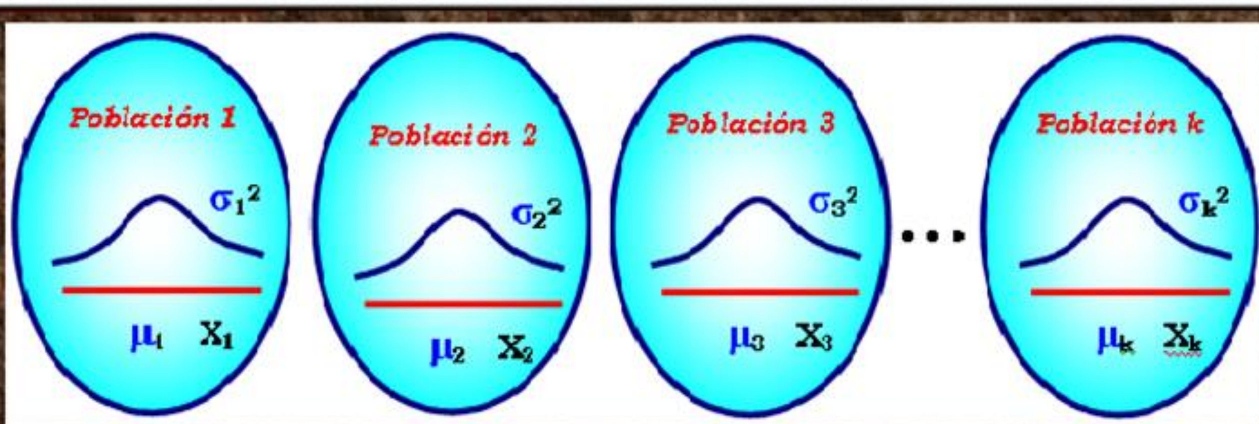
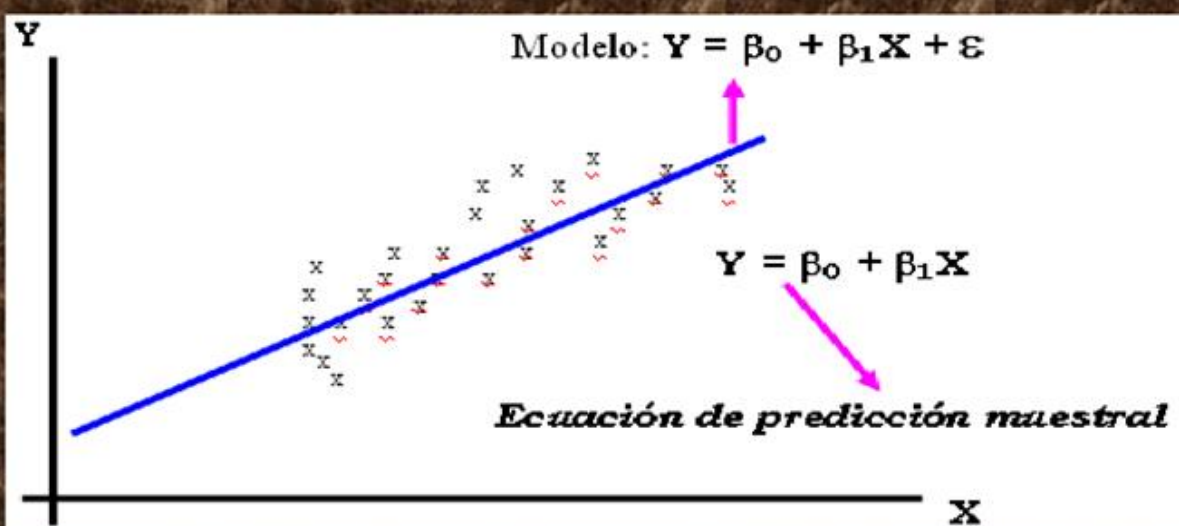
"Estadística II"
Código: 834501T
Ingeniería Industrial y en Informática

"Guía de la Unidad # 3"





Fecha: 22 de Junio del 2025

Fernando A. Contreras J.

"Aplicaciones de la χ^2"

"Prueba de Buen Ajuste"

Es una prueba que se utiliza para determinar si una población estadística, tiene una distribución teórica específica. Esta prueba se basa en que tan buen ajuste se puede tener entre las frecuencias de ocurrencia de las observaciones en una muestra observada y las frecuencias esperadas que se obtienen a partir de la distribución hipotética. Esto se realiza a partir del estadístico de prueba:

