Distribución de muestreo de la estadística $\frac{(n-1)S^2}{\sigma^2}$

Uno de los parámetros de más interés es la varianza σ^2 de una población normal, se sabe que el estimador insesgado de σ^2 es

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \overline{X}\right)^2$$

por lo que es necesario establecer cuál es la distribución de muestreo del estadístico S^2 o encontrar un estadístico de apoyo cuya distribución sea conocida.

Para encontrar un estadístico de apoyo para efectuar inferencias sobre σ^2 debemos suponer que las observaciones x_i provienen de una población con distribución normal con media μ y varianza σ^2 y por lo tanto $\overline{X}$ también tiene distribución normal con media μ y varianza σ^2/n.

Para cada observación x_i de la muestra, el estadístico $\frac{x_i - \mu}{\sigma}$ tiene distribución normal estándar y en consecuencia, el estadístico: G=$\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{\sigma^2}$ tiene distribución Ji-cuadrada con n grados de libertad.

Para simplificar las operaciones algebraicas multiplicamos ambos lados de la ecuación por σ^2, así:

$\sigma^2 G = \sum_{i=1}^{n}(x_i-\mu)^2$ restando y sumando la misma cantidad $\overline{X}$ dentro del paréntesis:

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

$$\sigma^2 G = \sum_{i=1}^{n}\left(x_i-\overline{X}\right)^2 + n(\overline{X}-\mu)^2 + 2(\overline{X}-\mu)\sum_{i=1}^{n}(x_i-\overline{X}),$$

pero $\sum_{i=1}^{n}(x_i-\overline{X}) = 0$ de donde:

$$\sigma^2 G = \sum_{i=1}^{n}\left(x_i-\overline{X}\right)^2 + n(\overline{X}-\mu)^2$$

Dividiendo ambos lados de la ecuación entre σ^2 queda:

$$G = \sum_{i=1}^{n} \frac{(x_i - \overline{X})^2}{\sigma^2} + \frac{n(\overline{X} - \mu)^2}{\sigma^2}$$

Reacomodando algunos términos:

$$G = \sum_{i=1}^{n} \left(\frac{x_i - \overline{X}}{\sigma^2} \right)^2 + \left(\frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \right)^2$$

Como se vio antes, G (el lado izquierdo de la ecuación) es una variable aleatoria con distribución Ji-cuadrada con n grados de libertad; el segundo termino del lado derecho es el cuadrado de una VA normal estándar, por lo que tiene distribución Ji-cuadrada con un grado de libertad. Se deduce del teorema de aditividad de la distribución Ji-cuadrada que el primer término del lado derecho de la ecuación tiene necesariamente distribución Ji-cuadrada con (n-1) grados de libertad, es decir:

$$\sum_{i=1}^{n} \left(\frac{x_i - \overline{X}}{\sigma} \right)^2 \sim \chi^2_{n-1}$$

Ahora, multiplicando y dividiendo este último estadístico por (n-1) y sacando de la sumatoria el parámetro σ^2 tenemos:

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

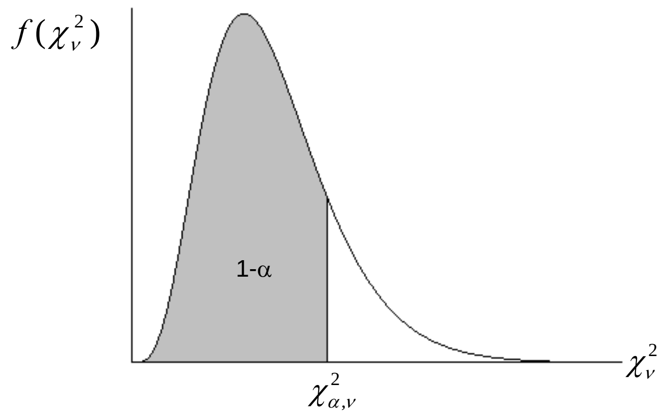
pero antes se definió $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{X})^2$, sustituyendo: $\sum_{i=1}^{n} \left(\frac{x_i - \overline{X}}{\sigma} \right)^2 = \frac{(n-1)S^2}{\sigma^2}$

Por lo tanto, la variable aleatoria $\frac{(n-1)S^2}{\sigma^2}$ tiene distribución Ji-cuadrada con υ=(n-1) grados de libertad.

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

De lo anterior se tiene que:

$$P(S^2 < s^2) = P\left(\frac{(n-1)S^2}{\sigma^2} < \chi^2_{\alpha,\nu}\right) = 1-\alpha \quad \text{para toda } s^2 \in S^2$$



Es conveniente aclarar que la variable aleatoria $\frac{(n-1)S^2}{\sigma^2}$ es sumamente sensible a la suposición de normalidad de la población de la que se extrae la muestra, si la población en estudio no tiene distribución normal, los resultados aquí obtenidos no serán válidos.

