INTERVALOS DE CONFIANZA Y PRUEBAS DE HIPÓTESIS

Carrera : Ingenierias Varias (IME-261)

Una estimación por intervalo de un parámetro problacional θ es un intervalo de la forma $\theta_I < \theta < \theta_S$, donde θ_I y θ_S, dependen del valor del estadístico $\widehat{\Theta}$ para una muestra partícular y también de su distribución muestral. De esta forma, una muestra aleatoria de las calificaciones verbales de la P.A.A. de los estudiantes universitarios de una clase de cuarto año medio podría producir un intervalo de 530 a 550, dentro del cual se esperaría encontrar el promedio real de todas las calificaciones verbales de la P.A.A. $(\theta = \mu)$ de dicho grupo. Los valores de los puntos extremos, 530 y 550, dependerán de la media muestra calculada $\overline{X}$ y de su respectiva distribución muestral. Conforme el tamaño de la muestra se hace más grande, se sabe que $\sigma^2_{\bar{x}} = \frac{\sigma^2}{n}$ disminuye, y por consiguiente la estimación posiblemente se acerca más al parámetro μ, lo que resulta en un intervalo más pequeño. Así el intervalo de estimación probabilística indica, por su longitud, la precisión de la estimación puntual.

Ya que muestras distintas generalmente dan valores distintos de $\widehat{\Theta}$ y, por tanto, de θ_I y de θ_S ,estos puntos extremos del intervalo son los valores de las variables aleatorias correspondientes $\widehat{\Theta}_I$ y $\widehat{\Theta}_S$. A partir de la distribución muestral de $\widehat{\Theta}$ será posible determinar y $\widehat{\theta}_I$ y $\widehat{\theta}_S$ tales que la $P(\widehat{\Theta}_I < \theta < \widehat{\Theta}_S)$ sea igual para cualquier valor fraccional positivo que se desee especificar. Si, por ejemplo, se encuentran $\widehat{\theta}_I$ y $\widehat{\theta}_S$, tales que:

$$P(\widehat{\Theta}_I < \theta < \widehat{\Theta}_S) = 1 - \alpha$$

para $0 < \alpha < 1$, entonces se tiene una probabilidad de $(1 - \alpha)$ de seleccionar una muestra aleatoria que produzca un intervalo que contenga a θ. El intervalo $\widehat{\theta}_I < \theta < \widehat{\theta}_S$, se llaman límites de confianza inferior y superior. De tal manera que cuando $\alpha = 0.05$, se tiene un intervalo de confianza del 95%, y cuando $\alpha = 0.01$ se obtiene otro más amplio del 99%. Entre mayor es el intervalo de confianza, se tiene más seguridad de que el intervalo contiene el parámetro desconocido. Por supuesto, es mejor tener un intervalo de confianza de que la vida promedio de determinado transistor de televisor está entre 6 y 7 años, que un 99, 9 de confianza que esté entre 3 y 10 años. Desde un punto de vista ideal, es preferible un intervalo pequeño con un alto grado de confianza. Algunas veces, las restricciones en el tamaño de la muestra evitan la obtención de intervalos pequeños sin sacrificar el grado de confianza.

Todo intervalo de confianza tiene asociado un par de hipótesis del tipo:

$$H_0 : \theta = \theta_0 \qquad \text{v/s} \qquad H_1 : \theta \neq \theta_0$$

Una hipótesis estadística es una afirmación o conjetura sobre alguna característica desconocida de la población bajo estudio. Dichas afirmaciones o conjeturas, son respecto de los parámetros de una distribución de probabilidades o la distribución en si. En esta asignatura, el interés esta en los parámetros de las distribuciones de probabilidad, lo cual se define como: **prueba de hipótesis.**

$\boldsymbol{H_0}$: Denominada *Hipótesis Nula,* la cual normalmente representa la posición conservativa en una investigación o pregunta de investigación, esto pues es necesario que se considere verdadera, a menos que exista suficiente evidencia de lo contrario.

$\boldsymbol{H_1}$: Denominada *Hipótesis Alternativa,* esta se construye como negación o opusto de la hipótesis nula, y normalmente corresponde a la posición rupturista de una investigación o pregunta de investigación. Esta estrategia, asegura que la aceptación de la hipótesis alternativa como verosimil, es producto de la negación de la hipótesis nula, y por tanto, se tendrá suficiente evidencia para dicha decisión.

α : Se denomina nivel de significación, y corresponde a la probabilidad de rechazar la hipótesis nula H_0 cuando ésta es verdadera.