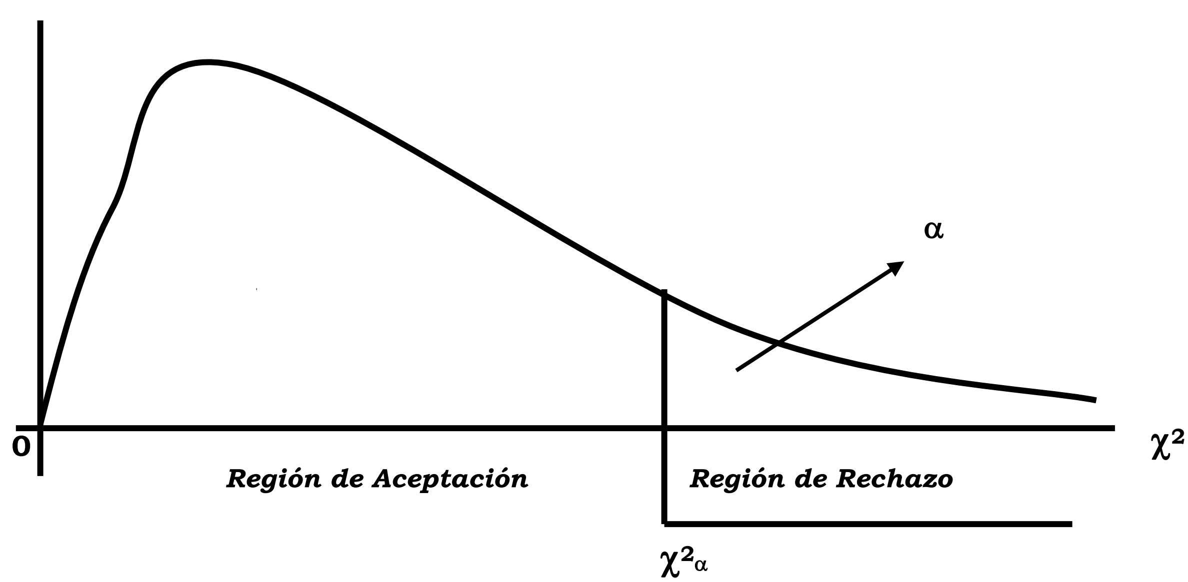
$$\chi^2_c = \sum_{i=1}^{k} \frac{(o_i - e_i)^2}{e_i} \quad ;$$

donde "χ^2_c" es un valor de una variable aleatoria cuya distribución muestral se aproxima muy de cerca con la distribución ***"ji cuadrada"*** con $\nu = k - 1$ grados de libertad. Los símbolos o_i y e_i representan las frecuencias observadas y esperadas, respectivamente, para la **i-ésima** celda y **k** representa el número de celdas.

Las hipótesis que se plantean son:

H_0: La variable **"X"** tiene una distribución **" . . . "**

H_1: La variable **"X"** no tiene una distribución **" . . . "**



"Luego"

- Si $\chi^2_c \in$ Región de Rechazo, entonces se Rechaza a $\mathbf{H_0}$.
- Si $\chi^2_c \in$ Región de Aceptación, entonces se Acepta a $\mathbf{H_0}$.

"Ejemplo 1": Considérese el lanzamiento de un dado. Elabore un contraste de hipótesis para determinar si el dado es legal. Para contrastar éstas hipótesis, suponga que se lanzó el dado **120** veces y se obtuvieron los siguientes resultados:

Tabla de ***"frecuencia observadas"*** del lanzamiento del dado:

Cara Superior	***1***	***2***	***3***	***4***	***5***	***6***	***Total***
Frecuencias Observada	***20***	***22***	***17***	***18***	***19***	***24***	***120***

Utilice para ello un nivel de significancia de α **= 0.05**.

"Desarrollo"

Para determinar si el dado es legal **(*un cubo perfecto*)**, se realizo un experimento que generó los resultados mostrados en la tabla anterior. Por otro lado el hecho de que el dado sea legal quiere decir que cuando se lanza, todas las caras del mismo tienen la misma posibilidad de ocurrir; de ahí que la frecuencia esperada de cada cara es igual. Esto quiere decir que los resultados esperados del lanzamiento de un dado se ajustan a una distribución teórica específica. Esta distribución teórica específica es la ***"Distribución Uniforme Discreta"***.

De acuerdo a lo anterior se planteará la hipótesis de que el dado es legal, y por tanto sus resultados se ajustan a la distribución uniforme discreta.

Sea "**X**" la variable aleatoria cuya distribución se supone ***"Uniforme Discreta"***.

Se plantean las hipótesis:

H_0: La variable **"X"** tiene una distribución "***Uniforme Discreta***".

H_1: La variable **"X"** no tiene una distribución "***Uniforme Discreta***".

Se construye una tabla de ***"frecuencias esperadas"*** del lanzamiento del dado:

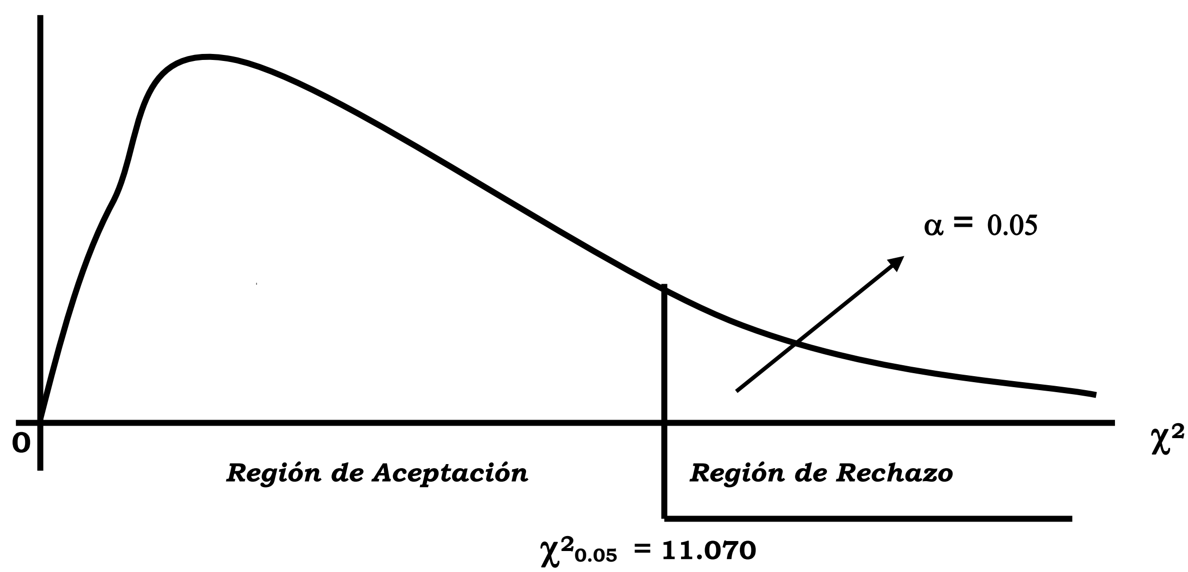
Cara Superior	***1***	***2***	***3***	***4***	***5***	***6***	***Total***
Frecuencias Esperadas	***20***	***20***	***20***	***20***	***20***	***20***	***120***

