

DISTRIBUCION JI-CUADRADA (X^2)

En realidad la distribución ji-cuadrada es la distribución muestral de s^2 . O sea que si se extraen todas las muestras posibles de una población normal y a cada muestra se le calcula su varianza, se obtendrá la distribución muestral de varianzas.

Para estimar la varianza poblacional o la desviación estándar, se necesita conocer el estadístico X^2 . Si se elige una muestra de tamaño n de una población normal con varianza σ^2 , el estadístico:

$$\frac{(n-1)s^2}{\sigma^2}$$

tiene una distribución muestral que es una **distribución ji-cuadrada** con $gl=n-1$ **grados de libertad** y se denota X^2 (X es la minúscula de la letra griega ji). El estadístico ji-cuadrada está dado por:

$$X^2 = \frac{(n-1)s^2}{\sigma^2}$$

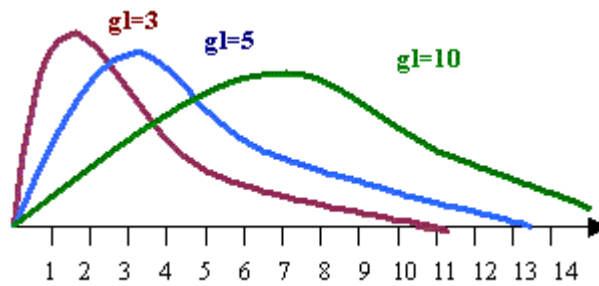
donde n es el tamaño de la muestra, s^2 la varianza muestral y σ^2 la varianza de la población de donde se extrajo la muestra. El estadístico ji-cuadrada también se puede dar con la siguiente expresión:

$$X^2 = \frac{\sum (x - \bar{x})^2}{\sigma^2}$$

Propiedades de las distribuciones ji-cuadrada

1. Los valores de X^2 son mayores o iguales que 0.
2. La forma de una distribución X^2 depende del $gl=n-1$. En consecuencia, hay un número infinito de distribuciones X^2 .
3. El área bajo una curva ji-cuadrada y sobre el eje horizontal es 1.
4. Las distribuciones X^2 no son simétricas. Tienen colas estrechas que se extienden a la derecha; esto es, están sesgadas a la derecha.
5. Cuando $n > 2$, la media de una distribución X^2 es $n-1$ y la varianza es $2(n-1)$.
6. El valor modal de una distribución X^2 se da en el valor $(n-3)$.

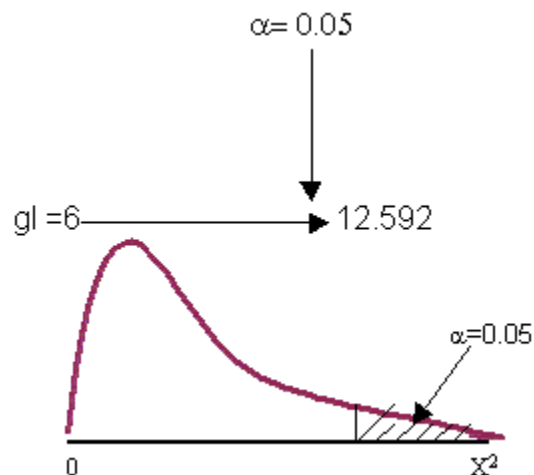
La siguiente figura ilustra tres distribuciones X^2 . Note que el valor modal aparece en el valor $(n-3) = (gl-2)$.



La función de densidad de la distribución X^2 esta dada por:

$$f(x) = \frac{1}{2^{v/2} \Gamma(v/2)} x^{v/2-1} e^{-x/2} \quad \text{para } x > 0$$

La tabla que se utilizará para estos apuntes es la del libro de probabilidad y estadística de Walpole, la cual da valores críticos $X^2_{\alpha}(gl)$ para veinte valores especiales de α . Para denotar el valor crítico de una distribución X^2 con gl grados de libertad se usa el símbolo $X^2_{\alpha}(gl)$; este valor crítico determina a su derecha un área de α bajo la curva X^2 y sobre el eje horizontal. Por ejemplo para encontrar $X^2_{0.05}(6)$ en la tabla se localiza 6 gl en el lado izquierdo y $\alpha = 0.05$ a lo largo del lado superior de la misma tabla.



