# 7. Estadística no paramétrica

## Introducción

Las pruebas de hipótesis no solamente pueden estar distribuidas normalmente o ser numéricas; sino que pueden ser nominales (también llamadas categóricas) sino que su distribución puede no ser normal. A este tipo de información se le conoce como no paramétrica.

## Chi-cuadrada

El estadístico de prueba que se utiliza en estos casos es la distribución chi-cuadrada o ji-cuadrada cuya fórmula se denota de la siguiente manera:

$$\boxed{\chi^2 = \sum \frac{(f_0 - f_e)^2}{f_e}}$$

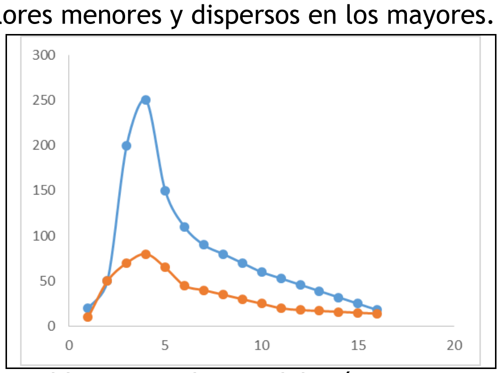
Para k-1 grados de libertad

Donde:
$\chi^2$ : Chi-cuadrada
$f_o$ : Frecuencia observada
$f_e$ : Frecuencia esperada

## Características de Chi-cuadrada

Entre las principales característica que se pueden identificar son:

a. **Valores no negativos**: Al elevar al cuadrado la variación entre la frecuencia observada y la frecuencia esperada, el resultado siempre es positivo.
b. **Familia de distribuciones**: Al contar variables con múltiples tipos de valores, las gráficas resultantes son variadas tanto en la forma como en la altura. Debido a los múltiples valores que puede tomar, se trabaja con k-1 grados de libertad para darle un mejor ajuste a los resultados de la prueba de hipótesis.
c. **Sesgada por la derecha**: tiene sesgo positivo, porque se puede concluir que los valores están concentrados en los valores menores y dispersos en los mayores.




Para cada característica de la variable en estudio, se deberá tener un pronóstico definido.

UNIVERSIDAD DE SAN PEDRO SULA

## *Ejemplo 7.1*

1. En una encuesta que se realizó en Espresso Americano, se dio a probar a 60 clientes la nueva quesadilla y se les pidió valorar si el nuevo producto les gustaba, los resultados que se obtuvieron son los siguientes:

| Resultados | Encuestas |
|---|---|
| Me gusta | 10 |
| No me gusta | 20 |
| Sin comentarios | 30 |

Si las respuestas se esperaban de manera uniforme, calcular el valor de chi-cuadrada.

### Desarrollo

Se aplicaron 60 encuestas; por lo que se esperaba que cada respuesta tendría un total uniforme; es decir, 20 por cada característica de la variable.

| Resultados | Encuestas $f_o$ | Pronóstico $f_e$ | $f_0 - f_e$ | | $(f_0 - f_e)^2$ | $\frac{(f_0 - f_e)^2}{f_e}$ |
|---|---|---|---|---|---|---|
| Me gusta | 10 | 20 | 10 - 20 = | -10 | 100 | 5 |
| No me gusta | 20 | 20 | 20 - 20 = | 0 | 0 | 0 |
| Sin comentarios | 30 | 20 | 30 - 20 = | 10 | 100 | 5 |
| | 60 | 60 | | | | 10 |

$$\chi^2 = \mathbf{10}$$

El valor de chi-cuadrada es 10.

2. En una distribuidora de vehículos se ha consultado a los clientes sobre su opinión con relación a las nuevas instalaciones. Una de las preguntas se refiere al tamaño de la nueva agencia. Se supone que las respuestas deben ser uniformes y que los resultados que se obtuvieron son:

| *NUEVAS INSTALACIONES* | *CLIENTES* |
|---|---|
| Muy grandes | 100 |
| Regular | 120 |
| Muy pequeñas | 74 |
| Adecuadas | 110 |

### Desarrollo

Se aplicaron 404 encuestas, por lo tanto se espera que cada característica debiera tener 101 respuestas cada una.

| *NUEVAS INSTALACIONES* | *CLIENTES* $f_o$ | *Pronóstico* $f_e$ | $\frac{(f_o - f_e)^2}{f_e}$ |
|---|---|---|---|
| Muy grandes | 100 | 101 | 0.010 |
| Regular | 120 | 101 | 3.574 |
| Muy pequeñas | 74 | 101 | 7.218 |
| Adecuadas | 110 | 101 | 0.802 |
| | | Σ | 11.604 |

El valor de chi-cuadrada es 11.604

USAP
UNIVERSIDAD DE SAN PEDRO SULA

# Prueba de bondad de ajuste: Frecuencias esperadas iguales

En una investigación cualitativas, las variables no son numéricas; pero, las frecuencias si lo son. Se hace el conteo de los resultados obtenidos y se asume que se esperaba que todas las respuestas fueran iguales.

Para determinar si los resultados son similares o no, se recurre a la prueba de hipótesis, para comprobar si existen diferencia o no entre lo observado y lo esperado.

Se siguen los mismos pasos que para estadísticos anteriores que son:

1. Establecer la hipótesis nula y la alternativa
2. Seleccionar un nivel de significancia
3. Seleccionar el estadístico de prueba
4. Formular la regla de decisión
5. Tomar una decisión

## *Ejemplo 7.2*

1. La gerente de marketing de un fabricante de tarjetas deportivas planea iniciar la venta de una serie de tarjetas deportivas con fotografías y estadística de juego de la liga nacional. Uno delos problemas es la selección de exjugadores. En una exhibición de tarjetas de futbol en el Estadio Morazán el pasado fin de semana se instaló un puesto y ofreció tarjetas de Danilo Toselo, Carlos Pavón, Carlo Costly, Diego Vásquez, Wilmer Velásquez y Rambo de León. Al final del día vendió 120 tarjetas. El número de tarjetas vendidas de cada jugador es la siguiente:

| # | Jugador | Tarjetas Vendidas |
|---|---|---|
| 1 | Danilo Toselo | 13 |
| 2 | Carlos Pavón | 33 |
| 3 | Carlo Costly | 14 |
| 4 | Diego Vásquez | 7 |
| 5 | Wilmer Velásquez | 36 |
| 6 | Rambo de León | 17 |

