Sea $\underline{X} = \{X_1, X_2 ... X_n\}$ una muestra aleatoria simple de una población con distribución normal, con media μ y varianza σ^2. y sean $\hat{\sigma}_0^2$, $\hat{\sigma}_1^2$ y $\hat{\sigma}_2^2$ tres estimadores de la varianza poblacional σ^2.

Donde: $\hat{\sigma}_0^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$, $\hat{\sigma}_1^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{X})^2$ y $\hat{\sigma}_2^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{X})^2$

Determinar para cada uno de ellos

a) su media
b) su varianza
c) si es el estimador eficiente de σ^2
d) si es consistente
e) si es suficiente

Resolución:

Antes de encontrar la media y la varianza de cada estimador conviene ver que con un poco de álgebra se puede demostrar que: $\sum_{i=1}^{n}(x_i - \bar{X})^2 = \sum_{i=1}^{n}(x_i - \mu)^2 - (\bar{X} - \mu)^2$. y por otro lado, recordar que:

1. $V(x_i) = E\left[(x_i - \mu)^2\right] = \sigma^2$

2. $V(\bar{X}) = E\left[(\bar{X} - \mu_{\bar{X}})^2\right] = E\left[(\bar{X} - \mu)^2\right] = \frac{\sigma^2}{n}$

3. Si la población tiene distribución normal entonces las variables aleatorias

$$\sum_{i=1}^{n}\left(\frac{x_i - \mu}{\sigma}\right)^2 \sim \chi_n^2 , \quad \text{y} \quad \left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}\right)^2 \sim \chi_1^2$$

tienen distribución Jí-cuadrada con *n* y *1* grados de libertad respectivamente.

4. como las dos variables anteriores tiene distribución Ji-cuadrada y la varianza de un variable aleatoria Ji-cuadrada con ν grados de libertad es 2ν, entonces sus varianzas son:

$$V\left[\sum_{i=1}^{n}\left(\frac{x_i - \mu}{\sigma}\right)^2\right] = 2n , \quad \text{y} \quad V\left[\left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}\right)^2\right] = 2$$

a) las medias:

Para $\hat{\sigma}_0^2$ se tiene:

$$E\left[\hat{\sigma}_0^2\right] = E\left[\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2\right] = \frac{1}{n}\sum_{i=1}^{n}E\left[(x_i-\mu)^2\right] = \frac{1}{n}\sum_{i=1}^{n}\sigma^2 = \frac{1}{n}n\sigma^2 = \sigma^2$$

por lo tanto es insesgado.

Para $\hat{\sigma}_1^2$ se tiene:

$$\hat{\sigma}_1^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{X})^2 = \frac{1}{n}\left[\sum_{i=1}^{n}(x_i-\mu)^2 - n(\overline{X}-\mu)^2\right]$$

$$E\left[\hat{\sigma}_1^2\right] = \frac{1}{n}\left[\sum_{i=1}^{n}E\left[(x_i-\mu)^2\right] - nE\left[(\overline{X}-\mu)^2\right]\right] = \frac{1}{n}\left[\sum_{i=1}^{n}\sigma^2 - n\frac{\sigma^2}{n}\right] = \frac{1}{n}\left[n\sigma^2-\sigma^2\right] = \frac{(n-1)}{n}\sigma^2$$

por lo tanto, es sesgado.

Para $\hat{\sigma}_2^2$ se tiene:

$$\hat{\sigma}_2^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{X})^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n}(x_i-\mu)^2 - n(\overline{X}-\mu)^2\right]$$

$$E\left[\hat{\sigma}_2^2\right] = \frac{1}{n-1}\left[\sum_{i=1}^{n}E\left[(x_i-\mu)^2\right] - nE\left[(\overline{X}-\mu)^2\right]\right] = \frac{1}{n-1}\left[\sum_{i=1}^{n}\sigma^2 - n\frac{\sigma^2}{n}\right] = \frac{1}{n-1}\left[n\sigma^2-\sigma^2\right] = \sigma^2$$

por lo tanto, es insesgado.

b) las varianzas:

Para $\hat{\sigma}_0^2$ se tiene:

$$V\left[\hat{\sigma}_0^2\right] = V\left[\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2\right] = \frac{1}{n^2}V\left[\sum_{i=1}^{n}(x_i-\mu)^2\right]$$

Del recordatorio 3 y 4 :

$$V\left[\hat{\sigma}_0^2\right] = \frac{1}{n^2}V\left[\sum_{i=1}^{n}\sigma^2\left(\frac{x_i-\mu}{\sigma}\right)^2\right] = \frac{\sigma^4}{n^2}V\left[\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)^2\right] = \frac{\sigma^4}{n^2}2n = \frac{2\sigma^4}{n}$$

Para $\hat{\sigma}_1^2$ se tiene:

$$V\left[\hat{\sigma}_1^2\right] = V\left\{\frac{1}{n}\left[\sum_{i=1}^{n}(x_i-\mu)^2 - n(\overline{X}-\mu)^2\right]\right\}$$

$$V\left[\hat{\sigma}_1^2\right] = \frac{1}{n^2}\left\{V\left[\sum_{i=1}^{n}(x_i-\mu)^2\right] - n^2 V\left[(\overline{X}-\mu)^2\right]\right\}$$

Del recordatorio 3 y 4 :

$$V\left[\hat{\sigma}_1^2\right] = \frac{1}{n^2}\left\{V\left[\sum_{i=1}^{n}\sigma^2\left(\frac{x_i-\mu}{\sigma}\right)^2\right] - n^2 V\left[\frac{\sigma^2}{n}\left(\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}\right)^2\right]\right\}$$

$$= \frac{1}{n^2}\left\{\sigma^4 V\left[\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)^2\right] - n^2\frac{\sigma^4}{n^2} V\left[\left(\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}\right)^2\right]\right\}$$

$$= \frac{1}{n^2}\left\{\sigma^4(2n) - \sigma^4(2)\right\} = \frac{2\sigma^4(n-1)}{n^2}$$

Para $\hat{\sigma}_2^2$ se tiene:

$$V\left[\hat{\sigma}_2^2\right] = V\left\{\frac{1}{n-1}\left[\sum_{i=1}^{n}(x_i-\mu)^2 - n(\overline{X}-\mu)^2\right]\right\}$$

$$E\left[\hat{\sigma}_1^2\right] = \frac{1}{(n-1)^2}\left\{V\left[\sum_{i=1}^{n}(x_i-\mu)^2\right] - n^2 V\left[(\overline{X}-\mu)^2\right]\right\}$$

Procediendo de manera similar al caso anterior se tiene:

$$V\left[\hat{\sigma}_2^2\right] = \frac{1}{(n-1)^2}\left\{\sigma^4(2n) - \sigma^4(2)\right\} = \frac{2\sigma^4(n-1)}{(n-1)^2} = \frac{2\sigma^4}{(n-1)}$$

c) la eficiencia

Para determinar si alguno de los tres estimadores es el estimador eficiente de σ^2 debe encontrarse la cota inferior de Cramér-Rao, Si la varianza de alguno de los tres estimadores coincide con la cota inferior de Cramér-Rao, entonces el estimador correspondiente es el estimador eficiente de σ^2.

La cota inferior de Cramer Rao es:

$$I(\theta) = \frac{1}{nE\left\{\left[\frac{\partial \ln f(x,\theta)}{\partial\theta}\right]^2\right\}} \quad \text{o bien} \quad I(\theta) = \frac{1}{-nE\left[\frac{\partial^2 \ln f(x,\theta)}{\partial\theta^2}\right]}$$

(Nota: bajo ciertas condiciones generales, ambas expresiones son equivalentes)

Considerando $f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)$ y que se busca un límite inferior para la varianza del estimador de σ^2, al realizar los cálculos se encuentra que:

$$I(\sigma^2) = \frac{2\sigma 4}{n}$$

Al compararla con las varianzas de los estimadores obtenidas en el inciso (b) se concluye que el estimador eficiente de σ^2 es $\hat{\sigma}_0^2$, sin embargo, este estimador tiene la desventaja de ser función de la varianza poblacional μ, que generalmente es desconocida. Por tanto, se tendrá que recurrir a uno de los otros dos estimadores propuestos. Para evaluar a los otros dos estimadores, podemos calcular la eficiencia relativa de $\hat{\sigma}_2^2$ con respecto a $\hat{\sigma}_1^2$, para ello, debemos encontrar antes el error cuadrático medio de cada uno de los dos estimadores.