Se calcula el estadístico de prueba χ^2_c:

$$\chi^2_c = \frac{(20-20)^2}{20} + \frac{(22-20)^2}{20} + \frac{(17-20)^2}{20} + \frac{(18-20)^2}{20} + \frac{(19-20)^2}{20} + \frac{(24-20)^2}{20}$$

Luego, $\chi^2_c = 1.7$

$\nu = 6 - 1 = 5$ Grados de Libertad.



Luego como $\chi^2_c \in$ Región de Aceptación, entonces se Acepta a $\mathbf{H_0}$. Es decir, se concluye con un nivel de significancia de **0.05** que no existe suficiente evidencia estadística para decir que el dado no esta balanceado.

<u>"Ejemplo 2"</u> Las personas que mueren como consecuencia de accidentes de tráfico en un cierto cruce de vías en un año **(52** semanas**)**; esta dada por la siguiente estadística,

Tabla de ***<u>"frecuencias observadas"</u>***:

# de personas muertas	**0**	**1**	**2**	**3**	**4**	**5**	***Total***
Frecuencia Observada por Semanas	**6**	**10**	**20**	**10**	**6**	**0**	***52 Semanas***

Elabore un contraste de hipótesis para determinar si los resultados de ésta estadística se ajustan a alguna distribución teórica específica. Utilice un nivel de significancia de **0.05**.

<u>"Desarrollo"</u>

Como se puede observar que el número de ocurrencias es aleatorio en un intervalo de tiempo; se puede sugerir que las observaciones se ajustan al modelo de ***"Poisson"***.

Sea **"X"** el **#** de personas muertas en cruce de vías por semana.

Se plantean las hipótesis:

$\mathbf{H_0}$: La variable **"X"** tiene una distribución de ***"Poisson"***.

$\mathbf{H_1}$: La variable **"X"** no tiene una distribución de ***"Poisson"***.

Se construye una tabla de ***<u>"frecuencias esperadas"</u>***:

Para ello se calcula el parámetro λ de la distribución de Poisson.

$$\lambda \approx X_p = \frac{0 \times 6 + 1 \times 6 + 2 \times 20 + 3 \times 10 + 4 \times 6 + 5 \times 0}{52} = 2 \text{ Muertes/Semana}$$

Luego se utiliza el modelo de Poisson para construir el cuadro de frecuencias esperadas.

$$\mathbf{P(X = x) = \frac{e^{-\lambda}\lambda^{x}}{x!}}$$

P(X = 0**) =** 0.1353 **(**Se multiplica por $n = \sum f_i = 52$**)**; se tiene: 7,03 ≈ 7

P(X = 1**) =** 0.2706 **(**Se multiplica por $n = \sum f_i = 52$**)**; se tiene: 14,07 ≈ 14

P(X = 2**) =** 0.2706 **(**Se multiplica por $n = \sum f_i = 52$**)**; se tiene: 14,03 ≈ 14

P(X = 3**) =** 0.1804 **(**Se multiplica por $n = \sum f_i = 52$**)**; se tiene: 9,38 ≈ 9

P(X = 4**) =** 0.0902 **(**Se multiplica por $n = \sum f_i = 52$**)**; se tiene: 4,69 ≈ 5

P(X = 5**) =** 0.0368 **(**Se multiplica por $n = \sum f_i = 52$**)**; se tiene: 1,91 ≈ 2

Tabla de ***"frecuencias esperadas"***:

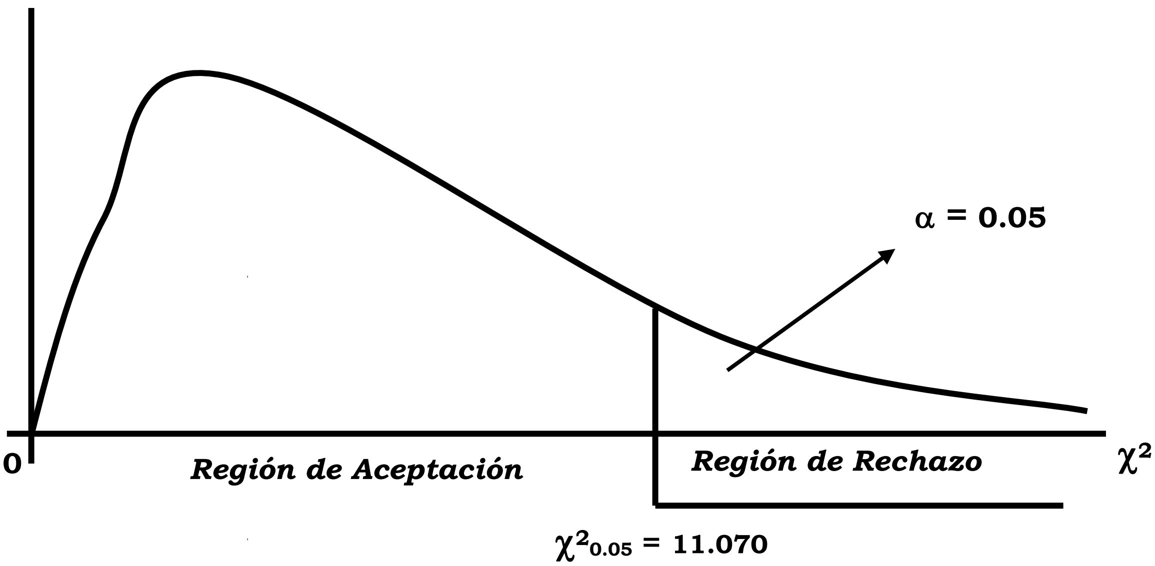
# de personas muertas	**0**	**1**	**2**	**3**	**4**	**5**	***Total***
Frecuencia Observada por Semanas	**7**	**14**	**14**	**9**	**5**	**2**	***51 Semanas***