$1 - \alpha$: Se denomina nivel de confianza y corresponde a la probabilidad de aceptar la hipótesis nula cunado ésta es verdadera.

Asociado al juego de hipótesis, se tiene un criterio de decisión, el cual es:

Rechazar H_0 *ssi* $\theta_0 \notin (a, b)$

Cuando el parámetro es la media poblacional, el total poblacional o la proporción poblacional entonces un intervalo de confianza se puede construir de la siguiente forma: $P(\hat{\Theta} + \epsilon \leq \hat{\Theta} + \epsilon) = 1 - \alpha,$ donde como por ejemplo el parámetro $\Theta = \mu,$ por lo tanto se debe consturir el intervalo de confianza para la media poblacional μ cuando σ^2 es desconocido. La expresión para el intervalo de confianza al $(1 - \alpha)\,\%$ es:

$$P(\overline{X} - t_{(n-1, 1-\frac{\alpha}{2})} * \frac{S}{\sqrt{n}} \leq \mu \leq \overline{X} + t_{(n-1, 1-\frac{\alpha}{2})} * \frac{S}{\sqrt{n}}) = 1 - \alpha$$

Las fórmulas para estos casos la encontrará en APUNTES ANEXOS.

PROBLEMA

En el marco de una investigación sobre religiosidad de alumnos de enseñanza media de la region de la Araucania, se seleccionaron ala azar 30 colegios, 10 municipalizados, 10 subvencionados y 10 particulares. En cada colegio, se aplicó a una muestra aleatorizada y representativa de alumnos una ancuesta sobre reliogiosidad, la cual proporciona un indice de 0 a 100, donde "0" signfica nada de religiosidad y "100" la maxima religiosidad.

Los datos que se entregan en la tabla siguiente y corresponden a la media del indice de religiosidad de cada colegio.

Subvencionado	**Municipal**	**Particular**
58.00	74.50	31.20
56.00	64.80	19.10
61.00	74.40	26.70
53.80	64.00	25.20
64.00	76.10	16.70
65.20	56.00	16.00
62.30	81.50	16.00
54.40	65.00	20.80
61.50	75.30	22.10
64.20	73.60	29.50

ACTIVIDADES

a) Caracterizar el indice de religisidad para cada tipo de colegio. Utilice recursos gráficos para apoyar esta caracterización.

b) Determine el tipo de colegio donde el indice de religiosidad es mayor. Argumente estadísticamente, con alguno de los métodos conocidos.

SOLUCION.-

INFORMACION GENERAL PREVIA.-

	Subvencionado	**Municipal**	**Particular**
SUMA	600.40	705.20	223.30
MEDIA	60.04	70.52	22.33
VARIANZA	17.67	59.05	31.34
D. ESTAND	4.20	7.68	5.60

Definición de variable aleatoria:

I.- Indice de religiosidad en porcentaje.

Como el objetivo final del estudio es determinar cuál de los 3 tipos de colegio presenta el mayor indice de religiosidad, entonces parece lógico comparar el tipo de colegio que presenta la respuesta mayor respecto del punto de vista del indice de religiosidad promedio, con la respuesta promedio de los otros 2 tipos de colegio, es decir solo es necesario hacer 2 comparaciones, los cuales son:

i) MUNICIPAL- SUBVENCIONDO
ii) MUNICIPAL- PARTICULAR

a) Caracterizar el indice de religiosidad de los tipo de colegio. Utilice recursos gráficos para apoyar esta caracterización.

Caracterizar el indice de religisidad para cada tipo de colegio, significa describir su comportamiento en términos de tendencia central y dispersión. Esto puede materializarse adecuadamente si suponemos simetría y un coeficiente de kurtosis cercano a 0.263, que se resume al pedir normalidad del indice de religiosidad para cada tipo de colegio. Si aceptamos que la distribución de probabilidades de la variable indice de religiosidad es normal, entonces podemos caracterizar a la variable en estudio construyendo intervalos de confianza para μ y σ^2

Observación:

Probar normalidad del indice de religiosidad para tipo de colegio, es fundamental para construir los intervalos de confianza de μ y σ^2. Sin esta característica y un tamaño de muestra pequeño $(n < 25)$, se desconoce la distribución de probabilidades asociada en cada caso. Sin embargo, en este ejemplo asumiremos la normalidad.

