

# Prueba de Chi cuadrada y covarianza

Gabriela Sánchez Y.

5064

El presente trabajo se divide en secciones, cada una de ellas se enfoca en un objetivo en particular. La sección 1 presenta una aplicación de la prueba de Chi cuadrada y finalmente la sección 2 valida de forma numérica y analítica dos propiedades relacionadas a la covarianza.

Los experimentos realizados para encontrar los resultados numéricos se realizan con el apoyo del lenguaje de programación R [4] cuyo código puede consultarse en el archivo `t11.R` [5].

## 1. Prueba de Chi cuadrada

Existen varios tipos de pruebas de Chi cuadrada: la prueba de bondad de ajuste y las pruebas de asociación e independencia. La prueba de bondad de ajuste se utiliza para probar qué tan bien se ajusta una muestra de datos categóricos a una distribución teórica. Las pruebas de asociación e independencia, tal y como su nombre lo dice, se usan para determinar si una variable está asociada con otra y para determinar si el valor observado de una variable depende del valor observado de otra, respectivamente [1].

Por el momento no fue posible determinar una aplicación de esta prueba a los datos actuales de la tesis por lo que se usan los datos proporcionados por el INEGI [2] sobre Seguridad pública y justicia. Específicamente datos sobre la percepción de la población mayor de 18 años, por tipo de autoridad. Se tienen dos tipos de percepción *algo efectivo* y *muy efectivo* y se analizan sólo dos tipos de autoridad: policía federal y estatal. Los datos se muestran en el cuadro 1.

Cuadro 1: Percepción de la efectividad del trabajo de diferentes tipos de autoridades.

	Muy efectivo	Algo efectivo
Policía federal	16.3	49.7
Policía estatal	8.7	45.1

Con estos datos, se aplica la prueba de asociación de Chi cuadrada para determinar si el tipo de autoridad afecta al nivel de percepción. La hipótesis nula es que no hay asociación entre estas dos variables, es decir, que el tipo de autoridad no se asocia con la percepción de efectividad. Para poder analizar el valor de *chi cuadrada*, primero se aclara que se tiene una tabla de valores de  $2 \times 2$ , es decir, se tienen dos grados de libertad. En este caso el valor crítico con el que se compara el valor de *chi cuadrada* es 3.841.

Los resultados obtenidos se muestran a continuación:

Pearson's Chi-squared test

data: prueba  
X-squared = 1.3047, df = 1, p-value = 0.2534

El  $p$ -valor es mayor que el nivel de significación, lo que indica que no se puede rechazar la hipótesis nula. Observe también que  $\chi^2 < 3.841$ , de acuerdo a esto, la hipótesis nula no se rechaza.

## 2. Covarianza

En esta sección se desea validar de forma numérica y analítica dos propiedades relacionadas a la covarianza. Para lograr el objetivo, primero es necesario definir la covarianza, recordar el concepto de varianza y algunas propiedades del valor esperado. En adelante  $\mu_x$  y  $\mu_y$  representan el valor esperado de las variables aleatorias  $X$  y  $Y$ , respectivamente

Sean  $X$  y  $Y$  dos variables aleatorias, la covarianza de éstas se define mediante la ecuación 1 o, de manera alternativa, mediante la ecuación 2

$$\text{Cov}[X, Y] = E[(X - \mu_x)(Y - \mu_y)] \quad (1)$$

$$= E[XY] - \mu_x \mu_y. \quad (2)$$

Por su parte, la varianza de una variable aleatoria  $X$  se define por la ecuación 1

$$V[X] = E[(X - \mu_x)^2] = E[X^2] - \mu_x^2. \quad (3)$$

El valor esperado cumple con diferentes propiedades [3]: el valor esperado de la suma de dos variables aleatorias es la suma de los valores esperados y el valor esperado de una variable aleatoria multiplicada por una constante es esa constante multiplicada por el valor valor esperado de la variable aleatoria. Es decir,

$$E[X + Y] = E[X] + E[Y], \quad (4)$$

$$E[cX] = cE[X]. \quad (5)$$

### 2.1. Propiedad 1

Primero se demuestra que  $\text{Cov}[aX + b, cY + d] = ac \cdot \text{Cov}[X, Y]$  de forma analítica. De acuerdo a la definición de covarianza, es necesario calcular el valor esperado de las variables  $aX + b$  y  $cY + d$ . De esta manera, utilizando las propiedades del valor esperado, se tiene que

