

# Pruebas estadísticas

Gabriela Sánchez Y.

5064

Esta práctica se divide en dos partes, una teórica y una práctica. La parte teórica consiste en contestar una serie de preguntas relacionadas con pruebas estadísticas, mientras que la práctica muestra la aplicación de las mismas al un conjunto de datos de *seguridad pública y justicia* obtenidos del INEGI [1].

## 1. Parte teórica

En esta sección se responde una serie de preguntas realizando una lectura previa sobre pruebas estadísticas en tres sitios: [Centro de Ayuda XLSTAT](#), [EcuRed](#) y [Máxima formación](#).

### Relación entre contraste de hipótesis y pruebas estadísticas

Ambas formulan hipótesis y evalúan la evidencia estadística que proporcionan los datos para aceptar o rechazar dichas hipótesis.

#### ¿Qué indicaría rechazar la hipótesis nula?

Para poder rechazar una hipótesis nula, es necesario que el  $p$ -valor obtenido de la prueba estadística aplicada sea menor que el nivel de significación  $\alpha$  de la misma. Al rechazar la hipótesis nula, se acepta la hipótesis alternativa.

#### ¿Cómo se interpreta la salida de una prueba estadística?

Antes de realizar la prueba se debe especificar un nivel de significación  $\alpha$ , este valor indica la probabilidad de rechazar la hipótesis nula cuando es verdadera (error de tipo I). Para interpretar la salida de la prueba, se compara el  $p$ -valor con el valor de  $\alpha$ : si  $p\text{-valor} < \alpha$ , se rechaza la hipótesis nula  $H_0$ ; en cambio, si  $p\text{-valor} > \alpha$  no se rechaza la hipótesis nula. Este último resultado no necesariamente implica que se debe aceptar la hipótesis nula, más bien se tiene que no existe suficiente evidencia estadística para rechazar dicha hipótesis.

#### ¿Cómo seleccionar el valor de $\alpha$ ?

Anteriormente se mencionó que  $\alpha \in [0, 1]$  indica la probabilidad de rechazar la hipótesis nula cuando ésta es verdadera, por lo que la elección de su valor depende de los datos que se estudian. El qué tan peligroso es cometer un error de este tipo, determinará el valor de  $\alpha$  a usar. Por ejemplo, si se desea analizar el efecto de algún tratamiento médico, el valor elegido debe ser muy pequeño.

#### ¿Cuáles son los errores frecuentes de interpretación del $p$ -valor?

Tomar un valor de significación no adecuado, conduciría a cometer el error de rechazar la

hipótesis nula cuando ésta es verdadera.

### ¿Qué es la potencia estadística y para qué sirve?

La potencia de una prueba estadística es la capacidad de la prueba para rechazar la hipótesis nula cuando la hipótesis alternativa es verdadera. En otras palabras, la potencia es la probabilidad de no cometer un error de tipo II. Es por esto que su valor es  $1 - \beta$ , donde  $\beta$  es la probabilidad de aceptar la hipótesis nula cuando es falsa (error de tipo II).

Tener un buen valor para la potencia estadística, indica que la prueba detectará un efecto que realmente existe en los datos.

### ¿Cuáles son los supuestos para aplicar técnicas paramétricas?

Se usa una prueba paramétrica si los datos son cuantitativos. Un requisito indispensable es que los datos tengan una distribución de probabilidad normal, además deben tener varianzas iguales o similares (homocedasticidad).

### Ejemplos de pruebas estadísticas paramétricas y no paramétricas

Dentro de las pruebas paramétricas más comunes se encuentran la prueba  $t$  de Student que se utiliza para comparar las medias de dos poblaciones, el análisis de varianza (ANOVA), el análisis de covarianza (ANCOVA), entre otros.