Cálculo de Probabilidad

El cálculo de probabilidad en una distribución muestral de varianzas nos sirve para saber como se va a comportar la varianza o desviación estándar en una muestra que proviene de una distribución normal.

Ejemplos:

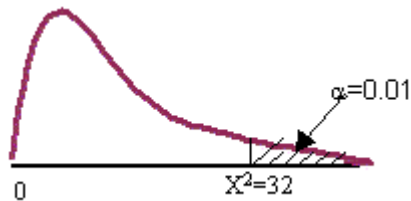
- Suponga que los tiempos requeridos por un cierto autobús para alcanzar un de sus destinos en una ciudad grande forman una distribución normal con una desviación estándar $\sigma = 1$ minuto. Si se elige al azar una muestra de 17 tiempos, encuentre la probabilidad de que la varianza muestral sea mayor que 2.

Solución:

Primero se encontrará el valor de ji-cuadrada correspondiente a $s^2=2$ como sigue:

$$X^2 = \frac{(n-1)s^2}{\sigma^2} = \frac{(17-1)(2)}{(1)^2} = 32$$

El valor de 32 se busca adentro de la tabla en el renglón de 16 grados de libertad y se encuentra que a este valor le corresponde un área a la derecha de 0.01. En consecuencia, el valor de la probabilidad es $P(s^2 > 2)$



- Encuentre la probabilidad de que una muestra aleatoria de 25 observaciones, de una población normal con varianza $\sigma^2 = 6$

, tenga una varianza muestral:

- Mayor que 9.1
- Entre 3.462 y 10.745

Solución.

- Primero se procederá a calcular el valor de la ji-cuadrada:

$$X^2 = \frac{(n-1)s^2}{\sigma^2} = \frac{(25-1)(9.1)}{6} = 36.4$$

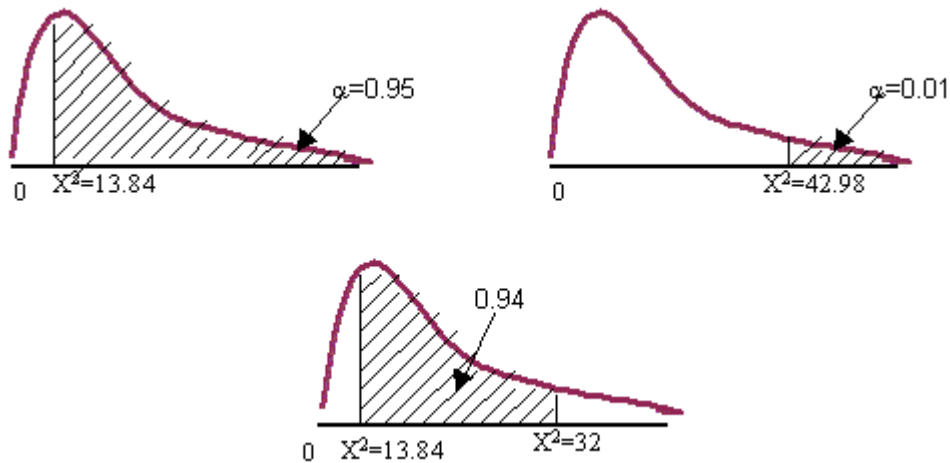
Al buscar este número en el renglón de 24 grados de libertad nos da un área a la derecha de 0.05. Por lo que la $P(s^2 > 9.1) = 0.05$

- Se calcularán dos valores de ji-cuadrada:

$$X^2 = \frac{(n-1)s^2}{\sigma^2} = \frac{(25-1)(3.462)}{6} = 13.847 \quad \text{y} \quad X^2 = \frac{(25-1)(10.745)}{6} = 42.98$$

Aquí se tienen que buscar los dos valores en el renglón de 24 grados de libertad. Al buscar el valor de 13.846 se encuentra un área a la derecha de 0.95. El valor de 42.98 da un área a la derecha de 0.01. Como se está pidiendo la probabilidad entre dos valores se resta el área de 0.95 menos 0.01 quedando 0.94.

