Intervalos de confianza para la media y la proporción, cálculo del tamaño muestral

Dada la naturaleza de ciertas variables, se puede asociar a estas una distribución de probabilidad específica. La más conocida es la distribución normal. Sin embargo, existen otras distribuciones, como la distribución binomial, poisson, hipergeométrica, entre otras.

Distribución	**Uso**	Fórmula
Binomial	Discreta	$P(X=k)=\binom{N}{k}\cdot p^k q^{N-k}$
Geométrica	Discreta	$P(X=k)=(1-p)^{k-1}p$
Poisson	Discreta	$P(X=x)=\frac{e^{-\lambda}\lambda^x}{x!}$
Hipergeométrica	Discreta	$P(X=k)=\frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$

1. En el artículo de Araya, V. et al, Frecuencia de periodontitis en una muestra de diabéticos tipo 2 y no diabéticos de Santiago de Chile. Rev. chil. endocrinol. diabetes 2011; 4 (4): 251-256, se presenta un cuadro con los diagnóstico de periodontitis severa, según grupo y entre los diferentes exámenes de laboratorio se entregan los resultados obtenidos en triglicéridos para los pacientes con Periodontitis Severa.

Diabéticos	**Total casos**	**Periodontitis Severa**	**Pacientes con P. Severa Triglicéridos mg/dl**	
			Prom	DesvEst
SI	62	38	230	30
NO	65	32	150	70

a) Obtenga los Intervalos de confianza al 99% para la media de los triglicéridos de cada grupo. ¿qué puede concluir?

Respuesta
En este caso, en ambos grupos tenemos $n > 30$, por lo tanto, un intervalo de confianza para la media es:

$$(\bar{x} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}})$$

En este caso $\alpha = 0.01$ pues queremos un intervalo de confianza del 99%. Ahora reemplazamos valores para cada grupo.

Intervalo de confianza para la media de triglicéridos en diabéticos:

$$(230 - z_{0.995}\frac{30}{\sqrt{38}}, 230 + z_{0.995}\frac{30}{\sqrt{38}})$$

$$(217.4927, 242.5073)$$

Intervalo de confianza para la media de triglicéridos en no diabéticos:

$$(150 - z_{0.995}\frac{70}{\sqrt{32}}, 150 + z_{0.995}\frac{70}{\sqrt{32}})$$

$$(118.1979, 181.8021)$$

Podemos observar que los intervalos son disjuntos, es decir, pertenecer al grupo diabético o no diabético genera diferencias importantes en el nivel de triglicéridos al presentar Periodontitis Severa.

b) Se ha decidido reiterar el estudio, dado que algunos indicadores presentan demasiado error estadístico. Determine el tamaño muestral mínimo de pacientes diabéticos que le permita estimar el nivel medio de triglicéridos con un error no mayor a 6 mg/dl y un 90% de confianza en pacientes diabéticos con diagnóstico de periodontitis severa. Puede usar los resultados previos de referencia.

Respuesta
Como sólo tenemos información de los diabéticos con periodontitis obtenemos que el número de pacientes de periodontitis debe ser:

$$n_p = (\frac{z_{1-\alpha/2} \cdot \sigma}{error})^2$$

En este caso, $\alpha = 0.1$, luego $z_{1-\alpha/2} = z_{0.95} = 1.645$. Por lo tanto,

$$n_p = (\frac{1.645 \cdot 30}{6})^2 = 67.65062$$

Para saber el n note que la proporción de diabéticos que tiene periodontitis es $p = \frac{38}{62} = 0.6129032$, por lo tanto, el número de pacientes diabéticos necesario es:

$$n = \frac{n_p}{p} = \frac{67.65062}{0.6129032} = 110.3773$$

Por lo tanto, basta tomar $n = 111$ para asegurar que el error sea menor a 6 mg/dl.

2. Se tiene información de 61 pacientes hospitalizados en el Hospital Clínico de la UC por diagnóstico de neumonía adquirida en la comunidad (NAC). De estos pacientes, 6 fallecieron hasta 30 días después de la hospitalización.

a) Construya un intervalo de confianza de 95% para la proporción de fallecidos por NAC hasta 30 días después de la hospitalización. Indique el o los supuestos necesarios para la construcción del intervalo. Interprete el intervalo construido.