Algunos ejemplos de pruebas no paramétricas son la prueba de Wilcoxon que se puede usar para comparar las medias de dos grupos que no siguen una distribución normal, Kruskal-Wallis (el equivalente a un ANOVA de una vía) y para comparar dos grupos de variables cualitativas nominales y ordinales se puede usar la prueba Chi-cuadrada.

### Guía para encontrar la prueba estadística que buscas

Si se tiene una sola muestra y se quiere determinar si la media de la misma tiene un valor específico, se puede seguir el diagrama de la figura 1, en este caso lo único que se debe verificar para elegir el tipo de prueba es si los datos o no siguen una distribución normal.

En la figura 2 se presenta una guía para determinar el tipo de prueba a usar cuando se desea comparar la media de dos muestras. En este diagrama se usa ND para preguntar si los datos son normalmente distribuidos.

## 2. Parte práctica

En esta sección se aplican distintas pruebas estadísticas usando el lenguaje de programación R [2] a los datos sobre *seguridad pública y justicia* obtenidos de la página del Instituto Nacional de Estadística y Geografía [1]. Para todas las pruebas se fija el nivel de significación  $\alpha = 0.05$ .

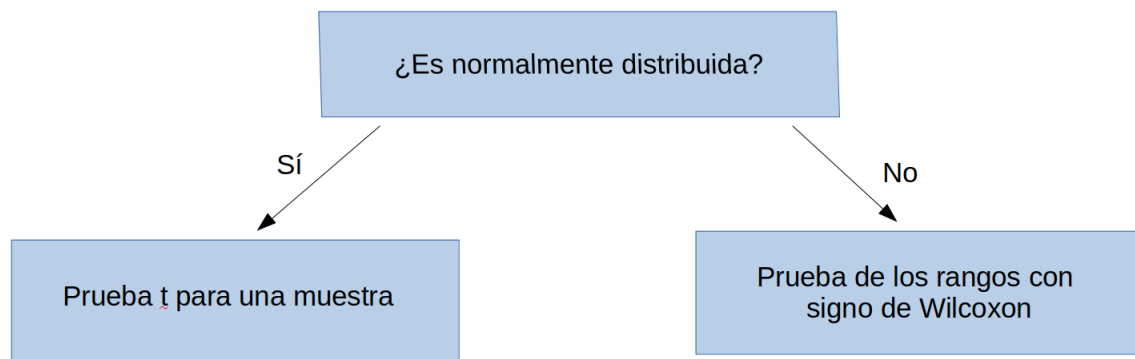


Figura 1: Guía para determinar el tipo de prueba para una sola muestra.