Si la importancia de los jugadores es similar, debería haberse vendido la misma cantidad de cada uno de ellos; sin embargo, podría suceder que el muestreo haya generado un sesgo; pero, que la población sí mantenga las mismas preferencias.

| # | Jugador | Tarjetas Vendidas | Pronóstico |
|---|---|---|---|
| 1 | Danilo Toselo | 13 | 20 |
| 2 | Carlos Pavón | 33 | 20 |
| 3 | Carlo Costly | 14 | 20 |
| 4 | Diego Vásquez | 7 | 20 |
| 5 | Wilmer Velásquez | 36 | 20 |
| 6 | Rambo de León | 17 | 20 |
| | Total tarjetas | 120 | 120 |

La información obtenida durante la venta piloto es la "frecuencia observada" y se esperaba que las 120 tarjetas vendidas se distribuyeran de manera uniforme entre todos los jugadores; esta la "frecuencia esperada".

¿Se puede determinar que hay diferencia entre las tarjetas vendidas y las esperadas, con un nivel de significancia del 5%?

UNIVERSIDAD DE SAN PEDRO SULA

## Desarrollo

- Paso 1: Hipótesis nula y alternativa

$$H_0: f_0 \leq f_a$$
$$H_0: f_0 > f_a$$

- Paso 2: Nivel de significancia

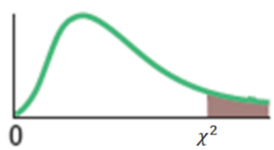
$$\alpha = 0.05$$

- Paso 3: Estadístico de prueba

$$\chi^2 = \sum \frac{(f_0 - f_e)^2}{f_e}$$

- Paso 4: Regla de decisión

$$\alpha = 0.05$$
$$1 cola$$
$$gl = 6 - 1 = 5$$




| Grados de libertad, *gl* | Área de cola derecha 0.10 | 0.05 | 0.02 | 0.01 |
|---|---|---|---|---|
| 1 | 2.706 | 3.841 | 5.412 | 6.635 |
| 2 | 4.605 | 5.991 | 7.824 | 9.210 |
| 3 | 6.251 | 7.815 | 9.837 | 11.345 |
| 4 | 7.779 | 9.488 | 11.668 | 13.277 |
| 5 | 9.236 | 11.070 | 13.388 | 15.086 |

Valor crítico $\chi^2 = 11.07$

- **Paso 5:** Toma de decisión

| # | Jugador | Tarjetas Vendidas $f_o$ | Pronóstico $f_e$ | $f_0$ - $f_e$ | | $(f_0 - f_e)^2$ | $\frac{(f_0 - f_e)^2}{fe}$ |
|---|---|---|---|---|---|---|---|
| 1 | Danilo Toselo | 13 | 20 | 13 - 20 = | -7 | 49 | 2.45 |
| 2 | Carlos Pavón | 33 | 20 | 33 - 20 = | 13 | 169 | 8.45 |
| 3 | Carlo Costly | 14 | 20 | 14 - 20 = | -6 | 36 | 1.8 |
| 4 | Diego Vásquez | 7 | 20 | 7 - 20 = | -13 | 169 | 8.45 |
| 5 | Wilmer Velásquez | 36 | 20 | 36 - 20 = | 16 | 256 | 12.8 |
| 6 | Rambo de León | 17 | 20 | 17 - 20 = | -3 | 9 | 0.45 |
| | Total tarjetas | 120 | 120 | | | | 34.4 |

$$\chi^2 = 34.4$$

La hipótesis nula se rechaza.

Se puede concluir que sí hay diferencia entre las tarjetas vendidas y las tarjetas pronosticas.

Es improbable que las ventas de las tarjetas sean las mismas entre los 6 jugadores.

2. La directora de recursos humanos de Georgetown Paper, Inc., está preocupada por el ausentismo entre los trabajadores por hora, por lo que decide tomar una muestra de los registros de la compañía y determinar si el ausentismo está distribuido de manera uniforme en toda la semana de seis días. Las hipótesis son:

   $H_0$: El ausentismo está distribuido de manera uniforme en toda la semana de trabajo.

   $H_1$: El ausentismo no está distribuido de manera uniforme en toda la semana de trabajo.

| Día | Número de ausencias |
|---|---|
| Lunes | 12 |
| Martes | 9 |
| Miércoles | 11 |
| Jueves | 10 |
| Viernes | 9 |
| Sábado | 9 |

Con una significancia del 1% ¿qué le indican los resultados a la directora de recursos humanos?

## Desarrollo

- **Paso 1: Hipótesis nula y alternativa**

$$H_0: f_0 \leq f_a$$
$$H_1: f_0 > f_a$$

- Paso 2: Nivel de significancia

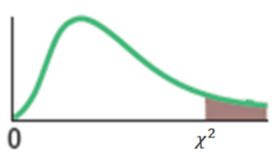
$$\alpha = 0.01$$

- Paso 3: Estadístico de prueba

$$\chi^2 = \sum \frac{(f_0 - f_e)^2}{f_e}$$

- Paso 4: Regla de decisión

$$H_0: f_0 \leq f_a$$
$$\alpha = 0.01$$
$$1\ cola$$
$$n = 6$$
$$gl = 6 - 1 = 5$$




| Grados de libertad, *gl* | Área de cola derecha 0.10 | 0.05 | 0.02 | 0.01 |
|---|---|---|---|---|
| 1 | 2.706 | 3.841 | 5.412 | 6.635 |
| 2 | 4.605 | 5.991 | 7.824 | 9.210 |
| 3 | 6.251 | 7.815 | 9.837 | 11.345 |
| 4 | 7.779 | 9.488 | 11.668 | 13.277 |
| 5 | 9.236 | 11.070 | 13.388 | 15.086 |