$$E[aX + b] = E[aX] + E[b] = aE[X] + b = a\mu_x + b,$$

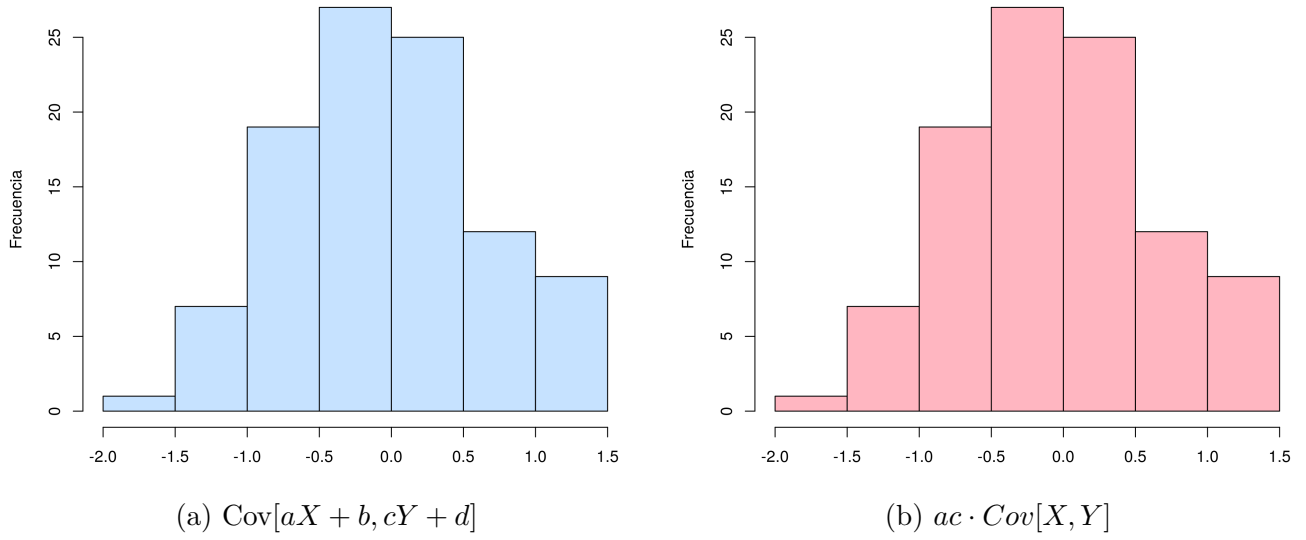


Figura 1: Demostración numérica de la propiedad  $\text{Cov}[aX + b, cY + d] = ac \cdot \text{Cov}[X, Y]$ .

procediendo de manera análoga se puede determinar que  $E[cY + d] = c\mu_y + d$ . Entonces

$$\begin{aligned}
 \text{Cov}[aX + b, cY + d] &= E[\{aX + b - (a\mu_x + b)\}\{cY + d - (c\mu_y + d)\}] \\
 &= E[(aX - a\mu_x)(cY - c\mu_y)] \\
 &= E[acXY - ac\mu_x Y - ac\mu_y X + ac\mu_x \mu_y] \\
 &= ac \cdot E[XY - \mu_x Y - \mu_y X + \mu_x \mu_y] \\
 &= ac \cdot E[(X - \mu_x)(Y - \mu_y)] \\
 &= ac \cdot \text{Cov}[X, Y].
 \end{aligned}$$

La demostración numérica consiste en tomar, para la variable  $X$  mil elementos que siguen una distribución de probabilidad normal con media cero y desviación estándar igual a cinco. Mientras que los elementos de la variable  $Y$  siguen una distribución uniforme. Después se crean las variables aleatorias  $aX + b$  y  $cY + d$ .

Sin perder la generalidad se definen  $a = 3$ ,  $b = 1$ ,  $c = 5$  y  $d = 2$ . Se calculan por separado  $\text{Cov}[aX + b, cY + d]$  y  $ac \cdot \text{Cov}[X, Y]$  mediante el uso de la función `cov` de R. El procedimiento se repite cien veces y los resultados se muestran en los histogramas de la figura 2. Prácticamente se tiene el mismo histograma por lo que se puede concluir que la propiedad sí se cumple.

## 2.2. Propiedad 2

Se desea demostrar que  $\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y] + 2\text{Cov}[X, Y]$ . Dada la definición de varianza, es necesario encontrar el valor esperado de la suma de las variables aleatorias  $X$  y  $Y$ :