Se calcula el estadístico de prueba χ^2_c:

$$\chi^2_c = \frac{(6-7)^2}{7} + \frac{(10-14)^2}{14} + \frac{(20-14)^2}{14} + \frac{(10-9)^2}{9} + \frac{(6-4)^2}{4} + \frac{(0-2)^2}{2} =$$

= 4.97

ν **= 6 – 1 = 5** Grados de Libertad



Luego como $\chi^2_c \in$ Región de Aceptación, entonces se Acepta a H_0.

∴ Se concluye con un nivel de significancia de **0.05** que la variable aleatoria "**X**" tiene una distribución de Poisson.

"Ejemplo 3": Las estaturas de **200** empleados de una empresa se distribuye de acuerdo a la tabla adjunta.

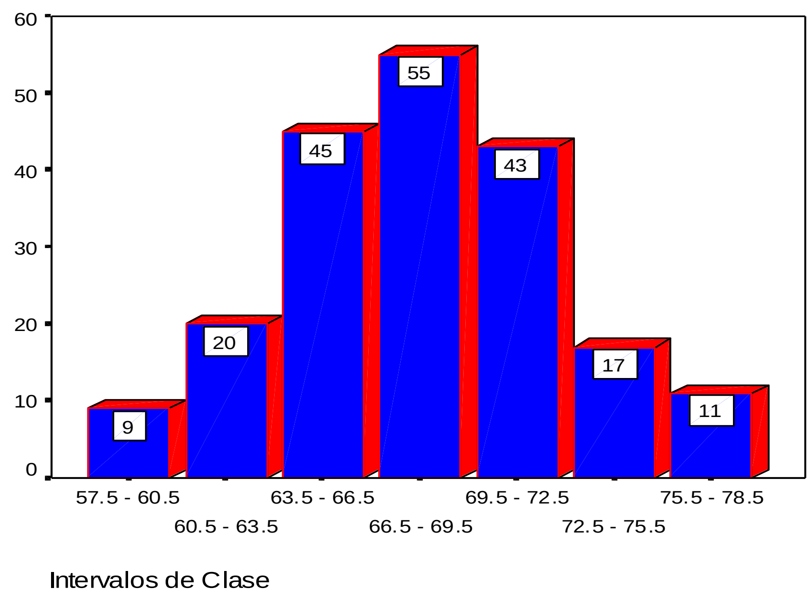
Tabla de ***"frecuencias observadas"***:

Pulgadas	***Frecuencia Observada***	***Marca de Clase (Xi)***
57.5 – 60.5	***9***	***59***
60.5 – 63.5	***20***	***62***
63.5 – 66.5	***45***	***65***
66.5 – 69.5	***55***	***68***
69.5 – 72.5	***43***	***71***
72.5 – 75.5	***17***	***74***
75.5 – 78.5	***11***	***77***
	$n = \sum f_{observ.} = \sum f_i = 200$	

Elabore un contraste de hipótesis para determinar si los resultados de ésta estadística se ajustan a alguna distribución teórica específica. Utilice un nivel de significancia de **0.05**.

“Desarrollo”

Al graficar la tabla de datos se tiene:



De acuerdo al análisis gráfico anterior de las observaciones, se puede sugerir que las observaciones se ajustan al modelo ***“Normal”***.

Sea **“X”** la estatura en pulgadas de los empleados.

Se plantean las hipótesis:

$\mathbf{H_0}$**:** La variable **“X”** tiene una distribución ***“Normal”***.

$\mathbf{H_1}$**:** La variable **“X”** no tiene una distribución ***“Normal”***.