Valor crítico $\chi^2 = 15.086$

- **Paso 5:** Toma de decisión

| Día | Número de ausencias | Pronóstico $f_e$ | $f_0 - f_e$ | | $(f_0 - f_e)^2$ | $\frac{(f_0 - f_e)^2}{f_e}$ |
|---|---|---|---|---|---|---|
| Lunes | 12 | 10 | 12 - 10 = | 2 | 4 | 0.4 |
| Martes | 9 | 10 | 9 - 10 = | -1 | 1 | 0.1 |
| Miércoles | 11 | 10 | 11 - 10 = | 1 | 1 | 0.1 |
| Jueves | 10 | 10 | 10 - 10 = | 0 | 0 | 0 |
| Viernes | 9 | 10 | 9 - 10 = | -1 | 1 | 0.1 |
| Sábado | 9 | 10 | 9 - 10 = | -1 | 1 | 0.1 |
| | 60 | 60 | | | | 0.8 |

$$\chi^2 = 0.80$$

USAP
UNIVERSIDAD DE SAN PEDRO SULA

**La hipótesis nula se acepta.**
**El ausentismo se distribuye de manera uniforme durante la semana.**

## Prueba de bondad de ajuste: Frecuencias esperadas desiguales

Si las frecuencias esperadas son desiguales, se debe definir los valores exactos para cada una de las características de la variable o en su lugar, la ponderación correspondiente.

Los pasos a seguir seguirán siendo los mismos:

### *Ejemplo 7.3*

1. Según las políticas de una aseguradora, las atenciones en los hospitales para el ingreso de los adultos mayores en el período de un año, tiene el siguiente comportamiento.
   40% no requiere hospitalización
   30% es hospitalizado una vez
   20% es hospitalizado dos veces
   10% es hospitalizado 3 o más veces

   Una encuesta de 150 adultos mayores, miembros de la aseguradora reveló que:
   55 no requirieron hospitalización
   50 fueron admitidos una vez
   32 fueron admitidos dos veces
   13 fueron admitidos tres o más veces

   Con un nivel de significancia de 0.05, probar si es posible que los resultados sean congruentes con las ponderaciones.

### Desarrollo

Construir la tabla base para iniciar el análisis.

| Admisión | Encuestados | Ponderación poblacional |
|---|---|---|
| No requiere hospitalización | 55 | 40% |
| Hospitalizado 1 vez | 50 | 30% |
| Hospitalizado 2 veces | 32 | 20% |
| Hospitalizado 3 o más veces | 13 | 10% |

**PASO 1: Hipótesis nula y alternativa**

$$\boldsymbol{H_0: f_0 \leq f_a}$$
$$\boldsymbol{H_0: f_0 > f_a}$$

**PASO 2: Nivel de significancia**

$$\boldsymbol{\alpha = 0.05}$$

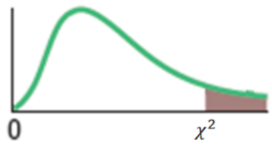
**PASO 3: Estadístico de prueba**

$$\chi^2 = \sum \frac{(f_0 - f_e)^2}{f_e}$$

**PASO 4: Regla de decisión**

$$\boldsymbol{\alpha = 0.05}$$
$$\boldsymbol{1cola}$$
$$\boldsymbol{gl = 4 - 1 = 3}$$

UNIVERSIDAD DE SAN PEDRO SULA



| Grados de libertad, *gl* | Área de cola derecha | | | |
|---|---|---|---|---|
| | 0.10 | 0.05 | 0.02 | 0.01 |
| 1 | 2.706 | 3.841 | 5.412 | 6.635 |
| 2 | 4.605 | 5.991 | 7.824 | 9.210 |
| 3 | 6.251 | 7.815 | 9.837 | 11.345 |

Valor crítico $\chi^2 = \mathbf{7.815}$

**PASO 5: Toma de decisión**

- Calcular las frecuencias esperadas

| **Admisión** | **Encuestados $f_o$** | **Ponderación poblacional** | **Esperados $f_e$** |
|---|---|---|---|
| No requiere hospitalización | 55 | 40% | 60 |
| Hospitalizado 1 vez | 50 | 30% | 45 |
| Hospitalizado 2 veces | 32 | 20% | 30 |
| Hospitalizado 3 o más veces | 13 | 10% | 15 |
| | 150 | | |

- Calcular el cociente entre la variación cuadrada y la frecuencia esperada.

| **Admisión** | **Encuestados $f_o$** | **Ponderación poblacional** | **Esperados $f_e$** | **$f_0 - f_e$** | $\frac{(f_0 - f_e)^2}{f_e}$ |
|---|---|---|---|---|---|
| No requiere hospitalización | 55 | 40% | 60 | -5 | 0.417 |
| Hospitalizado 1 vez | 50 | 30% | 45 | 5 | 0.556 |
| Hospitalizado 2 veces | 32 | 20% | 30 | 2 | 0.133 |
| Hospitalizado 3 o más veces | 13 | 10% | 15 | -2 | 0.267 |
| | 150 | | 150 | | 1.372 |

$$\chi^2 = \mathbf{1.372}$$

La hipótesis nula se acepta.
No hay evidencia de una diferencia entre la población y la muestra.
No hay diferencia entre los que son hospitalizados y los que no lo son.

2. El departamento de Ventas del centro comercial Mall Oriental sabe que el 10% de sus clientes son clase media de la zona rural, el 20% son ganaderos, 25% empleados de empresas grandes y 45% empleados de empleados de pequeña y mediana empresa. De los 1000 clientes a los que se les entregó un cupón con el 20% de descuento en todas las tiendas del centro comercial, 100 son de clase media de la zona rural, 240 de ganaderos, 360 de empresas grandes y 300 de empresa mediana y pequeña. ¿Se puede concluir que la distribución de los clientes que recibieron el cupón es diferente de las demás, con un nivel de significancia de 0.01?

