## Nombre del alumno: Alexa Yomara Téllez Méndez

## Nombre del profesor: ABEL ESTRADA DICHI

## Licenciatura: MVZ

## Materia:  CONTROL TOTAL DE CALIDAD

## Nombre del trabajo: T student y Chicuadrada

Ocosingo, Chiapas a 13 de octubre del 2021

## Introducción

El estadístico ji-cuadrado (o chi cuadrado), que tiene distribución de probabilidad del mismo nombre, sirve para someter a prueba hipótesis referidas a distribuciones de frecuencias. En términos generales, esta prueba contrasta frecuencias observadas con las frecuencias esperadas de acuerdo con la hipótesis nula. En este ensayo se describe el uso del estadístico ji-cuadrado para probar la asociación entre dos variables utilizando una situación hipotética y datos simulados. Luego se describe su uso para evaluar cual buena puede resultar una distribución teórica, cuando pretende representar la distribución real de los datos de una muestra determinada. A esto se le llama evaluar la bondad de un ajuste. Probar la bondad de un ajuste es ver en qué medida se ajustan los datos observados a una distribución teórica o esperada. Para esto, se utiliza una segunda situación hipotética y datos simulados.

Del mismo modo que los estadísticos “z”, con su distribución normal y “t”, con su distribución t de Student, nos han servido para someter a prueba hipótesis que involucran a promedios y porcentajes, el estadístico ji-cuadrado (o chi cuadrado), que tiene distribución de probabilidad del mismo nombre, nos servirá para someter a prueba hipótesis referidas a distribuciones de frecuencias.

En primer lugar usaremos el estadístico ji-cuadrado para probar la asociación entre dos variables, y luego lo usaremos para evaluar en qué medida se ajusta la distribución de frecuencias obtenida con los datos de una muestra, a una distribución teórica o esperada.

En términos generales, esta prueba contrasta frecuencias observadas con las frecuencias esperadas de acuerdo con la hipótesis nula. Al igual que en el caso de las pruebas anteriormente presentadas, ilustraremos con ejemplos.

**Ji- cuadrado como prueba de asociación**

Supongamos que un investigador está interesado en evaluar la asociación entre uso de cinturón de seguridad en vehículos particulares y el nivel socioeconómico del conductor del vehículo. Con este objeto se toma una muestra de conductores a quienes se clasifica en una tabla de asociación, encontrando los siguientes resultados

| Uso de cinturón | Nivel socioeconómico bajo | Nivel socioeconómico medio | Nivel socioeconómico alto | TOTAL |
|---|---|---|---|---|
| SI | 8 | 15 | 28 | 51 |
| NO | 13 | 16 | 14 | 43 |
| TOTAL | 21 | 31 | 42 | 94 |

Los pasos del análisis estadístico en este caso son los siguientes:

**1. En primer lugar se debe plantear las hipótesis que someteremos a prueba**

*H0: "El uso de cinturón de seguridad es independiente del nivel socioeconómico".*

*H1: "El uso de cinturón de seguridad depende del nivel socioeconómico".*

En esta prueba estadística siempre la hipótesis nula plantea que las variables analizadas son independientes.

**2. En segundo lugar, obtener (calcular) las frecuencias esperadas**

Estas son las frecuencias que debieran darse si las variables fueran independientes, es decir, si fuera cierta la hipótesis nula.

Las frecuencias esperadas se obtendrán de la distribución de frecuencias del total de los casos, 51 personas de un total de 94 usan el cinturón y 43 de 94 no lo usan.

Esa misma proporción se debería dar al interior de los tres grupos de nivel socioeconómico, de manera que el cálculo responde al siguiente razonamiento: si de 94 personas 51 usan cinturón; de 21 personas.

Estas son las frecuencias que debieran presentarse si la hipótesis nula fuera verdadera y, por consiguiente, las variables fueran independientes.

| Uso de cinturón | Nivel bajo | Nivel medio | Nivel alto | TOTAL |
|---|---|---|---|---|
| SI | *11,4* | *16,8* | *22,8* | 51 |
| NO | *9,6* | *14,2* | *19,2* | 43 |
| TOTAL | 21 | 31 | 42 | 94 |

**3. En tercer lugar se debe calcular el estadístico de prueba**

En este caso, el estadístico de prueba es Ji-cuadrado que, como dijimos al comienzo, compara las frecuencias que entregan los datos de la muestra (frecuencias observadas) con las frecuencias esperadas, y tiene la siguiente fórmula cálculo:

$$\chi^2 = \sum \frac{(o_i - e_i)^2}{e_i}$$

Donde $o_i$ representa a cada frecuencia observada y $e_i$ representa a cada frecuencia esperada.

De este modo el valor del estadístico de prueba para este problema será:

$$\chi^2 = \sum \frac{(o_i - e_i)^2}{e_i} =$$

$$\frac{(8-11{,}4)^2}{11{,}4} + \frac{(13-9{,}6)^2}{9{,}6} + \frac{(15-16{,}8)^2}{16{,}8} + \frac{(16-14{,}2)^2}{14{,}2} + \frac{(28-22{,}8)^2}{22{,}8} + \frac{(14-19{,}2)^2}{19{,}2} = 5{,}23$$

Entonces $\chi^2 = 5{,}23$ Este es el valor de nuestro estadístico de prueba que ahora, siguiendo el procedimiento de problemas anteriores, debemos comparar con un valor de la tabla de probabilidades para ji-cuadrado (**x**$^2$). Esta tabla es muy parecida a la tabla *t de student*, pero tiene sólo valores positivos porque ji-cuadrado sólo da resultados positivos.

