Probabilidad y estadística

MAT 041, Primer semestre

Francisco Cuevas Pacheco

5 de diciembre de 2022

Es por esto que dada una muestra aleatoria $\underline{X} = (X_1, X_2, \ldots, X_n)$ y una probabilidad dada $(1 - \alpha)$, resulta más adecuado encontrar estadísticas muestrales $L(\underline{X})$ y $U(\underline{X})$ tales que uno pueda establecer que

$$P(L(\underline{X}) \leq \theta \leq U(\underline{X})) = 1 - \alpha$$

Es por esto que dada una muestra aleatoria $\underline{X} = (X_1, X_2, \ldots, X_n)$ y una probabilidad dada $(1 - \alpha)$, resulta más adecuado encontrar estadísticas muestrales $L(\underline{X})$ y $U(\underline{X})$ tales que uno pueda establecer que

$$P(L(\underline{X}) \leq \theta \leq U(\underline{X})) = 1 - \alpha$$

Esto quiere decir que $IC(\theta) = [L(\underline{X})\ ;\ U(\underline{X})]$ **captura** al verdadero parámetro con probabilidad $1 - \alpha$.

Lectura equivocada

El verdadero parámetro se encuentra en $L(\underline{X}) \leq \theta \leq U(\underline{X})$ con un α porciento de confianza.

Construción del intervalo

Antertiormente hemos visto que un estadístico $T(\underline{X})$ nos permitía tener una estimación del parámetro buscado, además este estadístico al ser ua función de $\underline{X}$ corresponde a ser una variable aleatoria y por lo tanto tiene una función de densidad asociada. Usaremos entonces la distribución de $T(\underline{X})$ para construir nuestro intervalo.

El punto de partida para crear intervalos de confianza es mediante una **Cantidad Pivotal**

Algunos resultados

Invarianza estimador máximo verosímil

Considere $X_1, \ldots, X_n$ una muestra aleatoria de tamaño n proveniente de una densidad $f(x|\theta)$. Sea $\hat{\theta}$ el estimador máximo verosímil de θ, entonces, para cualquier función $h : \Theta \rightarrow \mathbb{R}$ se tiene que $h(\hat{\theta})$ es el estimador máximo verosimil de $h(\theta)$.

Algunos resultados pt.II

Distribución asintótica

Considere $X_1, \ldots, X_n$ una muestra aleatoria de tamaño n proveniente de una densidad f. Sea $\hat{\theta}$ el estimador máximo verosímil de θ y sea $h : \Theta \to \mathbb{R}$ una función contínua. Luego, se tiene que

$$\frac{(h(\theta) - h(\hat{\theta}))}{\sqrt{\nu(\theta)}} \to N(0,1)$$

donde $\nu(\theta) = \frac{(h'(\theta))^2}{\mathbb{E}\left[-\frac{\partial^2}{\partial \theta^2} \ell(\theta) \Big|_{\theta = \hat{\theta}} \right]}$

Algunos resultados pt.II

Distribución asintótica

Considere $X_1, \ldots, X_n$ una muestra aleatoria de tamaño n proveniente de una densidad f. Sea $\hat{\theta}$ el estimador máximo verosímil de θ y sea $h : \Theta \to \mathbb{R}$ una función continua. Luego, se tiene que

$$(h(\theta) - h(\hat{\theta})) \to N(0, \nu(\theta))$$

donde $\nu(\theta) = \frac{(h'(\theta))^2}{\mathbb{E}\left[-\frac{\partial^2}{\partial\theta^2}\ell(\theta)\Big|_{\theta=\hat{\theta}}\right]}$

Algunos comentarios:

1. Esto nos permite considerar una aproximación normal a funciones de los parámetros.
2. Estas aproximaciones valen para cualquier secuencia de variables aleatorias donde la segunda derivada de la log-verosimilitud existe.
3. La cantidad $\nu(\theta)$ se conoce como la *varianza asintótica*.