Se construye una tabla de ***“frecuencias esperadas”***. Para ello se usa el supuesto de que la variable **“X”** tiene una distribución normal con media $\mu = \mathbf{X}$ y desviación estándar $\sigma = \mathbf{S}$; donde:

k

$$\mathbf{X} = \frac{\sum_{i=1} \mathbf{f_i \times X_i}}{\sum_{i=1}^{k} \mathbf{f_i}} = \frac{9\mathbf{x}59 + 20\mathbf{x}62 + 45\mathbf{x}65 + 55\mathbf{x}68 + 43\mathbf{x}71 + 17\mathbf{x}74 + 11\mathbf{x}77}{200}$$

$\underline{X = 67.97}$

$$\mathbf{S^2} = \frac{1}{n-1}\left(\sum_{i=1}^{k} X_i^2 f_i - \frac{\left(\sum_{i=1}^{k} X_i \times f_i\right)^2}{n}\right) = \frac{1}{200-1}\left(927728 - \frac{(13594)^2}{200}\right) = 18.81$$

$\underline{S = 4.34}$

Por otro lado:

Si H_0 es verdad; entonces $f_{esperada}$ = Probabilidad del Intervalo x n

De ahí que:

$$* \ \mathbf{f_{e1}} = \mathbf{P}(57.5 < \mathbf{X} < 60.5) \times 200 = \mathbf{P}\left(\underbrace{\frac{57.5 - 67.97}{4.34} < Z < \frac{60.5 - 67.97}{4.34}}_{0.0347}\right) \times 200 = 6.9 \approx \mathbf{7}$$

$$* \ \mathbf{f_{e2}} = \mathbf{P}(60.5 < \mathbf{X} < 63.5) \times 200 = \mathbf{P}\left(\underbrace{\frac{60.5 - 67.97}{4.34} < Z < \frac{63.5 - 67.97}{4.34}}_{0.1088}\right) \times 200 = 21.8 \approx \mathbf{22}$$

$$* \ \mathbf{f_{e3}} = \mathbf{P}(63.5 < \mathbf{X} < 66.5) \times 200 = \mathbf{P}\left(\underbrace{\frac{63.5 - 67.97}{4.34} < Z < \frac{66.5 - 67.97}{4.34}}_{0.2154}\right) \times 200 = 43.1 \approx \mathbf{43}$$

$$* \; f_{e4} = P(66.5 < X < 69.5) \times 200 = P\left(\frac{66.5 - 67.97}{4.34} < Z < \frac{69.5 - 67.97}{4.34}\right) \times 200 = 54.0 \approx 54$$

0.2699

$$* \; f_{e5} = P(69.5 < X < 72.5) \times 200 = P\left(\frac{69.5 - 67.97}{4.34} < Z < \frac{72.5 - 67.97}{4.34}\right) \times 200 = 42.8 \approx 43$$

0.2140

$$* \; f_{e6} = P(72.5 < X < 75.5) \times 200 = P\left(\frac{72.5 - 67.97}{4.34} < Z < \frac{75.5 - 67.97}{4.34}\right) \times 200 = 21.7 \approx 22$$

0.1083

$$* \; f_{e7} = P(75.5 < X < 78.5) \times 200 = P\left(\frac{75.5 - 67.97}{4.34} < Z < \frac{78.5 - 67.97}{4.34}\right) \times 200 = 6.7 \approx 7$$

0.0334

Tabla de ***"frecuencias esperadas"***:

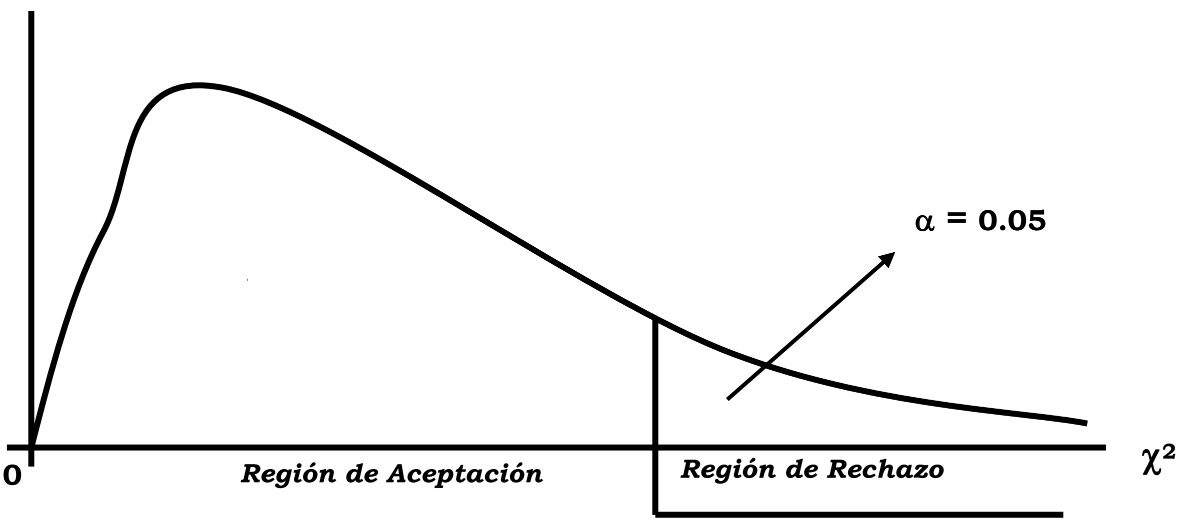
Pulgadas	Frecuencia Esperadas	Marca de Clase (Xi)
57.5 – 60.5	7	59
60.5 – 63.5	22	62
63.5 – 66.5	43	65
66.5 – 69.5	54	68
69.5 – 72.5	43	71
72.5 – 75.5	22	74
75.5 – 78.5	7	77
	$n = \Sigma f_{esperad.} = \Sigma f_i = 198$	

Se calcula el estadístico de prueba χ^2_c:

$$\chi^2_c = \frac{(9-7)^2}{7} + \frac{(20-22)^2}{22} + \frac{(45-43)^2}{43} + \frac{(55-54)^2}{54} + \frac{(43-43)^2}{43} + \frac{(17-22)^2}{22} + \frac{(11-7)^2}{7} =$$

= 4.29

$\nu = 7 - 1 = 6$ Grados de Libertad



$\chi^2_{0.05} = 12.592$

Luego como $\chi^2_c \in$ Región de Aceptación, entonces se Acepta a H_0.

