de libertad, con áreas $1 - \frac{\alpha}{2}$ y $\frac{\alpha}{2}$ respectivamente a su derecha

Ejemplo 10

Sean los pesos en decagramos de 10 paquetes de semillas de forrajes distribuidas por una agro-empresa:

46,4	46,1	45,8	47,0	46,1
45,9	45,8	46,9	45,2	46,0

Encuentra un intervalo de confianza del 95% para la varianza de los paquetes de semillas de forrajes que se distribuyeron por esta agro-empresa. Supone que los pesos se distribuyen normalmente.

La estimación puntual de S^2 es

$$s^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n}}{n-1} = \frac{21.273{,}12 - \frac{461{,}2^2}{10}}{9} = 0{,}286$$

Para n=10 y x_i $i = 1, 2, \ldots, 10$ los valores observados en la muestra

$(1 - \alpha)\,100\% = 95\,\% \rightarrow$ *grado de confianza* $\alpha = 0{,}05 \rightarrow \frac{\alpha}{2} = 0{,}025$

ν=9 $\quad \chi^2_{1-(\alpha/2),\nu} = \chi^2_{0,95;9} = 2{,}700 \qquad \chi^2_{\alpha/2,\nu} = \chi^2_{0,05;9} = 19{,}023$

El intervalo de confianza para σ_X^2 es:

$$\frac{9\,.\,0{,}286}{19{,}023} < \sigma_X^2 < \frac{9\,.\,0{,}283}{2{,}700}$$

$$0{,}135 < \sigma_X^2 < 0{,}953$$

Estimación de la razón de dos varianzas

Para determinar el intervalo de confianza para $\frac{\sigma^2_{X_1}}{\sigma^2_{X_2}}$ el cociente de dos varianzas poblacionales, con las poblaciones X_1 y X_2 Normales se emplea como estimador o función de decisión al estadístico $\frac{S_1^2}{S_2^2}$ el cociente de las varianzas muestrales para muestras aleatorias independientes de tamaños n_1 y n_2

F es el estadístico que se define mediante el cociente $F = \dfrac{\frac{U}{\nu_1}}{\frac{V}{\nu_2}}$

Donde U es una variable ji cuadrado con $\nu_1 = n_1 - 1$ grados de libertad : $U = \frac{(n-1)S_1^2}{\sigma^2_{X_1}} = \chi^2_{1,\nu_1}$

y V es una variable ji cuadrado con $\nu_2 = n_2 - 1$ grados de libertad : $V = \frac{(n-1)S_2^2}{\sigma^2_{X_2}} = \chi^2_{2,\nu_2}$

Entonces reemplazando de obtiene la variable F de Fisher $F = \dfrac{\frac{(n-1)S_1^2}{\sigma^2_{X_1}\,\nu_1}}{\frac{(n-1)S_2^2}{\sigma^2_{X_2}\,\nu_2}} = \frac{S_1^2}{S_2^2}\,\frac{\sigma^2_{X_2}}{\sigma^2_{X_1}}$

Se puede determinar el intervalo de confianza de $\frac{\sigma^2_{X_1}}{\sigma^2_{X_2}}$ empleando el estadístico F con un distribución muestral llamada F de Fisher con ν_1, ν_2 grados de libertad , $\nu_1 = n_1 - 1$ y $\nu_2 = n_2 - 1$

$$F = \frac{S_1^2 \quad \sigma^2_{X_2}}{S_2^2 \quad \sigma^2_{X_1}}$$

Se puede decir que $P\left(f_{1-(\alpha/2),\nu_1,\nu_2} < F < f_{\alpha/2,\nu_1,\nu_2} \right) = 1 - \alpha$

Donde $f_{1-(\alpha/2),\nu_1,\nu_2}$ y $f_{\alpha/2,\nu_1,\nu_2}$ son los valores de la distribución F con con ν_1, ν_2 grados de libertad , con áreas $1 - \frac{\alpha}{2}$ y $\frac{\alpha}{2}$ respectivamente a su derecha