## Desarrollo

| TIPO DE CLIENTE | CUPONES | % |
|---|---|---|
| Personas naturales | 200 | 10% |
| Media de zona rural | 100 | 8% |
| Ganaderos | 240 | 16% |
| Empresas grandes | 160 | 20% |
| Empresas medianas y pequeñas | 300 | 46% |
| Totales... | 1,000 | 100% |

**PASO 1: Hipótesis nula y alternativa**

$$H_0: f_0 \leq f_a$$
$$H_0: f_0 > f_a$$

**PASO 2: Nivel de significancia**

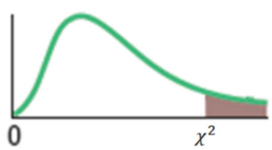
$$\alpha = 0.01$$

**PASO 3: Estadístico de prueba**

$$\chi^2 = \sum \frac{(f_0 - f_e)^2}{f_e}$$

**PASO 4: Regla de decisión**

$$\alpha = 0.01$$
$$1cola$$
$$gl = 5 - 1 = 4$$




| Grados de libertad, *gl* | Área de cola derecha 0.10 | 0.05 | 0.02 | 0.01 |
|---|---|---|---|---|
| 1 | 2.706 | 3.841 | 5.412 | 6.635 |
| 2 | 4.605 | 5.991 | 7.824 | 9.210 |
| 3 | 6.251 | 7.815 | 9.837 | 11.345 |
| 4 | 7.779 | 9.488 | 11.668 | 13.277 |
| 5 | 9.236 | 11.070 | 13.388 | 15.086 |

Valor crítico $\chi^2 = 13.277$

**PASO 5: Toma de decisión**

| TIPO DE CLIENTE | CUPONES $f_o$ | % | $f_e$ | $\frac{(f_o - f_e)^2}{f_e}$ |
|---|---|---|---|---|
| Personas naturales | 200 | 10% | 100 | 100.0 |
| Media de zona rural | 100 | 8% | 80 | 5.0 |
| Ganaderos | 240 | 16% | 160 | 40.0 |
| Empresas grandes | 160 | 20% | 200 | 8.0 |
| Empresas medianas y pequeñas | 300 | 46% | 460 | 55.7 |
| Totales… | 1,000 | 100% | 1,000 | 208.7 |

La hipótesis nula se rechaza.
Hay suficiente evidencia que indica que la cantidad de cupones entregados es diferente a la proporción de la población que es cliente del centro comercial.

## Tablas de contingencia

Se conoce como tabla de contingencia a toda distribución de frecuencias formada por dos variables cualitativas, también conocidos como datos bivariados o variables cruzadas.

| Variable 1 | Variable 2 | | | Total |
|---|---|---|---|---|
| | Característica 1 | Característica 2 | Característica 3 | |
| Característica 1 | | | | |
| Característica 2 | | | | |
| Característica 3 | | | | |
| Característica 4 | | | | |
| Totales … | | | | |

Una tabla de contingencia provee información variada, puesto que, proporciona datos en forma conjunta y en forma individual para cada variable, todo con una misma muestra.

USAP
UNIVERSIDAD DE SAN PEDRO SULA

## *Ejemplo 7.4*

1. En una investigación sobre los gustos de las personas en la prueba de un nuevo jugo combinado, se hizo diferencia entre las respuestas de los clientes del sexo femenino con las del masculino. Se encuestaron 100 personas que proporcionaron las siguientes respuestas:

| SABOR DEL JUGO | GÉNERO | | Total |
|---|---|---|---|
| | Masculino | Femenino | |
| Muy bueno | 20 | 15 | 35 |
| Bueno | 15 | 17 | 32 |
| Regular | 18 | 8 | 26 |
| Indeciso | 5 | 2 | 7 |
| Totales ... | 58 | 42 | 100 |

### Desarrollo

Análisis de la pregunta de cómo les había parecido el sabor del jugo, a 35 de ellos les pareció muy bueno el sabor y solamente 7 se mostraron indecisos. Los hombres que encontraron el sabor del jugo muy bueno fueron 20, lo que representa que son los que tienen mayor porcentaje de respuesta con respecto a todos los encuestados.

2. Vilma Cho es la directora de investigación de una empresa de productos químicos. En su proyecto actual Vilma debe determinar si existe alguna relación entre la clasificación de efectividad que los consumidores asignan a un nuevo insecticida y el sitio (urbano o rural) en el cual se utiliza. De los 150 consumidores a los que se les aplicó la encuesta, 95 vivían en la zona urbana y 55 en la rural, según se muestra en la tabla de dos variables.

| SITIO | URBANO | RURAL | TOTAL |
|---|---|---|---|
| Muy bueno | 30 | 20 | 50 |
| Regular | 50 | 15 | 65 |
| Malo | 15 | 20 | 35 |
| TOTAL | 95 | 55 | 150 |

### Desarrollo

La tabla está formada por dos columnas que indican que 50 de los consumidores describieron el producto como muy bueno, 65 como regular y 35 como malo. Los consumidores que viven en la zona urbana y que lo describieron como regular fueron 50, estos consumidores representan el porcentaje más alto de toda la muestra.

## Tablas de contingencia y chi-cuadrada

En una investigación con variables cruzadas, se analizan dos variables a las cuales se les calculará el valor de chi-cuadrada tomando de base una ponderación definida que pueden ser iguales o diferentes. Es de hacer notar que las características que se evalúan pueden ser variadas.

UNIVERSIDAD DE SAN PEDRO SULA

| VARIABLE 1 | VARIABLE 2 | | | | Total |
|---|---|---|---|---|---|
| | Carácterística 1 | | Carácterística 2 | | |
| | $f_o$ | $f_e$ | $f_o$ | $f_e$ | |
| Característica 1 | | | | | |
| Característica 2 | | | | | |
| Característica 3 | | | | | |
| **Totales ...** | **100** | **100** | **200** | **200** | **300** |

## *Ejemplo 7.5*

1. En una investigación piloto, con relación a la percepción de los consumidores sobre el sabor de un nuevo jugo que se quiere lanzar al mercado, se esperaba que el 60% contestara uniformemente que lo encontraba muy bueno o bueno, y que el resto lo encontrara regular o estuviera indeciso. Calcular el valor de chi-cuadrada para la distribución obtenida:

| SABOR DEL JUGO | GÉNERO | | Total |
|---|---|---|---|
| | Masculino | Femenino | |
| Muy bueno | 20 | 15 | 35 |
| Bueno | 15 | 17 | 32 |
| Regular | 18 | 8 | 26 |
| Indecisto | 5 | 2 | 7 |
| Totales ... | 58 | 42 | 100 |

### Desarrollo

- **Género:** **Masculino**
  a. Calcular el 30% esperado para los que respondieron “muy bueno” o “bueno” y el 20% para los que respondieron “regular” o “indeciso”.