∴ Se concluye con un nivel de significancia de **0.05** que existe alguna evidencia estadística para decir que la variable aleatoria **"X"** tiene una distribución aproximadamente normal.

"Prueba de Independencia"

Es una prueba que se utiliza para determinar la independencia de un grupo de variables con respecto a otro grupo de variables.

Se calcula el estadístico de prueba:

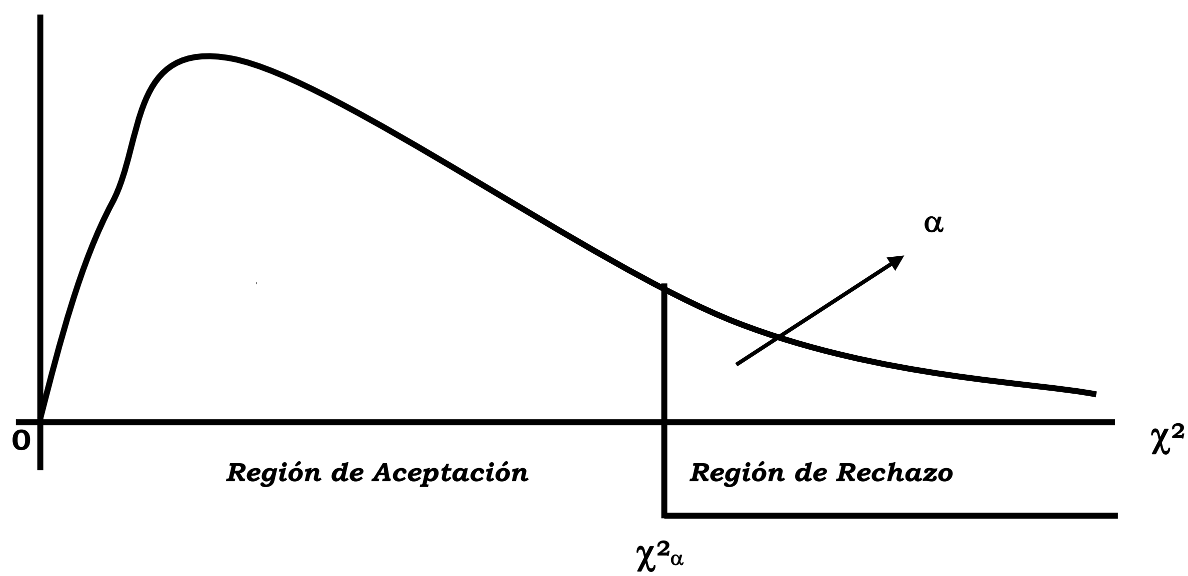
$$\chi^2_c = \sum_{i=1}^{k} \frac{(o_i - e_i)^2}{e_i} \quad ;$$

donde "χ^2_c" es un valor de una variable aleatoria cuya distribución muestral se aproxima muy de cerca con la distribución ***"ji cuadrada"*** con $\nu = (r - 1)(c - 1)$ grados de libertad. Las letras **r** y **c** representan las filas y las columnas de la tabla de contingencia **"r x c"** y **k** representa el número de celdas **"r x c"** de la tabla de contingencia.

Las hipótesis que se plantean son:

H_0: Son Independientes.

H_1: No son independientes.



"Luego"

- Si $\chi^2_c \in$ Región de Rechazo, entonces se Rechaza a **H_0**.
- Si $\chi^2_c \in$ Región de Aceptación, entonces se Acepta a **H_0**.

La regla general para obtener la tabla de frecuencias esperadas para cada una de las celdas, esta dada por la siguiente fórmula:

$$\textit{Frecuencia Esperada} = \frac{(\textit{Total de la Columna}) \times (\textit{Total de la Fila})}{\textit{Gran Total}}$$

"Ejemplo 4": Considérese una muestra aleatoria de **1000** individuos del Estado Táchira. Determine si existe independencia entre los consumidores de los estratos sociales bajo, medio y alto con respecto a la preferencia por una nueva marca de cerveza. Los datos recogidos se ilustran en la tabla de contingencia adjunta. Utilice un nivel de significancia de **0.05**.

Tabla de contingencia de ***"frecuencias observadas"***:

	Estrato Social			
Nueva Cerveza	***Bajo***	***Medio***	***Alto***	***Total***
La Prefieren	**182**	**213**	**203**	**598**
No la Prefieren	**154**	**138**	**110**	**402**
Total	**336**	**351**	**313**	**1000**

"Desarrollo"

Se plantean las hipótesis:

$\mathbf{H_0}$: Hay independencia entre los estratos.

$\mathbf{H_1}$: No hay independencia entre los estratos.