Intervalo de confianza asintótico

Notamos que el estimador natural de $\nu(\hat{\theta})$ es

$$\nu(\hat{\theta}) = \frac{(h'(\hat{\theta}))^2}{-\left.\frac{\partial^2}{\partial\theta^2}\ell(\theta)\right|_{\theta=\hat{\theta}}}$$

Considerando esto, la siguiente cantidad

$$\left(\frac{h(\theta) - h(\hat{\theta})}{\sqrt{\nu(\hat{\theta})}}\right)$$

es una cantidad pivotal para $h(\theta)$. Luego, un intervalo de confianza asintótico para $h(\theta)$ es

Intervalo de confianza asintótico

$$IC(h(\theta))_\gamma = \left[h(\hat{\theta}) - z_{1-\alpha/2}\sqrt{\nu(\hat{\theta})}; h(\hat{\theta}) + z_{1-\alpha/2}\sqrt{\nu(\hat{\theta})}\right]$$

Ejemplo I

Considere $X_1, \ldots, X_n$ una muestra aleatoria provenientes de una distribución

$$f(x|p) = p^x(1-p)^{1-x}, \quad x \in \{0,1\}, \quad p \in (0,1).$$

Se desea calcular un intervalo de confianza para p.

Ejemplo I

Recordamos la verosimilitud:

$$L(p, \underline{X}) = \prod_{i=1}^{N} p^{x_i}(1-p)^{1-x_i} = p^{\sum_{i=1}^{N} x_i}(1-p)^{N-\sum_{i=1}^{N} x_i}$$

y la log-verosimilitud

$$\ell(p, \underline{X}) = \log L(p, \underline{X}) = \log p \sum_{i=1}^{N} x_i + \log(1-p)(N - \sum_{i=1}^{N} x_i)$$

Ejemplo I: Estimador MV

Obtenemos la primera derivada de $\ell(\theta, \underline{X})$:

$$\frac{\partial \ell(p, \underline{X})}{\partial p} = \frac{\sum_{i=1}^{n} x_i}{p} - \frac{N - \sum_{i=1}^{n} x_i}{1-p}$$

obteniendo el estimador máximo verosímil:

$$\hat{p}_{MV} = \bar{X}$$

Ejemplo I: Varianza Asintótica

Notamos que $h(p) = p$, luego $h'(p) = 1$. Además se tiene que:

$$\frac{\partial^2 \ell(p, \underline{X})}{\partial p^2} = \frac{\sum_{i=1}^{n} x_i}{p^2} + \frac{N - \sum_{i=1}^{n} x_i}{(1-p)^2}$$

y en consecuencia:

$$-\left.\frac{\partial^2 \ell(p, \underline{X})}{\partial p^2}\right|_{=\hat{p}_{MV}} = \frac{N}{\hat{p}(1-\hat{p})}$$

Luego, la varianza asintótica $\nu(\theta) = \frac{\hat{p}(1-\hat{p})}{N}$.

Ejemplo I: Intervalo de confianza asintótico

Luego, un intervalo de confianza asintótico para p es

$$IC(p)_\gamma = \left[\hat{p}_{MV} - z_{1-\alpha/2}\sqrt{\frac{\hat{p}_{MV}(1-\hat{p}_{MV})}{n}}; \hat{p}_{MV} + z_{1-\alpha/2}\sqrt{\frac{\hat{p}_{MV}(1-\hat{p}_{MV})}{n}}\right]$$