| SABOR DEL JUGO | Masculino $f_o$ | $f_e$ |
|---|---|---|
| Muy bueno | 20 | 17.4 |
| Bueno | 15 | 17.4 |
| Regular | 18 | 11.6 |
| Indecisto | 5 | 11.6 |
| Totales ... | 58 | 58.0 |

  b. Calcular variación cuadrada entre el valor de la frecuencia observada y la esperada

| SABOR DEL JUGO | Masculino $f_o$ | $f_e$ | $(f_o - f_e)^2$ |
|---|---|---|---|
| Muy bueno | 20 | 17.4 | 6.76 |
| Bueno | 15 | 17.4 | 5.76 |
| Regular | 18 | 11.6 | 40.96 |
| Indecisto | 5 | 11.6 | 43.56 |
| Totales ... | 58 | 58 | |

  c. Calcular el cociente entre la variación cuadrada y la frecuencia esperada

USAP UNIVERSIDAD DE SAN PEDRO SULA

| SABOR DEL JUGO | Masculino $f_o$ | $f_e$ | $(f_o - f_e)^2$ | $\frac{(f_o - f_e)^2}{f_e}$ |
|---|---|---|---|---|
| Muy bueno | 20 | 17.4 | 6.76 | 0.3885 |
| Bueno | 15 | 17.4 | 5.76 | 0.3310 |
| Regular | 18 | 11.6 | 40.96 | 3.5310 |
| Indecisto | 5 | 11.6 | 43.56 | 3.7552 |
| Totales ... | 58 | 58 | | 8.0057 |

- **Género: Femenino**

d. Calcular el 30% esperado para los que respondieron "muy bueno" o "bueno" y el 20% para los que respondieron "regular" o "indeciso".

| SABOR DEL JUGO | Femenino $f_o$ | $f_e$ |
|---|---|---|
| Muy bueno | 15 | 12.6 |
| Bueno | 17 | 12.6 |
| Regular | 8 | 8.4 |
| Indecisto | 2 | 8.4 |
| Totales ... | 42 | 42 |

e. Calcular la variación cuadrada entre el valor de la frecuencia observada y la esperada.

| SABOR DEL JUGO | Femenino $f_o$ | $f_e$ | $(f_o - f_e)^2$ |
|---|---|---|---|
| Muy bueno | 15 | 12.6 | 5.76 |
| Bueno | 17 | 12.6 | 19.36 |
| Regular | 8 | 8.4 | 0.16 |
| Indecisto | 2 | 8.4 | 40.96 |
| Totales ... | 42 | 42 | |

f. Calcular el cociente entre la variación cuadrada y la frecuencia esperada.

| SABOR DEL JUGO | Femenino $f_o$ | $f_e$ | $(f_o - f_e)^2$ | $\frac{(f_o - f_e)^2}{f_e}$ |
|---|---|---|---|---|
| Muy bueno | 15 | 12.6 | 5.76 | 0.4571 |
| Bueno | 17 | 12.6 | 19.36 | 1.5365 |
| Regular | 8 | 8.4 | 0.16 | 0.0190 |
| Indecisto | 2 | 8.4 | 40.96 | 4.8762 |
| Totales ... | 42 | 42 | | 6.8889 |

g. Unir el cociente de ambos géneros

| SABOR DEL JUGO | Masculino $\frac{(f_o - f_e)^2}{f_e}$ | Femenino $\frac{(f_o - f_e)^2}{f_e}$ |
|---|---|---|
| Muy bueno | 0.3885 | 0.4571 |
| Bueno | 0.3310 | 1.5365 |
| Regular | 3.5310 | 0.0190 |
| Indecisto | 3.7552 | 4.8762 |
| Totales ... | 8.0057 | 6.8889 |

h. Sumar los totales de ambas característica

USAP
UNIVERSIDAD DE SAN PEDRO SULA

| SABOR DEL JUGO | Masculino $\frac{(f_o - f_e)^2}{f_e}$ | Femenino $\frac{(f_o - f_e)^2}{f_e}$ | |
|---|---|---|---|
| Muy bueno | 0.3885 | 0.4571 | |
| Bueno | 0.3310 | 1.5365 | |
| Regular | 3.5310 | 0.0190 | |
| Indecisto | 3.7552 | 4.8762 | |
| Totales ... | 8.0057 | 6.8889 | 14.895 |

$$\chi^2 = 14.895$$

2. Vilma Cho es la directora de investigación de una empresa de productos químicos. En su proyecto actual Vilma debe determinar si existe alguna relación entre la clasificación de efectividad que los consumidores asignan a un nuevo insecticida y el sitio (urbano o rural) en el cual se utiliza, según los registros históricos, el 50% de los consumidores lo clasificaron como muy bueno, 35% como regular y 15% como malo. Calcular el valor de chi-cuadrada para la distribución obtenida:

| CALIFICACIÓN | URBANO | RURAL | TOTAL |
|---|---|---|---|
| Muy bueno | 30 | 20 | 50 |
| Regular | 50 | 15 | 65 |
| Malo | 15 | 20 | 35 |
| TOTAL | 95 | 55 | 150 |

## Desarrollo

Determinar la tabla con las ponderaciones de la población:

| CALIFICACIÓN | URBANO $f_o$ | RURAL $f_o$ | % |
|---|---|---|---|
| Muy bueno | 30 | 20 | 50% |
| Regular | 50 | 15 | 35% |
| Malo | 15 | 20 | 15% |
| TOTAL | 95 | 55 | 100% |

Calcular el cociente de las variaciones cuadradas y la frecuencia esperada.

| CALIFICACIÓN | URBANO $f_o$ | RURAL $f_o$ | URBANO $f_e$ | RURAL $f_e$ | URBANO $\frac{(f_o - f_e)^2}{f_e}$ | RURAL $\frac{(f_o - f_e)^2}{f_e}$ |
|---|---|---|---|---|---|---|
| Muy bueno | 30 | 20 | 47.5 | 27.5 | 6.447 | 2.045 |
| Regular | 50 | 15 | 33.25 | 19.25 | 8.438 | 0.938 |
| Malo | 15 | 20 | 14.25 | 8.25 | 0.039 | 16.735 |
| TOTAL | 95 | 55 | 95.00 | 55.00 | 14.925 | 19.719 |

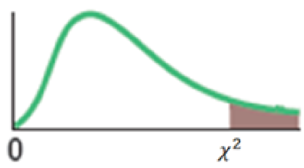
$\chi^2 = 14.925 + 19.719 = 34.64$

## Análisis de Tablas de contingencia

Existen investigaciones en las cuales se desea conocer el comportamiento con poblaciones que trabajan en conjuntos interceptados. Esta clasificación tiene como base la escala nominal debido a que no hay un orden natural para las clasificaciones.