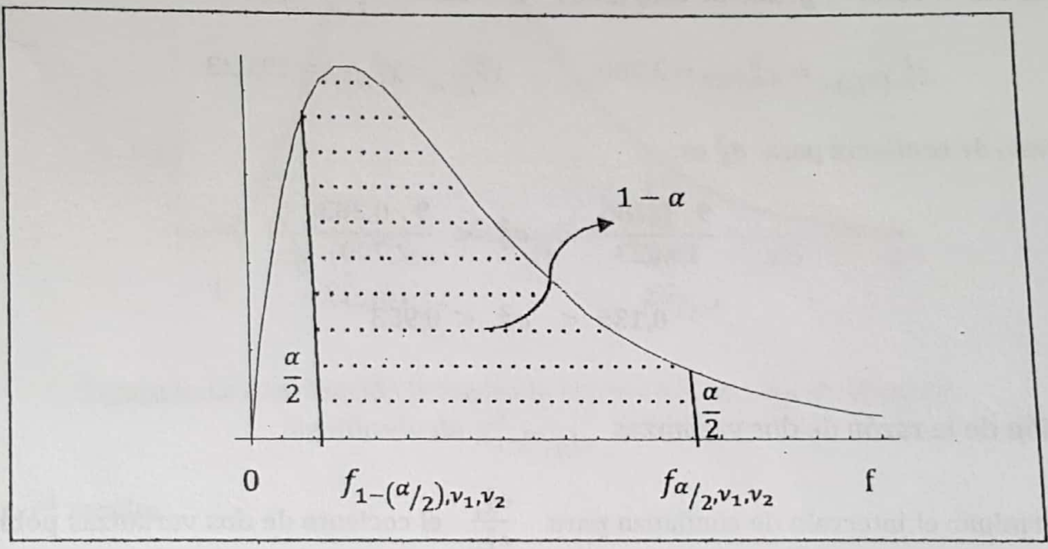

Figura 9 : La distribución F de Fischer con ν_1, ν_2 grados de libertad
Significado de $f_{1-(\alpha/2),\nu_1,\nu_2}$ y $f_{\alpha/2,\nu_1,\nu_2}$

Al sustituir F resulta:

$$P\left(f_{1-(\alpha/2),\nu_1,\nu_2} < \frac{S_1^2 \quad \sigma^2_{X_2}}{S_2^2 \quad \sigma^2_{X_1}} < f_{\alpha/2,\nu_1,\nu_2} \right) = 1 - \alpha$$

$$P\left(\frac{1}{f_{\alpha/2,\nu_1,\nu_2}} \frac{S_1^2}{S_2^2} < \frac{\sigma^2_{X_1}}{\sigma^2_{X_2}} < \frac{1}{f_{1-(\alpha/2),\nu_1,\nu_2}} \frac{S_1^2}{S_2^2} \right) = 1 - \alpha$$

Hay que recordar que el valor de F depende del orden de los grados de libertad y que

$$\frac{1}{f_{1-(\alpha/2),\nu_1,\nu_2}} = f_{\alpha/2,\nu_2,\nu_1}$$

Y se obtiene:

$$P\left(\frac{1}{f_{\alpha/2,\nu_1,\nu_2}} \frac{S_1^2}{S_2^2} < \frac{\sigma^2_{X_1}}{\sigma^2_{X_2}} < f_{\alpha/2,\nu_2,\nu_1} \frac{S_1^2}{S_2^2} \right) = 1 - \alpha$$

Para dos únicas muestras de tamaños n_1 y n_2 tomadas en forma independiente de las dos poblaciones normales, se calcula el cociente $\frac{s_1^2}{s_2^2}$ con las estimaciones puntuales s_1^2 y s_2^2 de las varianzas que se obtienen para cada una de las únicas muestras de las poblaciones cuyas varianzas se comparan.

De este modo, se obtiene el siguiente intervalo de confianza de $(1-\alpha)$ 100% para $\frac{\sigma_{X_1}^2}{\sigma_{X_2}^2}$

$$\frac{s_1^2}{s_2^2}\,\frac{1}{f_{\alpha/2,\,\nu_1,\,\nu_2}} < \frac{\sigma_{X_1}^2}{\sigma_{X_2}^2} < \frac{s_1^2}{s_2^2}\, f_{\alpha/2,\,\nu_2,\,\nu_1}$$

IC(12)

Donde $f_{\alpha/2,\nu_1,\nu_2}$ es el valor de la distribución F con con ν_1, ν_2 grados de libertad , con un área de $\frac{\alpha}{2}$ a su derecha y $f_{\alpha/2,\nu_2,\nu_1}$ es un valor que deja un área de $\frac{\alpha}{2}$ a su derecha para F con ν_2, ν_1 grados de libertad.