a.1.- **Intervalos de Confianza para μ para cada tipo de colegio**

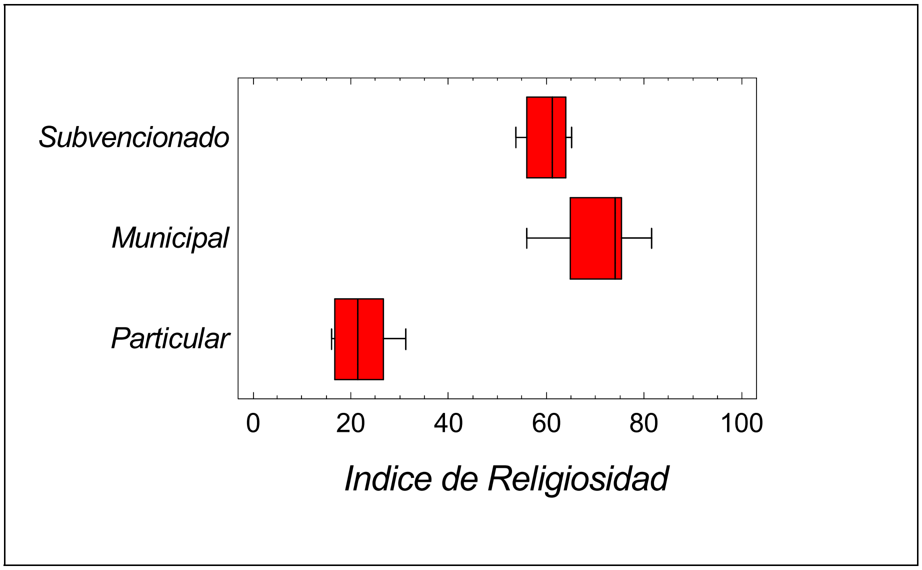
$P(\hat{\Theta} + \in \leq \hat{\Theta} + \in) = 1 - \alpha$, donde como por ejemplo el parámetro $\Theta = \mu$, por lo tanto se debe construir el intervalo de confianza para la media poblacional μ cuando σ^2 es desconocido. La expresión para el intervalo de confianza al $(1 - \alpha)\%$ es:

$$P\left(\overline{X} - t_{(n-1,1-\frac{\alpha}{2})} * \frac{S}{\sqrt{n}} \leq \mu \leq \overline{X} - t_{(n-1,1-\frac{\alpha}{2})} * \frac{S}{\sqrt{n}}\right) = 1 - \alpha$$

Consideremos un nivel de significación $\alpha = 0.05$, entonces

$$t_{(n-1,1-\frac{\alpha}{2})} = t_{(9.0.975)} = 2.262$$

Gráfico de Caja del Indice de Religiosidad por tipo de colegio



Colegio Subvencionado:

$$P\,(\,60.04\,-\,2.262*\frac{4.2}{\sqrt{10}}\,\leq\,\mu\,\leq 60.04+\,2.262*\frac{4.2}{\sqrt{10}})\,=\,0.95$$

$$P\,(\,57.04\,\leq\,\mu\,\leq 63.04\,)\,=\,0.95$$

Colegio Municipal:

$$P\,(\,70.52\,-\,2.262*\frac{7.68}{\sqrt{10}}\,\leq\mu\,\leq 70.52\,+\,2.262*\frac{7.68}{\sqrt{10}})\,=\,0.95$$

$$P\,(\,65.03\,\leq\,\mu\,\leq 76.01\,)\,=\,0.95$$

Colegio Particular:

$$P\,(\,22.33\,-\,2.262*\frac{5.6}{\sqrt{10}}\,\leq\,\mu\,\leq 22.33+\,+\,2.262*\frac{5.6}{\sqrt{10}}\,)\,=\,0.95$$

$$P\,(\,18.32\,\leq\,\mu\,\leq 26.34\,)\,=\,0.95$$

Por lo tanto como no existe intersección entre los intervalos de confianza construidos para μ con la información muestral de cada tipo de colegio, a priori podemos afirmar, que probabilisticamente los tipos de colegios estiman parametros distintos, entonces podemos asumir que es probable que el indice de religiosidad sea mayor en algún tipo de colegio respecto de los otros. Esto es evidente al observar que los colegios municipales son los que presentan la mayor media, en consecuencia existe evidencia estadística para sostener que el tipo de colegio donde se obtiene el mayor indice de religiosidad es el municipal.