Por lo tanto la $P(3.462 \leq s^2 \leq 10.745) = 0.94$



Estimación de la Varianza

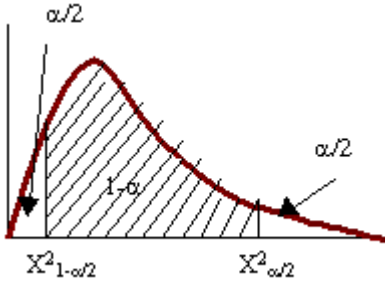
Para poder estimar la varianza de una población normal se utilizará la distribución ji-cuadrada.

$$X^2 = \frac{(n-1)s^2}{\sigma^2}$$

Al despejar esta fórmula la varianza poblacional nos queda:

$$\sigma^2 = \frac{(n-1)s^2}{X^2}$$

Los valores de X^2 dependerán de nivel de confianza que se quiera al cual le llamamos $1 - \alpha$. Si nos ubicamos en la gráfica se tiene:



Ejemplos:

1. Los siguientes son los pesos, en decagramos, de 10 paquetes de semillas de pasto distribuidas por cierta compañía: 46.4, 46.1, 45.8, 47.0, 46.1, 45.9, 45.8, 46.9, 45.2 y 46. Encuentre un intervalo de confianza de 95% para la varianza de todos los paquetes de semillas de pasto que distribuye esta compañía, suponga una población normal.

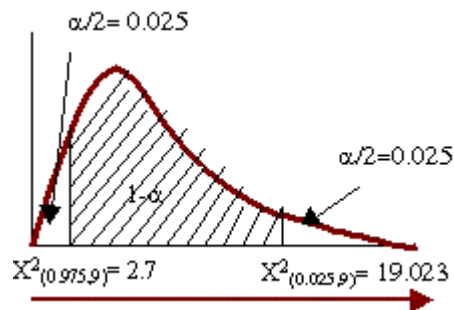
Solución:

Primero se calcula la desviación estándar de la muestra:

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n - 1}} = \sqrt{\frac{(46.4 - 46.12)^2 + (46.1 - 46.12)^2 + \dots + (46 - 46.12)^2}{10 - 1}} = 0.5347$$

al elevar este resultado al cuadrado se obtiene la varianza de la muestra $s^2 = 0.286$.

Para obtener un intervalo de confianza de 95% se elige un $\alpha = 0.05$. Después con el uso de la tabla con 9 grados de libertad se obtienen los valores de X^2 .



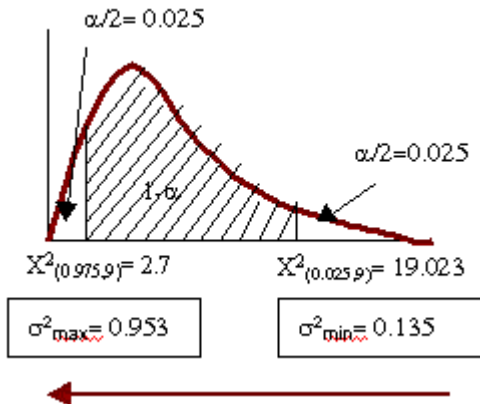
Se puede observar en la gráfica anterior que el valor de X^2 corre en forma normal, esto es de izquierda a derecha.

Por lo tanto, el intervalo de confianza de 95% para la varianza es:

$$\sigma^2_{\max} = \frac{(10-1)(0.286)}{2.7} = 0.953$$

$$\sigma^2_{\min} = \frac{(10-1)(0.286)}{19.023} = 0.135$$

Graficamente:



Se observa que la varianza corre en sentido contrario, pero esto es sólo en la gráfica. La interpretación quedaría similar a nuestros temas anteriores referentes a estimación. Con un nivel de confianza del 95% se sabe que la varianza de la población de los pesos de los paquetes de semillas de pasto esta entre 0.135 y 0.935 decagramos al cuadrado.

- En trabajo de laboratorio se desea llevar a cabo comprobaciones cuidadosas de la variabilidad de los resultados que producen muestras estándar. En un estudio de la cantidad de calcio en el agua potable, el cual se efectúa como parte del control de calidad, se analizó seis veces la misma muestra en el laboratorio en intervalos aleatorios. Los seis resultados en partes por millón fueron 9.54, 9.61, 9.32, 9.48, 9.70 y 9.26. Estimar la varianza de los resultados de la población para este estándar, usando un nivel de confianza del 90%.