El estadístico ji cuadrada es útil para probar de manera formal si hay una relación entre dos variables con escala nominal. En otras palabras, ¿es independiente una variable de la otra?

| Grados de libertad, *gl* | Área de cola derecha | | | |
|---|---|---|---|---|
| | 0.10 | 0.05 | 0.02 | 0.01 |
| 1 | 2.706 | 3.841 | 5.412 | 6.635 |
| 2 | 4.605 | 5.991 | 7.824 | 9.210 |
| 3 | 6.251 | 7.815 | 9.837 | 11.345 |
| 4 | 7.779 | 9.488 | 11.668 | 13.277 |
| 5 | 9.236 | 11.070 | 13.388 | 15.086 |
| 6 | 10.645 | 12.592 | 15.033 | 16.812 |
| 7 | 12.017 | 14.067 | 16.622 | 18.475 |
| 8 | 13.362 | 15.507 | 18.168 | 20.090 |
| 9 | 14.684 | 16.919 | 19.679 | 21.666 |
| 10 | 15.987 | 18.307 | 21.161 | 23.209 |




Es de hacer notar que al tener dos variables, los grados de libertad deben considerar las dos características y el producto de ambas variaciones lo determina.

$$gl = (filas - 1)(columnas - 1)$$

PASO 1: Hipótesis nula y alternativa

PASO 2: Nivel de significancia

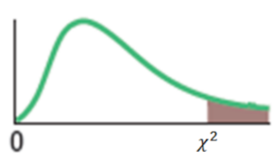
PASO 3: Estadístico de prueba

PASO 4: Regla de decisión

$\boldsymbol{\alpha = x.xx}$

$\mathbf{1}\ \boldsymbol{cola}$

$\boldsymbol{gl = (fila - 1)(columna - 1)}$




| Grados de libertad, *gl* | Área de cola derecha | | | |
|---|---|---|---|---|
| | 0.10 | 0.05 | 0.02 | 0.01 |
| 1 | 2.706 | 3.841 | 5.412 | 6.635 |
| 2 | 4.605 | 5.991 | 7.824 | 9.210 |
| 3 | 6.251 | 7.815 | 9.837 | 11.345 |

PASO 5: Toma de decisión

## *Ejemplo 7.6*

1. Una organización no gubernamental que trabaja con proyectos de inserción del privado de la libertad está investigando si los que recuperan su libertad tienen un mejor nivel de readaptación a la vida civil cuando lo hacen en la localidad que vivieron o en otra localidad que no conocen de antemano; es decir, Con una confiabilidad del 99%, ¿Hay una relación entre la adaptación a la vida civil en su localidad conocida o en una desconocida según los datos obtenidos en el 2013?

USAP UNIVERSIDAD DE SAN PEDRO SULA

En una encuesta realizada por psicólogos a 200 personas, 120 de las cuales residían en su localidad natal y 80 en otra localidad, se obtuvieron se los siguientes resultados:

| Residencia al salir de la prisión | Adaptación a la vida civil | | | | Total |
|---|---|---|---|---|---|
| | Sobresaliente | Buena | Regular | Insatisfactoria | |
| Localidad natal | 27 | 35 | 33 | 25 | 120 |
| En otra localidad | 13 | 15 | 27 | 25 | 80 |
| **Total** | **40** | **50** | **60** | **50** | **200** |

Tomar como base la ponderación del total según el tipo de residencia.

## Desarrollo

**PASO 1: Hipótesis nula y alternativa**

$$H_0: f_0 \leq f_a$$
$$H_0: f_0 > f_a$$

**PASO 2: Nivel de significancia**

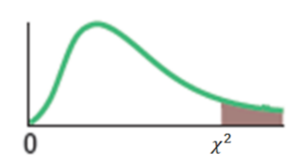
$$\alpha = 0.01$$

**PASO 3: Estadístico de prueba**

$$\chi^2 = \sum \frac{(f_0 - f_e)^2}{f_e}$$

**PASO 4: Regla de decisión**

$$\alpha = 0.01$$
$$1\ cola$$
$$gl = (2-1)(4-1) = 3$$




| Grados de libertad, *gl* | Área de cola derecha | | | |
|---|---|---|---|---|
| | 0.10 | 0.05 | 0.02 | 0.01 |
| 1 | 2.706 | 3.841 | 5.412 | 6.635 |
| 2 | 4.605 | 5.991 | 7.824 | 9.210 |
| 3 | 6.251 | 7.815 | 9.837 | 11.345 |

$$\chi^2 = 11.345$$

**PASO 5: Toma de decisión**

- Calcular el porcentaje que corresponde a cada uno de los tipos de residencia.
  Localidad natal: Porcentaje de 120 con relación a 200 = 120/200*100 = 60%
  En otra localidad: Porcentaje de 80 con relación a 200 = 80/200*100 = 40%

| Residencia depués de salir | Total | Ponderación | | % |
|---|---|---|---|---|
| Localidad natal | 120 | 120 / 200 = | 0.6 | 60% |
| En otra localidad | 80 | 80 / 200 = | 0.4 | 40% |
| Total | 200 | | 1.0 | 100% |

- **Calcular la frecuencia esperada en base a los porcentajes definidos.**

| Residencia depués de salir | Sobre-saliente | $f_e$ | |
|---|---|---|---|
| Localidad natal | 27 | 40 * 60% = | 24 |
| En otra localidad | 13 | 40 * 40% = | 16 |
| Total | 40 | | 40 |

**Igual para las demás características**

USAP
UNIVERSIDAD DE SAN PEDRO SULA

| Residencia depués de salir de la prisión | Adaptación a la vida civil ($f_o$) | | | | | Adaptación a la vida civil ($f_e$) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Sobre-saliente | Buena | Regular | Insatis-factoria | Total | Sobre-saliente | Buena | Regular | Insatis-factoria | Total |
| Localidad natal | 27 | 35 | 33 | 25 | 120 | 24 | 30 | 36 | 30 | 120 |
| En otra localidad | 13 | 15 | 27 | 25 | 80 | 16 | 20 | 24 | 20 | 80 |
| Total | 40 | 50 | 60 | 50 | 200 | 40 | 50 | 60 | 50 | 200 |