Respuesta
Sabemos que un intervalo para la proporción es:

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

En este caso, $\alpha = 0.05$,

$$\hat{p} = \frac{n^o\ de\ fallecidos}{pacientes\ hospitalizados} = \frac{6}{61} = 0.09836066$$

y $n = pacientes\ hospitalizados = 61$. Luego el intervalo es:

$$(0.09836066 \pm z_{0.975}\sqrt{\frac{0.09836066(1-0.09836066)}{61}})$$

$$(0.02362665, 0.1730947)$$

Supuestos: La variable de interés es dicotómica (muere o no muere), pero se asume que la proporción distribuye normal, para así poder crear el intervalo.

Interpretación: La proporción de pacientes hospitalizados fallecidos hasta 30 días después de la hospitalización varía entre 0,0233 y 0,1727 con confianza 95%.

b) Entre las variables que se asocian con la mortalidad por NAC está la temperatura corporal al ingreso al hospital, que para los 6 pacientes tuvo un promedio 37.7 grados y una desviación estándar 1.1 grados. Construya un intervalo de confianza de 90% para la temperatura promedio al ingreso al hospital de los fallecidos. Indique el o los supuestos necesarios para la construcción del intervalo. Interprete el intervalo construido.

Respuesta
En este caso tenemos sólo $n = 6$ pacientes, además de que la varianza no se asume conocida (sólo tenemos una estimación en la muestra). Como el $n < 30$ debemos utilizar la distribución t-student. El intervalo es de la siguiente forma:

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

En este caso, $\alpha = 0.1$ y reemplazamos el resto de los valores:

$$(37.7 \pm t_{(6-1,0.95)}\frac{1.1}{\sqrt{6}})$$

Finalmente el intervalo es:

$$(36.79512, 38.60488)$$

Interpretación: Con un 90% de confianza, la media de la temperatura de los fallecidos se encuentra en el intervalo construído.

Supuestos: Como no conocemos la varianza, por construcción debemos utilizar una t-student, bajo supuestos distribucionales de la varianza.

3. Un estudio con 34 niños es llevado a cabo con el objetivo de estimar los valores de referencia de las concentraciones séricas basales del factor de crecimiento similar a la Insulina-I (IGF-I) y Osteocalcina (ambos son medidos en ng/ml) en niños sanos. Los resultados, según grupo etario se presentan en la tabla siguiente:

Característica	**IGF - I**		**Osteocalcina**	
	4-6 años	7-10 años	4-6 años	7-10 años
Número de casos	15	19	15	19
Promedio	104	133	72	84
Desviación Estándar	46	35	16	18

a) ¿Hay evidencia, para alfa=5%, que permita afirmar que el nivel medio de IGF-I en niños de 7-10 años es superior a 120 ng/ml?

Respuesta
Planteamos las hipótesis:

$$H_0\colon\ \mu = 120 \qquad H_1\colon\ \mu > 120$$

En este caso, $\alpha = 0.05$ y sabemos que el estadístico del test es:

$$t_0 = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{133 - 120}{35/\sqrt{19}} = 1.61902$$

Y comparamos con una $t(n-1) = t(18) = 1.734$. Vemos que:

$$t_0 < t(n-1)$$

Por lo tanto, no hay evidencia para afirmar que el nivel medio de IGF-I es superior a 120 en niños de 7 a 10 años.

b) Basado en la información obtenida en este estudio, se desea implementar uno nuevo con niños de 4-6 años, de tal forma que permita estimar el nivel medio de Osteocalcina con un error no mayor a 3 ng/ml y el nivel medio de IGF-I con un error no mayor a 10 ng/ml, con un 95% de confianza. Para el tamaño de muestra propuesto por usted, ¿qué niveles de error se tendrían?

Respuesta
Calculamos los n para cada grupo:

Osteocalcina en niños de 4-6 años:
Sabemos que la desviación estándar es 16, además queremos un error o mayor a 3. Además $\alpha = 0.05$. Luego el n óptimo es:

$$n_1 = \left(\frac{z_{1-\alpha/2} \cdot \sigma}{3}\right)^2 = 109.2721$$

Tomamos $n_1 = 110$.

IGFI en niños de 4-6 años:
Sabemos que la desviación estándar es 46, además queremos un error o mayor a 10. Además $\alpha = 0.05$. Luego el n óptimo es:

$$n_2 = \left(\frac{z_{1-\alpha/2} \cdot \sigma}{10}\right)^2 = 81.28826$$

Tomamos $n_2 = 82$.
Como queremos que en ambos grupos se cumplan las condiciones, basta tomar $n = \max(n_1, n_2) = 110$.
El error esperable en este grupo, corresponde a

$$1.96 \cdot \frac{46}{\sqrt{110}} = 8.596419$$

En el otro grupo sería análogo pero considerando la desviación estándar correspondiente.