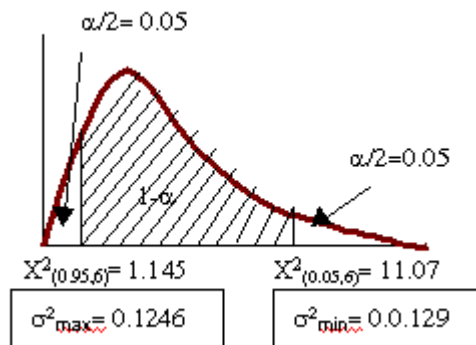
Solución:

Al calcular la varianza de la muestra se obtiene un valor de $s^2 = 0.0285$.

Se busca en la tabla los valores correspondientes con 5 grados de libertad, obteniéndose dos resultados. Para $X^2_{(0.95,5)} = 1.145$ y para $X^2_{(0.05,5)} = 11.07$.

Entonces el intervalo de confianza esta dado por:

$$\sigma^2_{\max} = \frac{(6-1)(0.0285)}{1.145} = 0.1246 \quad \text{y} \quad \sigma^2_{\min} = \frac{(6-1)(0.0285)}{11.07} = 0.0129$$



Ensayo de Hipótesis para la Varianza de una Población Normal

En la mayoría de los casos se tiene el problema de desconocer la varianza o desviación estándar de la población, en donde las distribuciones son normales. Si se desea probar una hipótesis acerca de la varianza se puede hacer utilizando las medidas estadísticas con las que se construyó el intervalo de confianza σ^2 , esto es con la distribución Ji- cuadrada.

Ejemplos:

1. Una compañía que produce una parte maquinada para un motor, afirma que tiene una varianza de diámetro no mayor a 0.0002 pulgadas. Una muestra aleatoria de 10 de dichas partes dio una varianza de muestra $s^2 = 0.0003$. Si se supone que las medidas del diámetro se distribuyen en forma normal, ¿hay evidencia para refutar lo que afirma el proveedor? Use $\alpha = 0.05$.

Solución:

Como en todos los ensayos de hipótesis que se han realizado anteriormente el procedimiento es el mismo. Después de que se identifican los datos, se plantea la hipótesis para determinar el tipo de ensayo.

Datos:

$$\sigma^2 = 0.0002$$

$$n = 10$$

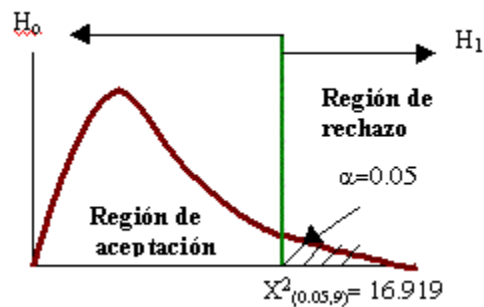
$$s^2 = 0.0003$$

$$\alpha = 0.05$$

Ensayo de hipótesis:

$$H_0; \sigma^2 = 0.0002$$

$$H_1; \sigma^2 > 0.0002$$



Regla de decisión:

Si $X^2_R \leq 16.919$ no se rechaza H_0 .

Si $X^2_R > 16.919$ se rechaza H_0 .

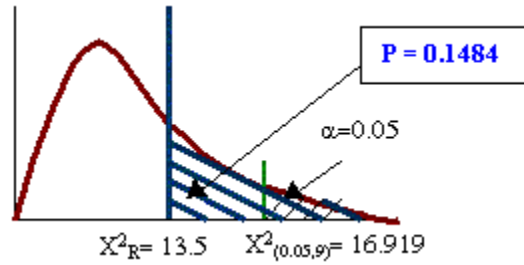
Cálculos:

$$X^2_R = \frac{(n-1)s^2}{\sigma^2} = \frac{(10-1)(0.0003)}{0.0002} = 13.5$$

Justificación y decisión:

Como 13.5 no es mayor que 16.919 por lo tanto no se rechaza H_0 y se concluye con un nivel de significancia de 0.05 que no se puede refutar la afirmación del proveedor.

Este ejercicio se puede aprovechar para calcular el valor de P. En la tabla se busca el valor de 13.5 en el renglón de 9 grados de libertad. Interpolando entre 0.10 y 0.20 se obtiene un valor de P de 0.1484.



2. El contenido de azúcar del almíbar de los duraznos enlatados tiene una distribución normal, donde se cree que la varianza es $\sigma^2 = 18 \text{ mg}^2$. Se toma una muestra de 10 latas dieron una desviación estándar de 4.8 mg. ¿Muestran estos datos suficiente evidencia para decir que la varianza ha cambiado?. Use un $\alpha = 0.05$ y calcule el valor de P.