Se construye la tabla de contingencia de ***"frecuencias esperadas"***:

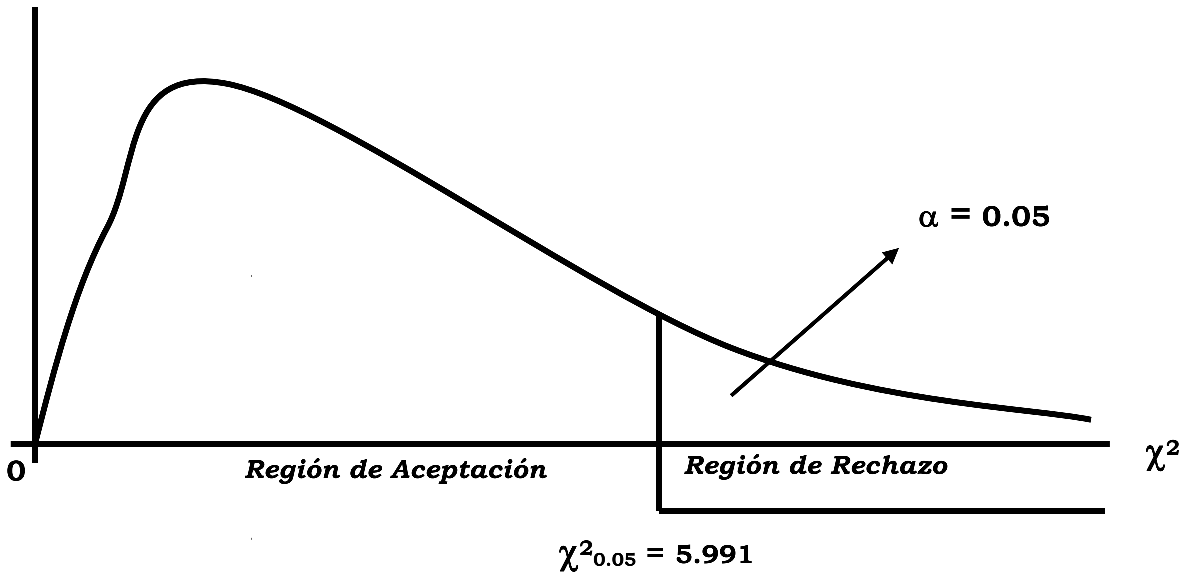
	Estrato Social			
Nueva Cerveza	***Bajo***	***Medio***	***Alto***	***Total***
La Prefieren	**201**	**210**	**187**	**598**
No la Prefieren	**135**	**141**	**126**	**402**
Total	**336**	**351**	**313**	**1000**

Se calcula el estadístico de prueba χ^2_c:

$$\chi^2_c = \frac{(182-201)^2}{201} + \frac{(213-210)^2}{210} + \frac{(203-187)^2}{187} + \frac{(154-135)^2}{135} + \frac{(138-141)^2}{141} + \frac{(110-126)^2}{126} =$$

$$= \mathbf{7.98}$$

$\nu = (2 - 1)(3 - 1) = 2$ Grados de Libertad



Luego como $\chi^2_c \in$ Región de Rechazo, entonces se Rechaza a H_0. $\therefore$ Se concluye con un nivel de significancia de **0.05** que la preferencia de los individuos del Estado Táchira con respecto a la nueva marca de cerveza no es independiente con respecto al nivel de los estratos sociales.

Nota: Cuando se tiene una tabla de contingencia "**2 x 2**" se aplica una corrección llamada ***"Corrección de Yates"*** :

$$\chi^2_{(corregida)} = \sum_{i=1}^{k} \frac{|(o_i - e_i) - 0.5|^2}{e_i}$$

"Prueba de Homogeneidad"

Es una prueba que se utiliza para determinar la homogeneidad entre un grupo de variables con respecto a otro grupo de variables.

Se calcula el estadístico de prueba:

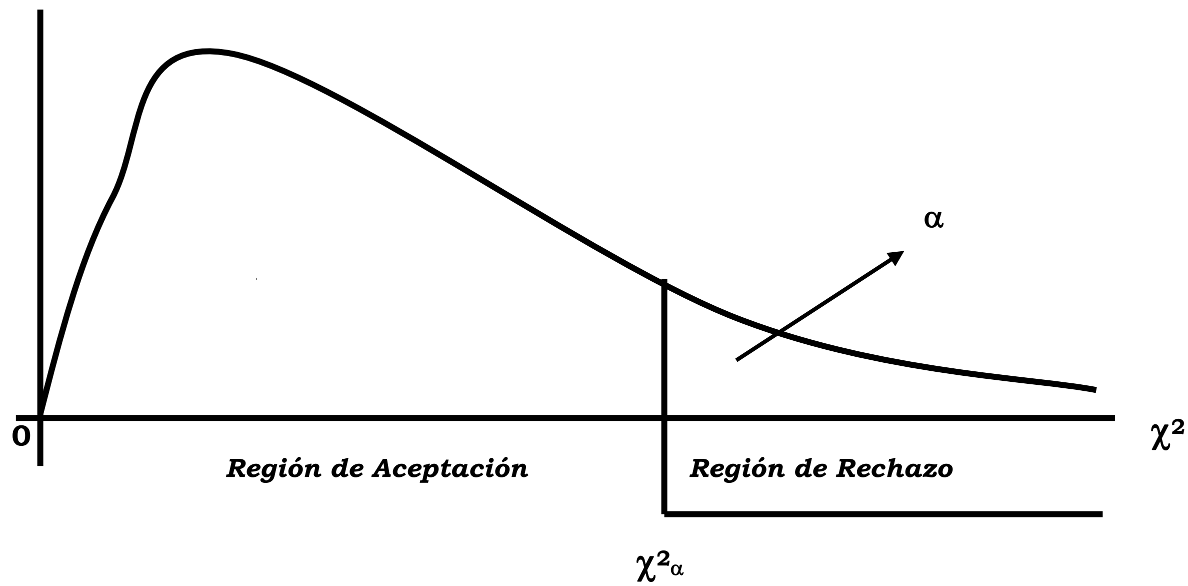
$$\chi^2_c = \sum_{i=1}^{k} \frac{(o_i - e_i)^2}{e_i} \quad ;$$

donde "χ^2_c" es un valor de una variable aleatoria cuya distribución muestral se aproxima muy de cerca con la distribución ***"ji cuadrada"*** con **ν = (r – 1)(c – 1)** grados de libertad. Las letras **r** y **c** representan las filas y las columnas de la tabla de contingencia **"r x c"** y **k** representa el número de celdas **"r x c"** de la tabla de contingencia.