- **Calcular la variación cuadrada sobre la frecuencia esperada y sumar todos los resultados:**
  - **Característica "sobresaliente"**

| depués de salir de la prisión | Sobresaliente | | $(f_o - f_e)^2 / f_e$ |
|---|---|---|---|
| | $f_o$ | $f_e$ | |
| Localidad natal | 27 | 24 | $(27 - 24)^2 / 24 =$ 0.3750 |
| En otra localidad | 13 | 16 | $(13 - 16)^2 / 16 =$ 0.5625 |
| Total | 40 | 40 | 0.9375 |

  - **Característica "Buena"**

| depués de salir de la prisión | Buena | | $(f_o - f_e)^2 / f_e$ |
|---|---|---|---|
| | $f_o$ | $f_e$ | |
| Localidad natal | 35 | 30 | $(35 - 30)^2 / 30 =$ 0.8333 |
| En otra localidad | 15 | 20 | $(15 - 20)^2 / 20 =$ 1.2500 |
| Total | 50 | 50 | 2.0833 |

  - **Característica "Regular"**

| depués de salir de la prisión | Regular | | $(f_o - f_e)^2 / f_e$ |
|---|---|---|---|
| | $f_o$ | $f_e$ | |
| Localidad natal | 33 | 36 | $(33- 36)^2 / 36 =$ 0.2500 |
| En otra localidad | 27 | 24 | $(27 - 24)^2 / 24 =$ 0.3750 |
| Total | 60 | 60 | 0.6250 |

  - **Característica "Insatisfactoria"**

| depués de salir de la prisión | Insatisfactoria | | $(f_o - f_e)^2 / f_e$ |
|---|---|---|---|
| | $f_o$ | $f_e$ | |
| Localidad natal | 25 | 30 | $(25 - 30)^2 / 30 =$ 0.8333 |
| En otra localidad | 25 | 20 | $(25 - 20)^2 / 20 =$ 1.2500 |
| Total | 50 | 50 | 2.08333 |

- **Sumar las sumatoria de cada columna**

| Residencia depués de salir de la prisión | Adaptación a la vida civil ($f_o$) | | | |
|---|---|---|---|---|
| | Sobre-saliente | Buena | Regular | Insatis-factoria |
| Localidad natal | 0.3750 | 0.8333 | 0.2500 | 0.8333 |
| En otra localidad | 0.5625 | 1.2500 | 0.3750 | 1.2500 |
| Total | 0.9375 | 2.0833 | 0.6250 | 2.0833 |
| Σ | 5.7292 | | | |

$$\chi^2 = \mathbf{5.729}$$

La regla de decisión indica que el valor crítico es 11.345 y el resultado de la muestra es 5.729.
La hipótesis nula se acepta

USAP
UNIVERSIDAD DE SAN PEDRO SULA

No hay evidencia de una diferencia entre la adaptación en la localidad o en una desconocida.

2. Una empresa de investigación de mercados ha sido contratada por la municipalidad de La Ceiba para realizar un proyecto para la construcción de una concha acústica en el casco histórico de la ciudad. Con una confiabilidad del 95% se puede establecer si hay diferencia entre la opinión de los hombres de las mujeres. Se levantó un encuesta en donde 300 fueron completadas por hombre y 200 por mujeres, obteniendo los siguientes resultados en una tabla de dos variables:

| GÉNERO | MUY BUENO $f_o$ | BUENO $f_o$ | MALO $f_o$ | INDECISO $f_o$ | NO CONTESTÓ $f_o$ | TOTAL |
|---|---|---|---|---|---|---|
| Hombres | 70 | 44 | 86 | 75 | 25 | 300 |
| Mujeres | 97 | 76 | 17 | 2 | 8 | 200 |
| Total | 167 | 120 | 103 | 77 | 33 | 500 |

## Desarrollo

**PASO 1: Hipótesis nula y alternativa**

$$H_0: f_0 \leq f_a$$
$$H_0: f_0 > f_a$$

**PASO 2: Nivel de significancia**

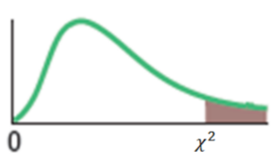
$$\alpha = 0.05$$

**PASO 3: Estadístico de prueba**

$$\chi^2 = \sum \frac{(f_0 - f_e)^2}{f_e}$$

**PASO 4: Regla de decisión**

$$\alpha = 0.05$$
$$1\ cola$$
$$gl = (2-1)(5-1) = 4$$




| Grados de libertad, $gl$ | Área de cola derecha 0.10 | 0.05 | 0.02 | 0.01 |
|---|---|---|---|---|
| 1 | 2.706 | 3.841 | 5.412 | 6.635 |
| 2 | 4.605 | 5.991 | 7.824 | 9.210 |
| 3 | 6.251 | 7.815 | 9.837 | 11.345 |
| 4 | 7.779 | 9.488 | 11.668 | 13.277 |
| 5 | 9.236 | 11.070 | 13.388 | 15.086 |

$$\chi^2 = 9.488$$

**PASO 5: Toma de decisión**

Definición de porcentajes por género

| GÉNERO | TOTAL | % |
|---|---|---|
| Hombres | 300 | 60% |
| Mujeres | 200 | 40% |
| Total | 500 | 100% |

Frecuencia observada en cada categoría.