Solución:

Datos:

$$\sigma^2 = 18$$

$$n = 10$$

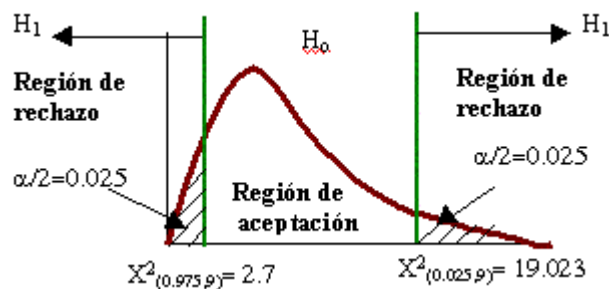
$$s = 4.8$$

$$\alpha = 0.05$$

Ensayo de hipótesis:

$$H_0: \sigma^2 = 18$$

$$H_1: \sigma^2 \neq 18$$



Regla de decisión:

Si $2.7 \leq X^2_R \leq 19.023$ no se rechaza H_0 .

Si $X^2_R < 2.7$ ó si $X^2_R > 19.023$ se rechaza H_0 .

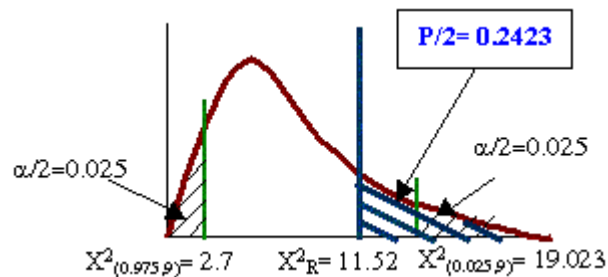
Cálculos:

$$X^2_R = \frac{(n-1)s^2}{\sigma^2} = \frac{(10-1)(4.8)^2}{18} = 11.52$$

Justificación y decisión:

Como 11.52 está entre 2.7 y 19.023, no se rechaza H_0 , y se concluye con un nivel de significancia de 0.05 que la varianza del contenido de azúcar del almíbar no ha cambiado, esto es es de 18 mg^2 .

Si recordamos al principio de este tema se dijo que la media de la distribución ji-cuadrada es $(n-1)$, por lo tanto la media de este ejercicio es de 9. Como el valor real de $X^2_R = 11.52$ este número se encuentra a la derecha de la media, lo cual quiere decir que el valor de $P/2$ será el área a la derecha del valor de X^2_R . Al buscar el valor de 11.52 en la tabla se obtiene un área de 0.2423, por lo tanto $P/2 = 0.2423$ y **$P = (2)(0.2423) = 0.4846$**



- Experiencia anterior indica que el tiempo que se requiere para que los estudiantes de último año de preparatoria completen una prueba estandarizada es una variable aleatoria normal con una desviación estándar de seis minutos. Se toma una muestra aleatoria de 20 estudiantes de último año de preparatoria y se obtiene una desviación estándar de 4.51. ¿Muestran estos datos suficiente evidencia para decir que la desviación estándar disminuyó?. Utilice el valor de P para su decisión.

Solución:

Datos:

$$\sigma = 6$$

$$n = 20$$

$$s = 4.51$$

Ensayo de hipótesis:

$$H_0; \sigma = 6$$

$$H_1; \sigma < 6$$

Cálculos:

$$X^2_R = \frac{(n-1)s^2}{\sigma^2} = \frac{(20-1)(4.51)^2}{(6)^2} = 10.735$$

Para obtener el valor de P, se busca en la tabla el 10.735 con 19 grados de libertad, y el área que se encuentra es la que está a la derecha de este valor. Como la media de esta distribución ji-cuadrada es de 19, por lo tanto el valor de 10.735 queda a la izquierda de la media. El valor de P es de 0.07, y con esto se puede concluir que si hubiéramos utilizado un nivel de significancia de 0.10, se rechaza H_0 y se concluye que la desviación estándar disminuyó, pero si se utiliza un valor de $\alpha = 0.05$, entonces no se rechaza H_0 y se concluiría que la desviación estándar no disminuyó. La decisión depende del error tipo I que esté dispuesto a tolerar el investigador.