Las hipótesis que se plantean son:

H_0: Hay Homogeneidad

H_1: No hay Homogeneidad



"Luego"

- Si χ^2_c $\in$ Región de Rechazo, entonces se Rechaza a **H_0**.
- Si χ^2_c $\in$ Región de Aceptación, entonces se Acepta a **H_0**.

"Ejemplo 5": Considérese la tabla de contingencia adjunta.

Tabla de contingencia de ***"frecuencias observadas"***:

	Estratificación por Edad			
Nuevo Refresco	***Niño***	***Joven***	***Adulto***	***Total***
Le Gusta	**82**	**70**	**62**	**214**
No le Gusta	**93**	**62**	**67**	**222**
Indecisos	**25**	**18**	**21**	**64**
Total	**200**	**150**	**150**	**500**

Pruebe la hipótesis de que las opiniones con respecto al ***"Nuevo Refresco"*** propuestas son las mismas (hay homogeneidad) dentro de cada estrato según la edad. Utilice un nivel de significancia de **0.05**.

"Desarrollo"

Se plantean las hipótesis:

$\mathbf{H_0}$: Hay Homogeneidad entre las Opiniones.

$\mathbf{H_1}$: No hay Homogeneidad entre las Opiniones.

Se construye la tabla de contingencia de ***"frecuencias esperadas"***:

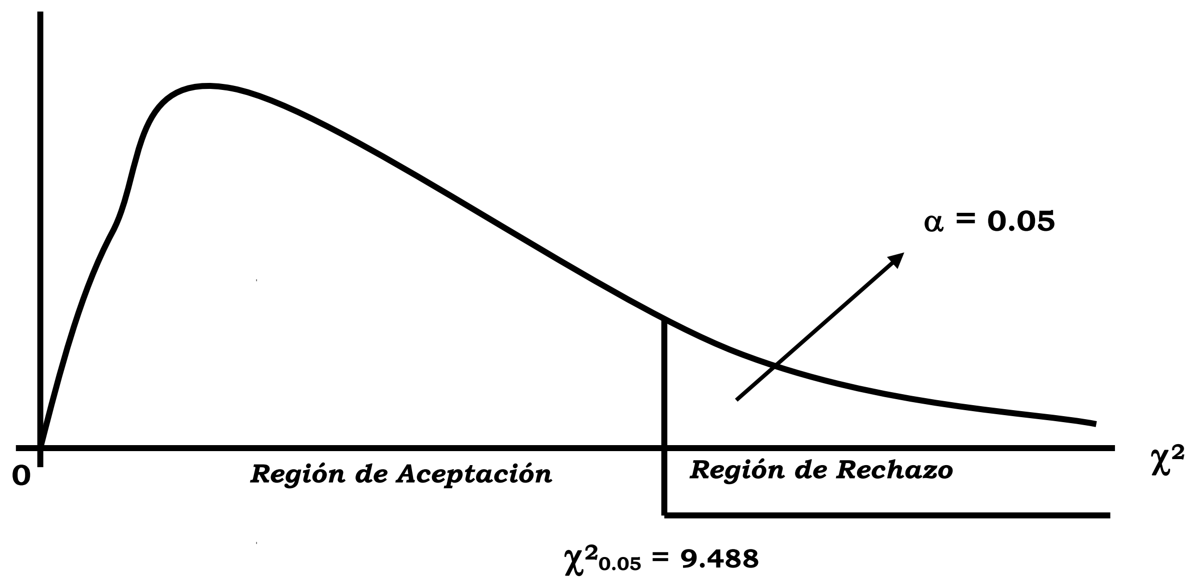
	Estratificación por Edad			
Nuevo Refresco	***Niño***	***Joven***	***Adulto***	***Total***
Le Gusta	**86**	**64**	**64**	**214**
No le Gusta	**89**	**67**	**67**	**223**
Indecisos	**26**	**19**	**19**	**64**
Total	**201**	**150**	**150**	**501**

Se calcula el estadístico de prueba χ^2:

$$\chi^2_c = \frac{(82-86)^2}{86} + \frac{(70-64)^2}{64} + \frac{(62-64)^2}{64} + \frac{(93-89)^2}{89} + \frac{(62-67)^2}{67} + \frac{(67-67)^2}{67} + \frac{(25-26)^2}{26} + \frac{(18-19)^2}{19} + \frac{(21-19)^2}{19}$$

$$= \mathbf{1.67}$$

ν **= (3 – 1)(3 – 1) = 4** Grados de Libertad



Luego como χ^2_c $\in$ Región de Aceptación, entonces se Acepta a $\mathbf{H_0}$.
$\therefore$ Se concluye con un nivel de significancia de **0.05** que no hay suficiente evidencia estadística para decir que las opiniones con respecto al nuevo refresco de los estratos de edad: Niño, Joven y Adulto difieran para cada opinión establecida.