USAP
UNIVERSIDAD DE SAN PEDRO SULA

| GÉNERO | MUY BUENO $f_o$ | BUENO $f_o$ | MALO $f_o$ | INDECISO $f_o$ | NO CONTESTÓ $f_o$ |
|---|---|---|---|---|---|
| Hombres | 70 | 44 | 86 | 75 | 25 |
| Mujeres | 97 | 76 | 17 | 2 | 8 |
| Total | 167 | 120 | 103 | 77 | 33 |

Frecuencia esperada en cada categoría

| GÉNERO | MUY BUENO $f_e$ | BUENO $f_e$ | MALO $f_e$ | INDECISO $f_e$ | NO CONTESTÓ $f_e$ |
|---|---|---|---|---|---|
| Hombres | 100.2 | 72.0 | 61.8 | 46.2 | 19.8 |
| Mujeres | 66.8 | 48.0 | 41.2 | 30.8 | 13.2 |
| Total | 167.0 | 120.0 | 103.0 | 77.0 | 33.0 |

Cálculo de chi-cuadrada

| GÉNERO | MUY BUENO $\frac{(f_o - f_e)^2}{f_e}$ | BUENO $\frac{(f_o - f_e)^2}{f_e}$ | MALO $\frac{(f_o - f_e)^2}{f_e}$ | INDECISO $\frac{(f_o - f_e)^2}{f_e}$ | NO CONTESTÓ $\frac{(f_o - f_e)^2}{f_e}$ | |
|---|---|---|---|---|---|---|
| Hombres | 9.10 | 10.89 | 9.48 | 17.95 | 1.37 | |
| Mujeres | 13.65 | 16.33 | 14.21 | 26.93 | 2.05 | |
| Total | **22.76** | **27.22** | **23.69** | **44.88** | **3.41** | **121.97** |

$$\chi^2 = 121.97$$

La hipótesis nula se rechaza.
Si hay diferencia en las respuestas de los hombres y las de las mujeres.

# Ejercicio

1. En una prueba de bondad de ajuste de ji cuadrada hay cuatro categorías y 200 observaciones. Utilice el nivel de significancia 0.05.
   a. ¿Cuántos grados de libertad hay?
   b. ¿Cuál es el valor crítico de ji cuadrada?
2. La hipótesis nula y la alternativa son:
   $H_0$ : Las frecuencias observadas son menores que las esperadas
   $H_a$ : Las frecuencias observadas no son iguales a las esperadas
   Los datos de la muestra obtenida son los siguientes:

| Categoría | $f_o$ |
|---|---|
| A | 10 |
| B | 20 |
| C | 30 |
| Total | 60 |

¿Cuál es su decisión con respecto a $H_0$?

UNIVERSIDAD DE SAN PEDRO SULA

3. Un dado se lanza 30 veces y los números 1 a 6 aparecen como muestra la siguiente distribución de frecuencia. Con un nivel de significancia de 0.10, ¿es posible concluir que el dado no está cargado? Con la prueba chi-cuadrada.

| Resultado | Frecuencia | Resultado | Frecuencia |
|---|---|---|---|
| 1 | 3 | 4 | 3 |
| 2 | 6 | 5 | 9 |
| 3 | 2 | 6 | 7 |

4. En un estudio que supone que el 40% de las observaciones se encuentran en la categoría A, el 40% en la categoría B y el 20% en la C. Una muestra de 60 observaciones dio los siguientes resultados:

| Categoría | $f_o$ |
|---|---|
| A | 30 |
| B | 20 |
| C | 10 |
| Total | 60 |

¿Cuál es su decisión con respecto a la hipótesis nula si aplica la prueba chi-cuadrada.

5. El departamento de tarjetas de crédito del Carolina Bank sabe por experiencia que 5% de sus tarjetahabientes terminó algunos años de la preparatoria, 15%, la preparatoria, 25%, algunos años de la universidad, y 55%, una carrera. De los 500 tarjetahabientes a quienes se les llamó por no pagar sus cargos en el mes, 50 terminaron algunos años de preparatoria, 100, la preparatoria, 190, algunos años de la universidad, y 160 se graduaron de la universidad. ¿Es posible concluir que la distribución de los tarjetahabientes que no pagan sus cargos es diferente a los demás? Utilice el nivel de significancia 0.01.

6. Un científico social tomó una muestra de 140 personas y las clasifica de acuerdo con su nivel de ingresos, y si jugaron o no en la lotería estatal el mes pasado. La información de la muestra aparece a continuación. ¿Es posible concluir que jugar a la lotería se relaciona con el nivel de ingresos? Utilice el nivel de significancia 0.05.

| | **Ingreso** | | | |
|---|---|---|---|---|
| **Descripción** | **Bajo** | **Medio** | **Alto** | **Total** |
| Jugaron | 46 | 28 | 21 | 95 |
| No jugaron | 14 | 12 | 19 | 45 |
| Total ... | 60 | 40 | 40 | 140 |

7. La directora de publicidad del Carolina Sun Times, el periódico más importante en Carolina del Norte y Carolina del Sur, estudia la relación entre el tipo de comunidad en que reside un suscriptor y la sección del periódico que lee primero. Para una muestra de lectores recopiló la siguiente información:

| **Ubicación** | **Noticias nacionales** | **Deportes** | **Tiras cómicas** |
|---|---|---|---|
| Ciudad | 170 | 124 | 90 |
| Suburbios | 120 | 112 | 100 |
| Rural | 130 | 90 | 88 |

Con un nivel de significancia de 0.05, ¿se puede concluir que hay una relación entre el tipo de comunidad donde reside la persona y la sección del periódico que lee primero?

8. El departamento de control de calidad de Food Town, Inc., cadena de abarrotes en el norte de Nueva York, realiza una verificación mensual sobre la comparación de los precios registrados con

los precios anunciados. La siguiente gráfica resume los resultados de una muestra de 500 artículos del mes pasado. La gerencia de la compañía quiere saber si hay una relación entre las tasas de error en los artículos con precios normales y los artículos con precios especiales. Utilice el nivel de significancia 0.01.

| Descripción | Precio regular | Precio especial anunciado |
|---|---|---|
| Precio bajo | 20 | 10 |
| Precio mayor | 15 | 30 |
| Precio correcto | 200 | 225 |

# BIBLIOGRAFÍA


- Lind, D.A., Marchal, W.G., Wathen, S.A. (15). (2012). *Estadística Aplicada a los Negocios y la Economía*. México: McGraw-Hill

- David M. Levine, Timothy C. Krehbiel, Mark L. Berenson. 2006. *Estadística para Administración*. (4° edición). Naucalpan de Juárez, México.: Pearson Prentice Hall

- Allen L. Webster. 2006. *Estadística aplicada a los Negocios y la Economía*. (3° edición). Sata Fe de Bogotá, Colombia.: Irwin McGraw-Hill