Ejemplo II

Supongamos que se desea crear un intervalo de confianza para el *odds ratio*

$$h(p) = \frac{p}{1-p}$$

Notamos entonces que

$$h'(p) = \frac{1}{(1-p)^2}$$

Luego se tiene que

$$\nu(\hat{p}) = \frac{\hat{p}}{n(1-\hat{p})^3}$$

En consecuencia se tiene que

$$IC\left(\frac{p}{1-p}\right)_\gamma = \left[\frac{\hat{p}_{MV}}{1-\hat{p}_{MV}} - z_{1-\alpha/2}\sqrt{\frac{\hat{p}_{MV}}{n(1-\hat{p}_{MV})^3}}; \frac{\hat{p}_{MV}}{1-\hat{p}_{MV}} + z_{1-\alpha/2}\sqrt{\frac{\hat{p}_{MV}}{n(1-\hat{p}_{MV})^3}}\right]$$

Ejemplo III

Sea $X_1, \ldots, X_N$ una muestra aleatoria proveniente de una distrubición Poisson de parámetro λ. Encuentre un intervalo de confianza para λ

Recordamos la verosimilitud:

$$L(\lambda, \underline{X}) = \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} = \lambda^{\sum_{i=1}^{n} x_i} e^{-n\lambda} \prod_{i=1}^{n} \frac{1}{x_i!}$$

y la log-verosimilitud

$$\ell(\lambda, \underline{X}) = \sum_{i=1}^{n} x_i \log \lambda - n\lambda + \sum_{i=1}^{n} \log \frac{1}{x_i!}$$

Ejemplo III: Derivadas y varianza asintótica

$$\frac{\partial}{\partial \lambda}\ell(\lambda, \underline{X}) = \frac{\sum_{i=1}^{n} x_i}{\lambda} - n$$

$$-\frac{\partial^2}{\partial \lambda^2}\ell(\lambda, \underline{X}) = \frac{\sum_{i=1}^{n} x_i}{\lambda^2}$$

En consecuencia se tiene que

$$\nu(\hat{\lambda}) = \frac{\hat{\lambda}}{n}$$

$$IC(\lambda)_\gamma = \left[\hat{\lambda} - z_{1-\alpha/2}\sqrt{\frac{\hat{\lambda}}{n}}; \hat{\lambda} + z_{1-\alpha/2}\sqrt{\frac{\hat{\lambda}}{n}}\right]$$

Ejemplo IV:

Sea $X_1, \ldots, X_N$ una muestra aleatoria proveniente de una distrubición exponencial de parámetro λ. Encuentre un intervalo de confianza para $\frac{\lambda^3}{\sigma^3/2}$ donde σ es conocido.

Ejemplo:

El directorio de una compañía productora de alimentos decidió hacer una investigación de mercado de modo que si el 40 % o más de los consumidores manifestaba la intención de comprar en producto, entonces la compañía consideraría comercilizarlo en el resto del pais. De 764 personas encuestadas 193 manifestaron que estarían dispuesto a comprar el producto.

¿Los datos de esta muestra sugieren apoyar la idea de comercializar el nuevo producto en un área más extensa? Use una confianza del 99 % para elaborar su respuesta.

Solución

La m.a. es de tamaño $n = 764$.

La v.a. es X :=Intención de compra de los consumidores clasificada como *si* o *no*.
Aquí el parámetro p corresponde a la verdadera proporción de consumidores que está dispuesto a comprar el nuevo producto y su estimación resulta ser
$\hat{p} = 193/764 = 0,253$.

Solución

La m.a. es de tamaño $n = 764$.

La v.a. es X :=Intención de compra de los consumidores clasificada como *si* o *no*. Aquí el parámetro p corresponde a la verdadera proporción de consumidores que está dispuesto a comprar el nuevo producto y su estimación resulta ser $\hat{p} = 193/764 = 0,253$.

Entonces el intervalo resulta ser,

$$\begin{aligned} IC(p) &= 0,253 \pm \sqrt{\frac{0,253(1-0,253)}{764}} Z_{0,995} \\ &= 0,253 \pm 0,0157(2,576) = 0,253 \pm 0,0405 \\ &= [0,212; 0,294] \end{aligned}$$

Entonces, a partir de la muestra y con un 99 % confianza, estimamos que la proporción de consumidores dispuestos a comprar el nuevo producto alimenticio está entre 21,2 % y 29,4 %, lejos del 40 % o más requerido para comercializar el nuevo producto en todo el país.

