Nombre: ________________________________ Fecha: ________________

Actividad *NUMB3RS*: Golosinas de colores

En “Final de la guardia” el FBI identifica a un sospechoso del asesinato de un policía. Cuando van a interrogarlo, lo encuentran drogado y muerto. Charlie piensa, dada la composición química de la droga, que puede aplicar un método estadístico de identificación para seguir el rastro hasta cierto narcotraficante que puede estar envuelto en los crímenes. En cualquier situación, los resultados (como la proporción de ingredientes en una muestra de heroína) varían al azar. Si lanzamos una moneda 1,000 veces, esperaríamos que salga cara alrededor de 500 veces. La prueba del *ji-cuadrado para el grado de coincidencia* nos permite establecer si hay una diferencia significativa entre las proporciones que se esperaría hallar en una muestra y las proporciones que realmente se hallan. Esta actividad encierra una introducción informal a la mecánica de dicho análisis.

Supongamos que ciertas golosinas de una marca popular vienen en cinco colores. Un estudiante contó las golosinas de cada color que hay en una bolsa y encontró los resultados de la siguiente tabla.

Según estos datos, ¿es probable que esta bolsa de golosinas fuera el resultado de un proceso fabril diseñado para producir proporciones iguales de cada color? Una manera de contestar esta pregunta es mediante el procedimiento estadístico llamado la prueba del ji-cuadrado para el grado de coincidencia, que indica en qué medida coincide la distribución de la muestra con la distribución teórica.

1. ¿Cuántas golosinas había en la bolsa? Si hubiera igual proporción de todas, ¿cuántas golosinas de cada color esperarías hallar? Estos valores se llaman los *valores previstos*. El recuento real de cada color en la tabla se llama el *valor observado.* Completa las dos columnas siguientes. ¿Crees que hay una diferencia considerable entre los valores observados y los previstos?

Color	**Observado**	**Previsto**	**Observado – Previsto**	$\frac{(O-E)^2}{E}$
Amarillo	11			
Rojo	19			
Azul	25			
Naranja	17			
Verde	13			

2. Esta prueba del ji-cuadrado cuantifica el grado de diferencia entre los valores observados y los previstos para cada color. ¿Tiene sentido hallar la suma de las diferencias (Observado – Previsto) para describir la diferencia total? ¿Por qué?

3. Para cada color, calcula “(Observado– Previsto)^2 / Previsto”; ingresa este valor en la última columna. Halla la suma de estos cinco valores. Esto es el *ji-cuadrado*, representado por χ^2.

Ahora la pregunta es si el ji-cuadrado que se halló es grande o pequeño. Para averiguarlo, puedes comparar tu ji-cuadrado observado con la distribución teórica de las probabilidades de que un ji-cuadrado sea mayor o igual a cualquier valor dado. La función de la densidad del ji-cuadrado se puede representar gráficamente con una calculadora gráfica. Configura la ventanilla así: **Xmín = 0**, **Xmáx = 20**, **Xescal = 1**, **Ymín = –0.05**, **Ymáx = 0.2**, **Yescal = 1**, **Xres = 1**. Presiona [2nd] [DISTR] para el menú de distribución. Corre a la derecha para seleccionar

John F. Mahoney, Escuela Superior Académica Benjamin Banneker, Washington, DC

DIB, escoge **Fdenχ^2(** e ingresa lo siguiente: **Fdenχ^2(Resp, 1000, 4)**. “Resp” es el ji-cuadrado calculado en la Pregunta #3. El número 1,000 se emplea como número grande, de modo que puedes hallar la probabilidad de que un ji-cuadrado caiga al azar entre tu valor y 1,000. El número final en el mando es el número de variables independientes en el problema, llamado grados de libertad. En este ejemplo, es 4. Como hay 85 golosinas y 5 colores, el número de golosinas posibles para el quinto color está determinado por el número de golosinas de los otros cuatro colores. Es decir, hay sólo cuatro variables independientes y el último color depende de cuántos hay de los otros cuatro. El mando “**Fdenχ^2**” calcula el área de la región debajo de la gráfica del valor del ji-cuadrado hasta (en este caso)1,000. Esta área se llama el valor *p*, la probabilidad de que un ji-cuadrado sea, simplemente al azar, igual o mayor que el valor observado.

4. ¿Cuál es el valor *p* para la bolsa de golosinas?