Para $\hat{\sigma}_1^2$

$$\begin{aligned}
ECM\left[\hat{\sigma}_1^2\right] &= Var\left[\hat{\sigma}_1^2\right] + Sesgo^2\left[\hat{\sigma}_1^2\right] \\
&= \frac{2\sigma^4(n-1)}{n^2} + \left[\sigma^2 - E(\sigma^2)\right]^2 \\
&= \frac{2\sigma^4(n-1)}{n^2} + \left[\sigma^2 - \frac{(n-1)}{n}\sigma^2\right]^2 \\
&= \frac{2\sigma^4(n-1)}{n^2} + \left[\frac{\sigma^2}{n}\right]^2 \\
&= \frac{2\sigma^4(n-1)}{n^2} + \frac{\sigma^4}{n^2} \\
&= \frac{\sigma^4(2n-2+1)}{n^2}
\end{aligned}$$

$$ECM\left[\hat{\sigma}_1^2\right] = \frac{\sigma^4(2n-1)}{n^2}$$

Para $\hat{\sigma}_2^2$ como es insesgado, $ECM\left[\hat{\sigma}_2^2\right] = V\left[\hat{\sigma}_2^2\right] = \frac{2\sigma^4}{(n-1)}$

y finalmente, la eficiencia relativa de $\hat{\sigma}_2^2$ con respecto a $\hat{\sigma}_1^2$ es:

$$\eta = \frac{ECM(\sigma_1^2)}{ECM(\sigma_2^2)} \qquad \eta = \frac{\sigma^4(2n-1)}{n^2} \Big/ \frac{2\sigma^4}{(n-1)} = \frac{(2n-1)(n-1)}{2n^2}$$

Debe observarse que para utilizar $\hat{\sigma}_2^2$ el tamaño de muestra debe ser mayor o igual a 2, pues de lo contrario, si n=1 su varianza queda indefinida.

Al analizar η resulta que para $n \geq 2$, $\eta > 1$, por lo que $\hat{\sigma}_2^2$ es mejor estimador que $\hat{\sigma}_1^2$, y además $\hat{\sigma}_2^2$ tiene la ventaja de ser insesgado.

d) la consistencia:

$$\lim_{n\to\infty} ECM\left[\hat{\sigma}_0^2\right] = \lim_{n\to\infty} V\left[\hat{\sigma}_0^2\right] = \lim_{n\to\infty}\left[\frac{2\sigma^4}{n}\right] = 0$$

$$\lim_{n\to\infty} ECM\left[\hat{\sigma}_1^2\right] = \lim_{n\to\infty}\left[\frac{\sigma^4(2n-1)}{n^2}\right] = 0$$

$$\lim_{n\to\infty} ECM\left[\hat{\sigma}_2^2\right] = \lim_{n\to\infty}\left[\frac{2\sigma^4}{(n-1)}\right] = 0$$

Por lo tanto, los tres estimadores son consistentes.

e) la suficiencia

Los tres estimadores son suficientes, dado que utilizan toda la información que proporciona la muestra, sin embargo deberá demostrarse mediante el teorema de factorización de Neyman-Pearson.

CONCLUSIÓN

Como los tres estimadores son consistentes y suficientes, los criterios que prevalecen son el sesgo y la eficiencia. El mejor estimador es $\hat{\sigma}_0^2$, pero no puede usarse cuando se desconoce μ, de los dos restantes el mejor es $\hat{\sigma}_2^2$, ya que $\hat{\sigma}_1^2$ es sesgado y menos eficiente que $\hat{\sigma}_2^2$