Caso de varianza desconocida

Si $X_1, \ldots, X_n$ es una muestra aleatoria de una población con distribución normal de media μ y varianza σ^2 desconocida y se quiere estimar μ mediante un intervalo de confianza. En este caso, la varianza poblacional σ^2 desconocida es reemplazada por su estimador insesgado $S^2 = \sum(X_i - \hat{X})^2/(n-1)$ y la cantidad pivotal que se emplea es,

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

De esta forma el intervalo de confianza del $(1-\alpha)100\,\%$ para estimar μ cuando σ^2 es desconocida resulta ser,

$$\begin{aligned} IC_{\alpha/2}(\mu) &= \left[\overline{X} - \frac{S}{\sqrt{n}} t_{1-\alpha/2}(n-1); \overline{X} + \frac{S}{\sqrt{n}} t_{1-\alpha/2}(n-1)\right] \\ &= \left[\bar{X} \pm \frac{S}{\sqrt{n}} t_{1-\alpha/2}(n-1)\right] \end{aligned}$$

Ejemplo

Los siguientes datos corresponden a la rentabilidad en porcentaje durante el último mes de 10 acciones escogidas aleatoriamente. Asuma que la rentabilidad es bien modelada por una distribución normal.

0,8	-0,3	-0,6	0,5	0,6	0,1	0,3	-0,5	-0,1	-0,2

a) Estime el valor de la rentabilidad media de todas las acciones con un intervalo de confianza del 95 %.

b) Ciertos expertos habían pronosticado que la rentabilidad de las acciones sería de $0,3\,\%$ en ese periodo de tiempo. ¿Los datos recolectados apoyan esa idea?.

c) Suponga que ahora el tamaño de la muestra aumenta a $n = 30$ y que la media y la desviación estándar de la muestra resultan ser las mismas. Si la confianza del intervalo también se mantiene, ¿cambia la conclusión respecto a la afirmación de los expertos?. Explique o comente su resultado.

Solución a) X : Rentabilidad de las acciones de la Bolsa de Valores. $X \sim N(\mu; \sigma^2)$ con μ y σ^2 desconocidos.

Del enunciado del problema se tiene que $n = 10$, la media muestral es $\bar{X} = 0,06$, la desviación estándar de la muestra es $S = 0,479$ y $\alpha = 0,05$. Así y el cuantil de la t-student es $t_{1-\alpha/2}(n-1) = t_{0,975}(9) = 2,262$. Entonces, el intervalo de confianza del 95 % para estimar la rentabilidad media de las acciones transadas en esa Bolsa de Valores es:

$$
\begin{aligned}
IC(\mu) &= \left[\bar{X} - \frac{S}{\sqrt{n}} t_{1-\alpha/2}(n-1) \ ; \ \bar{X} + \frac{S}{\sqrt{n}} t_{1-\alpha/2}(n-1)\right] \\
IC(\mu) &= \left[0,06 - 0,15(2,262) \ ; \ 0,06 + 0,15(2,262)\right] \\
IC(\mu) &\approx \left[0,06 - 0,34 \ ; \ 0,06 + 0,34\right] \\
&\approx \left[-0,28 \ ; \ 0,40\right]
\end{aligned}
$$

B) El intervalo estimado del 95 % de confianza para la rentabilidad media es $ic(\mu) = \big[-0,28 \ ; \ 0,40 \big]$ y contiene el valor pronosticado de $0,3\,\%$. Entonces, la información contenida en la muestra de tamaño $n = 10$ no descarta el pronóstico de los expertos.

B) El intervalo estimado del 95 % de confianza para la rentabilidad media es $ic(\mu) = \big[-0,28\ ;\ 0,40\big]$ y contiene el valor pronosticado de $0,3\,\%$. Entonces, la información contenida en la muestra de tamaño $n = 10$ no descarta el pronóstico de los expertos.