Un ji-cuadrado de 7.77 daría un valor *p* de 0.10. Este valor significa que si los colores de las golosinas realmente se distribuyeran por partes iguales, se presentaría un ji-cuadrado de 7.77 o mayor al azar aproximadamente el 10% de las veces. Algo que ocurre 10% de las veces no se considera muy raro, por lo cual un estadístico consideraría que no hay indicios suficientes para afirmar que las bolsas de golosina no vinieron de un proceso que produce los dulces de colores en igual número. Los estadísticos suelen llegar a esta conclusión si el valor *p* es más de 5%. Si el valor *p* es menor que 5%, el estadístico consideraría que hay evidencia suficiente para rechazar la suposición de que las bolsas de golosina vinieron de un proceso que produce los dulces de colores en igual número.

5. Según la respuesta a la Pregunta #4, ¿hay indicios suficientes para rechazas la hipótesis de que las bolsas de golosinas vinieron de un proceso que produce los colores en igual número? ¿Por qué?

6. Otro estudiante abrió una bolsa de golosinas de otra marca, contó las golosinas de cada color y halló los resultados que aparecen en la siguiente tabla.

Color	**Observado**	**Previsto**	$\frac{(O-E)^2}{E}$
Café	15		
Amarillo	14		
Rojo	16		
Azul	35		
Naranja	29		
Verde	24		

Repite las Preguntas #1 a 5, haciendo las modificaciones que sean necesarias, y aplica una prueba del ji-cuadrado para determinar si es probable que esta bolsa venga de un proceso de manufactura diseñado para producir igual número de dulces de cada color.

7. La tercera estudiante miró el sitio web del fabricante de esta segunda marca y vio la afirmación de que las golosinas se fabricaban en las siguientes proporciones: café 13%. amarillo 14%, rojo 13%, azul 24%, naranja 20% y verde 16%. Dada la distribución de colores en esta bolsa, ella dudó de la aseveración del fabricante. Repite las Preguntas #1–5 y usa la prueba del ji-cuadrado para determinar si es probable que esta bolsa de golosinas provino de un proceso fabril diseñado para producir dulces de cada color en estas proporciones.

John F. Mahoney, Escuela Superior Académica Benjamin Banneker, Washington, DC

El objeto de esta actividad es dar a los estudiantes un vistazo breve y sencillo de un tema matemático muy extenso. TI y NCTM lo invitan a usted y a sus estudiantes a aprender más sobre este tema con las extensiones que se ofrecen abajo y con su propia investigación independiente.

Extensiones

Para el estudiante

1. Explora la distribución de colores de varias marcas de golosinas, como Kissables®, variedades de M&Ms®, Skittles®, Reese’s Pieces®, etc., empleando las técnicas de esta actividad. ¿Hay indicios suficientes para aceptar que la distribución de colores es uniforme?

2. Las pautas administrativas para la distribución de calificaciones en cierta escuela secundaria son: A 10%, B 20%, C 40%, D 20% y F 10%. Una profesora de matemáticas dio las siguientes calificaciones a sus estudiantes en sus clases de estadística: 15 As, 25 Bs, 25 Cs, 15 Ds y 6 Fs. El director argumentó que la profesora no siguió las directrices. Ofrece evidencia estadística para respaldar o refutar lo que dice el director.

3. Haz una encuesta de por lo menos 20 estudiantes y pegúntales cuál de los siguientes es su color preferido: azul, rojo, verde o amarillo. Quizá esperes que la distribución de las preferencias sea igual, pero usualmente no lo es, cosa que se puede demostrar usando una prueba del ji-cuadrado.

Recursos adicionales

1. El sitio web de M&M® da el recuento de los colores de varios productos: **http://us.mms.com/us/about/products**

2. Para una explicación completa de la prueba del ji-cuadrado y una serie de ejercicios de simulación, visita: **http://www.math.uah.edu/stat/hypothesis/ChiSquare.xhtml**

3. Los libros de estadística AP® incluyen secciones sobre la prueba del ji-cuadrado para grado de coincidencia. Algunos de ellos son:

- Yates, D., Moore, D., & Starnes, D. (2003) The Practice of Statistics. New York: W. H. Freeman and Company.
- Bock, D., Velleman, P., & DeVeaux, R. (2004) Stats: Modeling the World. Boston: Pearson Education, Inc.
- Peck, R., Olsen, C., & Devore, J. (2005) Introduction to Statistics and Data Analysis, Second edition. Belmont: Thomson, Brooks/Cole.

John F. Mahoney, Escuela Superior Académica Benjamin Banneker, Washington, DC