a.2.- **Intervalos de Confianza para σ^2 de cada tipo de colegio**

Para caracterizar la varianza de los tres tipos de colegio que se estan evaluando respecto de sus condiciones para el indice de diversidad, es necesario construir intervalos de confianza para la varianza poblacional $P(\hat{\Theta}_{inf} \leq \Theta \leq \hat{\Theta}_{sup}) = 1 - \alpha,$ donde el parámetro $\Theta = \sigma^2$, por lo tanto se debe construir un intervalo de confianza para la varianza poblacional σ^2, cuando la distribución de probabilidades de la población de origen es normal. La expresión para el intervalo de confianza al $(1 - \alpha)$% es:

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

Consideremos un nivel de significación $\alpha = 0.05,$ entonces:

$$\chi^2_{(n-1,1-\frac{\alpha}{2})} = \chi^2_{(9,0.975)} = 19.023$$
$$\chi^2_{(n-1,\frac{\alpha}{2})} = \chi^2_{(9,0.025)} = 2.7$$

Colegio Subvencionado:

$$P\left(\frac{9*17.67}{19.023} \leq \sigma^2 \leq \frac{9*17.67}{2.7}\right) = 0.95$$

$$P(8.36 \leq \sigma^2 \leq 58.9) = 0.95$$

Colegio Municipal:

$$P\left(\frac{9*59.05}{19.023} \leq \sigma^2 \leq \frac{9*59.05}{2.7}\right) = 0.95$$

$$P(27.94 \leq \sigma^2 \leq 196.83) = 0.95$$

Colegio Particular:

$$P\left(\frac{9*31.34}{19.023} \leq \sigma^2 \leq \frac{9*31.34}{2.7}\right) = 0.95$$

$$P(14.83 \leq \sigma^2 \leq 104.47) = 0.95$$

Por lo tanto como existe intersección entre los intervalos de confianza construidos para σ^2 de cada tipo de colegio, a priori podemos afirmar que probabilisticamente los tipos de colegio tienen la misma varianza, esto es posible desde la perspectiva de estadística inferencial, porque los intervalos de confianza para las varianzas de los tres tipos de colegio en estudio tienen una zona de intersección común.

b) **Determine el tipo de colegio que presenta las mejores condiciones para la religiosidad. Argumente estadísticamente, con alguno de los métodos conocidos.**

Como el objetivo final del experimento es determinar cual de los 3 tipos de colegio es el que tiene el mayor indice de religiosidad, entonces parece lógico comparar el tipo de colegio que presenta la respuesta mayor respecto del punto de vista del indice de religiosidad, con la respuesta promedio de los otros 2 tipos de colegio, es decir solo es necesario hacer 2 comparaciones como máximo, los cuales son :

i) **Municipal - Subvencionada**
ii) **Municipal - Particular**

Dos tipos de colegio son comparables del punto de vista de la respuesta promedio, si las varianzas de los tipos de colegio involugrados son estadísticamente iguales. Por lo tanto la homocedasticidad la verificaremos construyendo los 2 intervalos de confianza respectivos, y verificando si ellos contienen la unidad (eso implica que $\sigma_1^2 = \sigma_2^2$).

El intervalo de confianza para la razón de 2 varianzas poblacionales es el siguiente :

$$P\left(\frac{S_1^2}{S_2^2} * f_{\left(\frac{\alpha}{2}\right)(n_1-1,n_2-1)} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq \frac{S_1^2}{S_2^2} * f_{\left(1-\frac{\alpha}{2}\right)(n_1-1,n_2-1)} \right) = 1-\alpha$$

Donde :

$$\alpha = 0.05$$

$$f_{\left(\frac{\alpha}{2}\right)(n_1-1,n_2-1)} = f_{(0.025)(9,9)} = \frac{1}{4.03} = 0.2481 \text{ , y}$$

$$f_{\left(1-\frac{\alpha}{2}\right)(n_1-1,n_2-1)} = f_{(0.975)(9,9)} = 4.03$$

1.- Entonces el intervalo de confianza de la razón de las varianzas para los colegios *Municipal/Subvencionado* es :

$$P\left(\frac{59.05}{17.67} * 0.2481 \leq \frac{\sigma_C^2}{\sigma_Q^2} \leq \frac{59.05}{17.67} * 4.03 \right) = 1-0.05$$

/ $P\left(0.8291 \leq \frac{\sigma_C^2}{\sigma_Q^2} \leq 13.4675 \right) = 0.95$, Por lo tanto como el intervalo de confianza al 95 % contiene la unidad, entonces podemos concluir que las varianzas de los tipos de colegio Municipal y Subvencionadao son iguales, es decir, los tipos de colegio involucrados en este caso son comparables estadísticamente.