C) Con $n = 30$ se tiene que $t_{1-\alpha/2}(n-1) = t_{0,975}(29) = 2,045$, el intervalo de confianza del 95 % para estimar la rentabilidad media de las acciones transadas en esa Bolsa de Valores es,

$$\begin{aligned} IC(\mu) &= \big[0,06 - 0,0874(2,045)\ ;\ 0,06 + 0,0874(2,045)\big] \\ &\approx \big[0,06 - 0,18\ ;\ 0,06 + 0,08743251(2,045)\big] \\ &\approx \big[-0,11\ ;\ 0,24\big] \end{aligned}$$

Ahora, el intervalo estimado del 95 % de confianza para la rentabilidad media es $IC(\mu) = \big[-0,11\ ;\ 0,24\big]$ y no contiene el valor pronosticado de $0,3\,\%$. Entonces, los datos de la muestra de tamaño $n = 30$ ahora permiten descartar o rechazar el pronóstico de los expertos.

Intervalo para estimar una varianza σ^2 en caso normal con media μ desconocida

Sea $X_1, \ldots, X_n$ una m.a. de una distribución normal de media y varianza desconocidos, y que el interés recae en estimar mediante un intervalo de confianza para la varianza poblacional σ^2.

Intervalo para estimar una varianza σ^2 en caso normal con media μ desconocida

Sea $X_1, \ldots, X_n$ una m.a. de una distribución normal de media y varianza desconocidos, y que el interés recae en estimar mediante un intervalo de confianza para la varianza poblacional σ^2.

El estimador insesgado de esta varianza es $S^2 = \sum (X_i - \bar{X})^2 / (n-1)$ y la cantidad pivotal adecuada es

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

Intervalo para estimar una varianza σ^2 en caso normal con media μ desconocida

Sea $X_1, \ldots, X_n$ una m.a. de una distribución normal de media y varianza desconocidos, y que el interés recae en estimar mediante un intervalo de confianza para la varianza poblacional σ^2.

El estimador insesgado de esta varianza es $S^2 = \sum (X_i - \bar{X})^2 / (n-1)$ y la cantidad pivotal adecuada es

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

Escogiendo colas simétricas de modo que queden probabilidades $\alpha/2$ a la derecha e izquierda de la distribución χ^2_{n-1} se obtiene que

$$\begin{aligned} P\left(\chi^2_{\alpha/2}(n-1) \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi^2_{1-\alpha/2}(n-1)\right) &= 1-\alpha \\ P\left(\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)}\right) &= 1-\alpha \end{aligned}$$

Así, un intervalo de confianza para estimar la varianza σ^2 es,

$$IC(\sigma^2) = \left[\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)} \quad ; \quad \frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)} \right]$$

Además, un intervalo de confianza *aproximado* para estimar la desviación estándar σ es,

$$IC(\sigma) = \left[\sqrt{\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)}} \quad ; \quad \sqrt{\frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)}} \right]$$

Ejemplo:

La Compañía de Aceros del Sur CAS revisa periódicamente su producción para evaluar el cumplimiento de las normas. En particular se encuentra vigilando la producción de planchas metálicas de 5 mm de espesor realizada por una máquina. Para este producto no se tolera una desviación estándar superior a los 0.02 mm. Un muestra de 22 planchas producidas recientemente revela los siguientes espesores en mm.

4,96	5,02	5,00	5,04	5,07	4,99	4,98	4,97	5,05	4,96	5,02
5,01	5,00	4,97	4,96	5,04	4,96	5,07	4,99	4,99	4,97	5,05

¿Que se puede concluir de esta muestra respecto a la variabilidad en el espesor de las planchas de 5 mm. producidas por esa máquina?

Solución

Aquí, $n = 22$, $\bar{X} = 5,00$, $S^2 = 0,001356$.