Ejemplo: Un fabricante utiliza un pieza de precisión en la maquinaria que produce, la longitud de esta pieza debe ser de 5 ± 0.05 cm. Ha establecido con su proveedor un contrato en el que se indica que de cada lote se tomará una muestra aleatoria de 25 piezas y si la desviación estándar de la muestra resulta mayor a 0.05cm devolverá todo el lote. Suponiendo que la longitud de la pieza se distribuye normalmente con desviación estándar de 0.04cm ¿Qué porcentaje de lotes serán rechazados? Suponer lotes de gran tamaño en relación con el tamaño de la muestra.

Solución: La probabilidad de rechazar un lote cualquiera está dada por:

$$P\left(S^2 > (0.05)^2\right) = P\left(\frac{(n-1)S^2}{\sigma^2} > \frac{(25-1)(0.05)^2}{(0.04)^2}\right) = P\left(\chi^2_{24} > 37.5\right)$$

Interpolando en la tabla de la variable aleatoria Ji-cuadrada con 24 grados de libertad se encuentra que:

$$P\left(S^2 > (0.05)^2\right) = P\left(\chi^2_{24} > 37.5\right) = 1 - 0.9592 = 0.0408$$

Utilizando la interpretación frecuentista de la probabilidad se concluye que, bajo las condiciones establecidas, se rechazarán 4.08 % de los lotes.

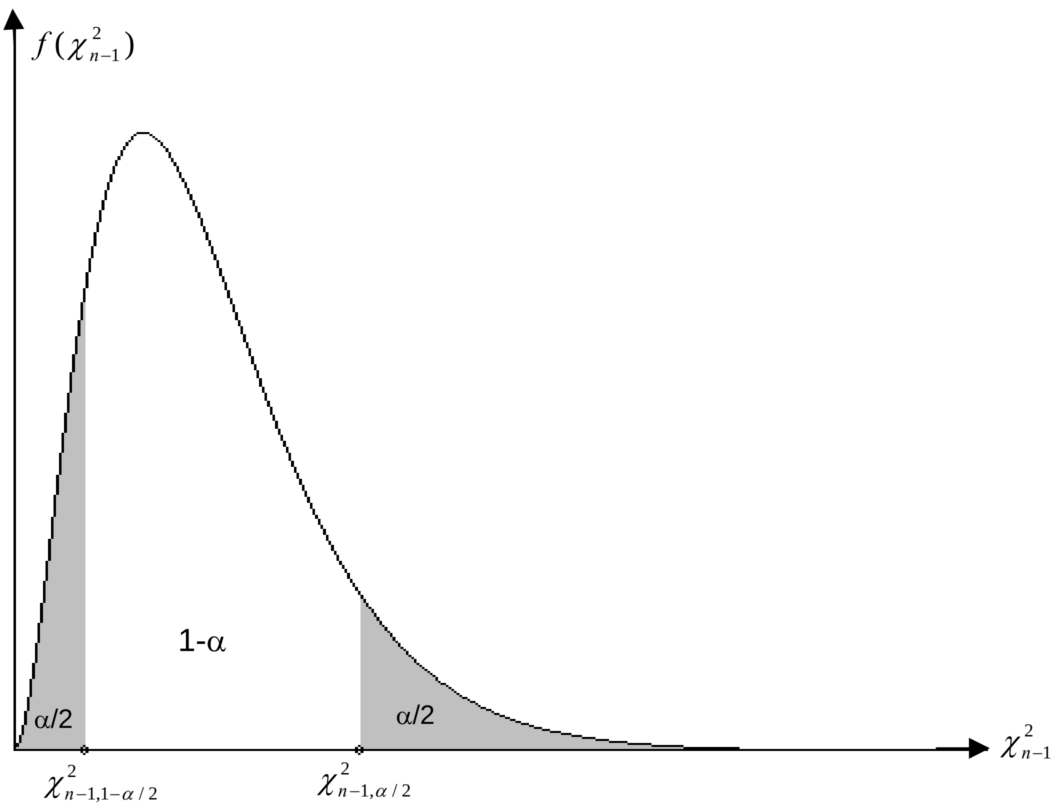
ESTIMACIÓN POR INTERVALOS PARA LA VARIANZA DE UNA POBLACIÓN CON DISTRIBUCIÓN NORMAL

El estimador insesgado de la varianza es: $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{X})^2$

Además el estadístico $\frac{(n-1)S^2}{\sigma^2}$ tiene distribución Ji-cuadrada con n-1 grados de libertad, por:

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

Para construir un intervalo de confianza bilateral de (1-α)100% , a partir de la gráfica de la distribución Ji-cuadrada se tiene:



$$P(\chi^2_{n-1,1-\alpha/2} < \chi^2_{n-1} < \chi^2_{n-1,\alpha/2}) = 1-\alpha$$

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

y utilizando el método del pivote para determinar el intervalo de confianza se tiene:

$$\frac{(n-1)S^2}{\chi^2_{n-1,\alpha/2}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{n-1,1-\alpha/2}}$$

con una confianza de (1-α)100%