Ejemplo 11

En el Ejemplo 4 , se obtuvo un intervalo de confianza para la diferencia en los contenidos promedios de ortofósforo en dos estaciones sobre un río y se supuso que las varianzas de las poblaciones normales eran distintas. Justifica este supuesto encontrando un intervalo de confianza del 98% para $\frac{\sigma_{X_1}^2}{\sigma_{X_2}^2}$ y para $\frac{\sigma_{X_1}}{\sigma_{X_2}}$ (razón de desviaciones estándar) *para* $\sigma_{X_1}^2$ *y* $\sigma_{X_2}^2$ *las varianzas de las poblaciones de contenidos de ortofósforo en las estaciones 1 y 2 respectivamente.*

En este caso $n_1 = 15$ y $s_1^2 = 3{,}07^2$; $n_2 = 12$ y $s_2^2 = 0{,}80^2$

$(1-\alpha)\,100\% = 98\,\% \rightarrow$ *grado de confianza* $\alpha = 0{,}02 \rightarrow \frac{\alpha}{2} = 0{,}01$ $\qquad \nu_1 = 14$, $\nu_2 = 11$

$$f_{\alpha/2,\nu_1,\nu_2} = f_{0{,}01,14,11} \cong 4{,}30 \;, \qquad f_{\alpha/2,\nu_2,\nu_1} = f_{0{,}01,11,14} \cong 3{,}87$$

(interpolando linealmente)

El intervalo de confianza es

$$\frac{3{,}07^2}{0{,}80^2}\,\frac{1}{4{,}30} < \frac{\sigma_{X_1}^2}{\sigma_{X_2}^2} < \frac{3{,}07^2}{0{,}80^2}\,3{,}87 \qquad \Rightarrow \qquad 3{,}425 < \frac{\sigma_{X_1}^2}{\sigma_{X_2}^2} < 56{,}991$$

El intervalo de confianza del 98 % para la razón de las desviaciones estándar de las poblaciones resulta:

$$1{,}851 = \sqrt{3{,}425} < \frac{\sigma_{X_1}}{\sigma_{X_2}} < \sqrt{56{,}991} = 7{,}549$$

Dado que los intervalos no contienen al valor 1 , dejan fuera la posibilidad de que los cocientes sean iguales a 1 : $\frac{\sigma_{X_1}^2}{\sigma_{X_2}^2} = 1$ y $\frac{\sigma_{X_1}}{\sigma_{X_2}} = 1$. Entonces son correctas las afirmaciones $\sigma_{X_1}^2 \neq \sigma_{X_2}^2$ y $\sigma_{X_2} \neq \sigma_{X_2}$, respectivamente: $\frac{\sigma_{X_1}^2}{\sigma_{X_2}^2} \neq 1$ y $\frac{\sigma_{X_1}}{\sigma_{X_2}} \neq 1$.

Tabla de Intervalos de Confianza

Se agrega una Tabla de Intervalos de Confianza para que se emplee como auxilio al momento de estar frente a un problema. En la tabla se indican, para cada parámetro a estimar por intervalo de confianza, las características de la población y de la muestra, el estadístico que se emplea como estimador y la acción o estimación puntual que se obtiene a partir de la muestra, la variable aleatoria empleada en la estimación y su distribución muestral y el intervalo de confianza del $(1-\alpha)100\%$ para el parámetro poblacional. En la última columna se indica el error estándar.

De esta forma, para resolver un problema, lo primero es identificar el parámetro a estimar por intervalo de confianza. Luego, se observan las características de la población o poblaciones de las que se extraen la o las muestras y el tamaño de la muestra. Aquí nos preguntamos: ¿La población es Normal o aproximadamente normal? ¿es no normal o desconocida? ¿ es binomial? ¿ el tamaño de la muestra es mayor o menor a 30 ? ¿ es la varianza conocida o desconocida ?. . . En el caso de dos muestras tomadas de dos poblaciones, además, nos preguntamos ¿las muestras son independientes o apareadas? ¿ pueden suponerse varianzas iguales? . . .

Una vez que decidimos la fila de la tabla que se ajusta a nuestro problema no hay más que calcular el intervalo de confianza, reemplazando con los valores que surgen de la muestra y de la distribución muestral para el valor del nivel de confianza α establecido.