Escogiendo $\alpha = 0,01$ se tiene que,

$$\begin{aligned} IC(\sigma) &= \left[\sqrt{\frac{(22-1)0,001356}{41,40}} \quad ; \quad \sqrt{\frac{(22-1)0,001356}{8,03}}\right] \\ &= [0,026 \quad ; \quad 0,060] \end{aligned}$$

Escogiendo $\alpha = 0,10$ se tiene que,

$$\begin{aligned} IC(\sigma) &= \left[\sqrt{\frac{(22-1)0,001356}{32,67}} \quad ; \quad \sqrt{\frac{(22-1)0,001356}{11,59}}\right] \\ &= [0,030 \quad ; \quad 0,050] \end{aligned}$$

Entonces, la estimación de la desviación estándar del espesor de las planchas de 5,00 mm., para confianzas iguales o inferiores a 99 %, está por sobre los 0,02 mm. que establece la norma. En consecuencia la muestra reciente sugiere revisar la máquina que produce esas planchas.

Ejercicios

Un artículo informa que se usó una muestra de tamaño 16 como base para calcular un intervalo de confianza de 95 % para la media μ de una población normal con varianza σ^2 conocida. El intervalo resultante fue $[0,34; 4,26]$. Pero, usted piensa que es mejor usar un nivel de confianza del 99 %. ¿Cuáles serían los límites de un intervalo de 99 % de confianza para estimar μ?

Un artículo informa que se usó una muestra de tamaño 16 como base para calcular un intervalo de confianza de 95 % para la media μ de una población normal con varianza σ^2 conocida. El intervalo resultante fue $[0,34; 4,26]$. Pero, usted piensa que es mejor usar un nivel de confianza del 99 %. ¿Cuáles serían los límites de un intervalo de 99 % de confianza para estimar μ?

Respuesta: $IC(\mu) = [-0,272; 2,576]$.

Para estimar la proporción de consumidores que prefieren un determinado refresco, por medio de un intervalo de confianza, se ha tomado una muestra al azar de 1075 consumidores, entre los que se han encontrado a 516 que lo prefieren.

a) Construya un intervalo de confianza del 95 % para la proporción de la población que prefiere el refresco.

b) Dado el costo de encuestar a 1075 personas, al repetir el estudio el año siguiente, el encargado del área de estudios decide reducir la muestra a 500. En este caso se encontraron 240 consumidores que prefieren este refresco. Calcule el nuevo intervalo de confianza al 95 % y compare con el obtenido en la parte a) indicando que se ganó y qué se perdió con la modificación.

Ejercicio:

La presión (en Pa) que soporta una placa de cierto material antes de agrietarse, es una variable aleatoria con distribución normal. En una muestra de 15 placas, la presión soportada por cada una de ellas fue la siguiente

30 40 35 38 37 37 38 42 31 34 35 36 37 34 32

Con base en esos datos:

a) Estime con un 90 % de confianza la presión media soportada por las placas de este material.

Respuesta: $IC(\mu) = [34,25; 37,21]$

b) ¿Cuántas placas deben seleccionarse en una muestra, si se desea estimar con un 95 % de confianza y con un margen de error no superior a 0,04, la proporción de placas que soportan una presión de al menos 35 (Pa).

Ejercicio:

La presión (en Pa) que soporta una placa de cierto material antes de agrietarse, es una variable aleatoria con distribución normal. En una muestra de 15 placas, la presión soportada por cada una de ellas fue la siguiente

30 40 35 38 37 37 38 42 31 34 35 36 37 34 32

Con base en esos datos:

a) Estime con un 90 % de confianza la presión media soportada por las placas de este material.

Respuesta: $IC(\mu) = [34,25; 37,21]$

b) ¿Cuántas placas deben seleccionarse en una muestra, si se desea estimar con un 95 % de confianza y con un margen de error no superior a 0,04, la proporción de placas que soportan una presión de al menos 35 (Pa).Respuesta: $n = 531$.