2.- Entonces el intervalo de confianza de la razon de las varianzas para los colegios *Municipal/Particular* es :

$$P\left(\frac{59.05}{31.34}*0.2481 \leq \frac{\sigma_C^2}{\sigma_{LB}^2} \leq \frac{59.05}{31.34}*4.03\right) = 1 - 0.05$$

$$P\left(0.4675 \leq \frac{\sigma_C^2}{\sigma_{LB}^2} \leq 7.5932\right) = 0.95$$, Por lo tanto como el intervalo de confianza al 95 % contiene la unidad, entonces podemos concluir que las varianzas para los colegios *Municipal y Particular* son iguales, es decir, los tipos de colegio involucrados en este caso son comparables.

Finalmente podemos afirmar que las respuestas promedio de los 3 tipos de colegio son estadísticamente comparables, esto porque sus varianzas son estadísticamente iguales.

El parámetro es $\frac{\sigma_1^2}{\sigma_2^2}$, y su estimador es -$\frac{S_1^2}{S_2^2}$. ¿ Porqué?., lo cuál parece ser natural, ¿ Verdad?.

El parámetro para este caso es $\mu_1 - \mu_2$, y el estimador es $\overline{x}_1 - \overline{x}_2$, y continuación se muestra como se obtiene aplicando el método de máxima Verosimilitud.

Para determinar cuál de dos tipo de colegio es mejor respecto del indice de religiosidad, se construye la estimación probabilistica con algún nivel de significación, en este caso $(\alpha = 0.05)$ para la diferencia de medias poblacionales, y se verifica si el cero esta contenido en el intervalo de confianza determinado. Si el cero esta contenido, entonces ambos tipos de colegio, estadísticamente son iguales respecto del indice de religiosidad, si el cero no esta contenido en el intervalo de confianza encontrado, entonces, el tipo de colegio que estadísticamente es mejor respecto del indice de religiosidad, es el de mayor media muestral.

La expresión general para la estimación probabilistica de la diferencia de 2 medias poblacionales , es la siguiente :

$$P\big((\overline{x}_1 - \overline{x}_2) + t_{(\frac{\alpha}{2})(n_1+n_2-2)} * S_p * \sqrt[2]{\frac{1}{n_1} + \frac{1}{n_2}} \leq \mu_1 - \mu_2 \leq$$

$$(\overline{x}_1 - \overline{x}_2) + t_{(1-\frac{\alpha}{2})(n_1+n_2-2)} * S_p * \sqrt[2]{\frac{1}{n_1} + \frac{1}{n_2}}) = 1 - \alpha$$

$$\text{Donde , } \quad Sp = \sqrt[2]{\frac{(n_1-1)*S_1^2+(n_2-1)*S_2^2}{(n_1+n_2-2)}}$$

Dado que las varianzas de las poblaciones son desconocidas y se estiman por S_1^2 y S_2^2 , varianzas muestrales.

Donde :

$\alpha = 0.05$ (nivel de significación)

$t_{(0.025)(18)} = -2.101$ (valor de tabla t-Student)

$t_{(0.975)(18)} = 2.101$ (valor de tabla t-Student)

1. – Estimación Probabilistica para la diferencia de medias de los tipos de colegio Municipal y Subvencionado. Entonces el Intervalo de confianza es :

$S_p = \sqrt[2]{\frac{9*59.05+9*17.67}{18}} = 6.1935$, entonces la estimación Probablistica al

95% de confianza es :

$$P(\,(70.52 - 60.04) - 2.101*6.1935*\sqrt[2]{\frac{1}{10} + \frac{1}{10}} \leq \mu_c - \mu_q \leq$$

$$(70.52 - 60.04) + 2.101*6.1935*\sqrt[2]{\frac{1}{10} + \frac{1}{10}}\,) = 1 - 0.05$$

$$P(\,10.48 - 5.8194 \leq \mu_c - \mu_q \leq 10.48 + 5.8194) = 0.95$$

$$P(4.6606 \leq \mu_c - \mu_q \leq 16.2994) = 0.95$$

Por lo tanto, como el intervalo de confianza no contiene el cero, entonces el tipo de colegio que presenta el mayor indice de religiosidad son los Municipales, entonces no es necesario hacer otra hipótesis estadística. Y hemos probado estadísticamente que los colegios municipalizados es donde se alcanza el mayor indice de religiosidad.