$$E[X + Y] = E[X] + E[Y] = \mu_x + \mu_y.$$

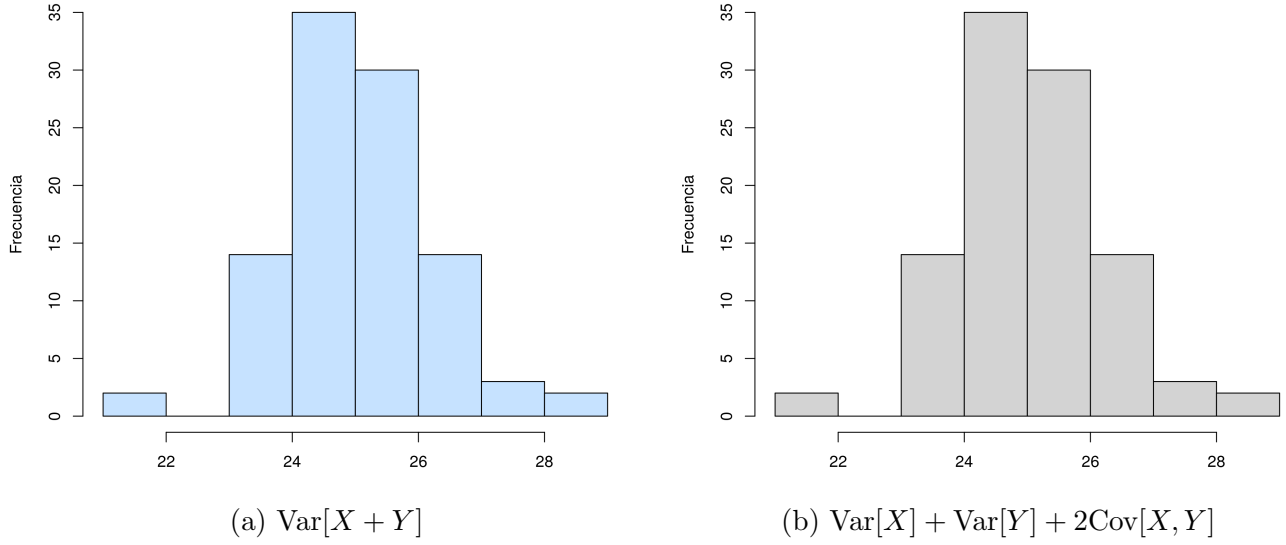


Figura 2: Demostración numérica de la propiedad  $\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y] + 2\text{Cov}[X, Y]$ .

Con este resultado, se procede a la demostración de la propiedad

$$\begin{aligned}
 \text{Var}[X + Y] &= E[(X + Y - \mu_x - \mu_y)^2] \\
 &= E[X^2 + XY - \mu_x X - \mu_y X + YX + Y^2 - \mu_x Y - \mu_y Y - \mu_x X - \mu_x Y + \mu_x^2 + \\
 &\quad \mu_x \mu_y - \mu_y X - \mu_y Y + \mu_x \mu_y + \mu_y^2] \\
 &= E[X^2 + 2XY - 2\mu_x X - 2\mu_y X - 2\mu_x Y - 2\mu_y Y + 2\mu_x \mu_y + \mu_x^2 + \mu_y^2 + Y^2] \\
 &= E[X^2 - 2\mu_x X + \mu_x^2] + E[Y^2 - 2\mu_y Y + \mu_y^2] + E[2XY - 2\mu_y X - 2\mu_x Y + 2\mu_x \mu_y] \\
 &= E[(X - \mu_x)^2] + E[(Y - \mu_y)^2] + 2E[(X - \mu_x)(Y - \mu_y)] \\
 &= \text{Var}[X] + \text{Var}[Y] + 2\text{Cov}[X, Y].
 \end{aligned}$$

La demostración numérica sigue una idea similar a la planteada para demostrar la propiedad 1. Nuevamente se definen  $X \sim \text{Norm}(0, 5)$  y  $Y \sim \text{Norm}(0, 1)$ . Mediante las funciones `var` y `cov` de R se realiza el cálculo de  $\text{Var}[X + Y]$  y  $\text{Var}[X] + \text{Var}[Y] + 2\text{Cov}[X, Y]$ . El proceso se repite un total de cien veces cuyos resultados se presentan en los histogramas de la figura 1. En este experimento, nuevamente se observa que se obtienen los mismos resultados en cada cálculo de manera que se apoya la demostración de la propiedad 2.

## Referencias

- [1] Minitab 18. ¿qué es una prueba de chi-cuadrada? <https://support.minitab.com/es-mx/minitab/18/help-and-how-to/statistics/tables/supporting-topics/chi-square/what-is-a-chi-square-test/>.
- [2] Instituto Nacional de Estadística y Geografía. Seguridad pública y justicia. <https://www.inegi.org.mx/datos/?t=0230>.

- [3] Charles M. Grinstead and J. Laurie Snell. *Introduction to Probability*. AMS, 2003.
- [4] Selva Prabhakaran. R-statistics. <http://r-statistics.co/Statistical-Tests-in-R.html>.
- [5] Gabriela Sánchez Y. Modelos probabilistas aplicados. <https://github.com/Saphira3000/MPA>.