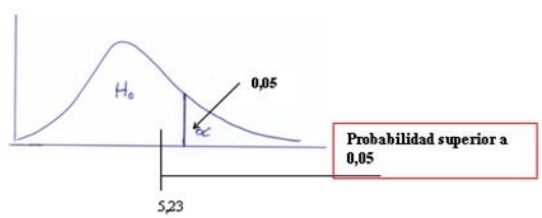
**Uso de tabla ji-cuadrado**

La tabla de ji-cuadrado tiene en la primera columna los grados de libertad y en la primera fila la probabilidad asociada a valores mayores a un determinado valor del estadístico.

Los grados de libertad dependen del número de celdas que tiene la tabla de asociación donde están los datos del problema y su fórmula de cálculo es muy sencilla:

Al comienzo elegimos un nivel de significación alfa=0,05. Entonces un valor de tabla para **x**$^2$ asociado a 2 grados de libertad y alfa 0,05 es 5,99.

Por lo tanto, como en el gráfico 2 vemos que 5,23 se encuentra a la izquierda de 5,99, la probabilidad asociada a valores superiores a 5,23 es mayor que alfa (0,05).




Según esto, debemos aceptar la hipótesis nula que plantea que las variables “uso de cinturón de seguridad” y “nivel socioeconómico” son independientes.

Limitación: como norma general, se exige que el 80% de las celdas en una tabla de asociación tengan valores esperados mayores de 5.

**Ji-cuadrado como prueba de bondad de ajuste**

También se puede usar el estadístico ji-cuadrado para evaluar cuán buena puede resultar una distribución teórica, cuando pretende representar la distribución real de los datos de una muestra determinada. A esto se le llama **evaluar la bondad de un ajuste**. Probar la bondad de un ajuste es ver en qué medida se ajustan los datos observados a una distribución teórica o esperada.

Tomemos como ejemplo la distribución esperada para los individuos de una población que son clasificados según grupo sanguíneo. Según estudios realizados en población, se espera que dicha distribución, en porcentajes.

| Grupo | Frecuencia esperada |
|---|---|
| AB | 2,0% |

| | |
|---|---|
| A | 30,5% |
| B | 9,3% |
| 0 | 58,2% |

En una muestra de 150 dadores de sangre se encontró la siguiente distribución:

| Grupo | Frecuencia observada |
|---|---|
| AB | 4 |
| A | 48 |
| B | 15 |
| 0 | 83 |

**Las hipótesis del problema son:**
*H0: los datos se ajustan a la distribución teórica.*
*H1: los datos no se ajustan a la distribución teórica.*

**Siguiendo el esquema general de solución propuesto para las pruebas de hipótesis, ahora corresponde elegir un nivel de significación**

Elegimos entonces alfa=0,01. El estadístico de prueba será ji-cuadrado, cuya fórmula es:

$$\chi^2 = \sum \frac{(o_i - e_i)^2}{e_i}$$

Debemos calcular las frecuencias esperadas en nuestro grupo. Si aplicamos los porcentajes esperados a la muestra de 150 casos podemos obtener las siguientes frecuencias esperadas (ei):

| Grupo | Frec. oi | Frec. ei |
|---|---|---|
| AB | 4 | 3,00 |
| A | 48 | 45,75 |
| B | 15 | 13,95 |
| 0 | 83 | 87,30 |
| Total | 150 | 150,00 |

Los grados de libertad de esta tabla se obtienen restando 1 al número de filas, en este caso: gl=4-1=3

Recordemos que la fila del total no se considera para los grados de libertad.

Si ya tenemos las frecuencias observadas y esperadas, podemos proceder a evaluar la diferencia entre ellas utilizando el estadístico ji-cuadrado. Si la diferencia entre frecuencias observadas y esperadas es grande, significará que la hipótesis nula es falsa, o sea, esta distribución no se ajusta a la distribución teórica y si, en cambio, resulta que la diferencia entre frecuencias observadas y esperadas no es muy grande, significará que la hipótesis nula es verdadera; por lo tanto, la distribución en la muestra se ajusta a la distribución teórica y diremos que no hay significación estadística.

El valor del estadístico de prueba ($\mathbf{x}^2$) es una medida de la diferencia entre frecuencias observadas y esperadas; por lo tanto, mientras mayor resulte, más fácil será rechazar la hipótesis nula.

## Conclusión

Dado que la probabilidad de $\chi^2 \geq 0,73$ es mayor que alfa, se acepta la hipótesis nula. Esto significa que los datos observados se ajustan a la distribución teórica, por lo tanto las diferencias observadas no son estadísticamente significativas.

En este ensayo se examinaron pruebas de hipótesis en las que la característica que se desconoce es alguna propiedad de la forma funcional de la distribución que se muestrea. Además se discutió pruebas de independencia de dos variables aleatorias en las cuales la evidencia muestral se obtiene mediante la clasificación de cada variable aleatoria en un cierto número de categorías. Este tipo de prueba recibe el nombre de bondad de ajuste. Para un tamaño específico del error de tipo I, la hipótesis nula será rechazada si existe una diferencia suficiente entre las frecuencias observadas y las esperadas.