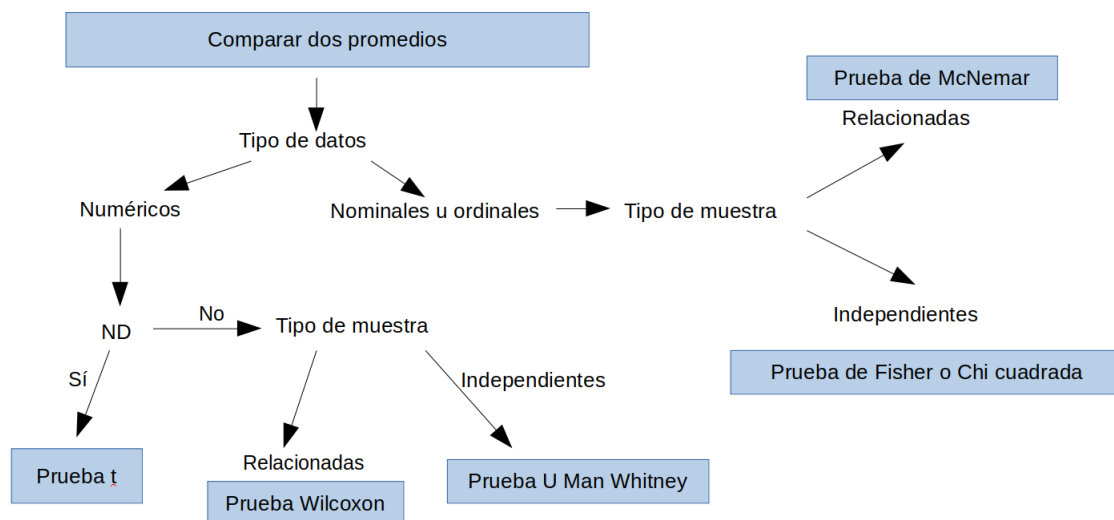


Figura 2: Guía para determinar el tipo de prueba para comparar la media de dos muestras.

## 2.1. Prueba de Shapiro

La prueba de *Shapiro* es usada para determinar si una muestra sigue una distribución normal. La hipótesis nula  $H_0$  de esta prueba es que los datos de la muestra están normalmente distribuidos. Si el  $p$ -valor es mayor que  $\alpha$ , no se rechaza la hipótesis nula y en este caso se confirma que la muestra cumple con la normalidad. En las siguientes subsecciones se muestran algunos ejemplos prácticos del uso de esta prueba.

## 2.2. Prueba $t$ de una muestra

Es una prueba paramétrica usada para determinar si la media de una muestra (con distribución normal), podría tener un valor específico.

Se toman los datos de la tasa de prevalencia delictiva por cada cien mil habitantes en las distintas entidades federativas del país en los años 2010 a 2013 y se verifica con la prueba si se cumple la hipótesis  $H_0$  que la población tiene una media igual a 23,600. Antes de aplicar la prueba se verifica que los datos sigan una distribución normal con la prueba de *Shapiro*. A continuación se muestran los resultados obtenidos de ambas pruebas.

Shapiro-Wilk normality test

```
data: delitos
W = 0.97897, p-value = 0.1259
-----
```

One Sample t-test

```
data: delitos
t = 0.022397, df = 95, p-value = 0.9822
alternative hypothesis: true mean is not equal to 23600
95 percent confidence interval:
22456.15 24769.96
sample estimates:
mean of x
23613.05
```

El  $p$ -valor obtenido de la prueba de *Shapiro* indica normalidad. Por otra parte, el  $p$ -valor obtenido de la prueba  $t$  indica que no se puede rechazar la hipótesis nula, sin embargo esto no quiere decir que se acepte. Es decir, esto no afirma que la media de los datos sea 23,600.

## 2.3. Prueba de los rangos con signo de Wilcoxon

Es la alternativa a la prueba anterior cuando no se cumple con la distribución normal.

Se analiza si la media de los datos de la tasa de prevalencia delictiva en el periodo comprendido desde 2010 hasta 2019 en todas las entidades federativas del país es igual a 23,600. Antes de aplicar la prueba, se verifica si hay normalidad en los datos. La salida que se obtiene es la siguiente:

Shapiro-Wilk normality test

```
data:  todos_delitos
W = 0.94654, p-value = 9.857e-09
-----
```

Wilcoxon signed rank test with continuity correction

```
data:  todos_delitos
V = 23976, p-value = 0.02514
alternative hypothesis: true location is not equal to 23600
```

El  $p$ -valor obtenido es menor a  $\alpha$ , por lo que se rechaza la hipótesis nula y se acepta la hipótesis alternativa que la media de los datos no es igual a 23,600.

## 2.4. Prueba $t$ para dos muestras y prueba de suma de rangos de Wilcoxon

Las dos pruebas son usadas para comparar las medias de dos muestras. La diferencia radica en que la prueba  $t$  requiere que se cumpla la normalidad, mientras que la prueba de *Wilcoxon* no.

Ahora se analiza si la media de la tasa de prevalencia delictiva en el periodo comprendido desde 2010 hasta 2019 en la zona norte del país es igual a la de la zona centro. Para la zona norte se consideran únicamente los estados que están en la frontera del país; es decir, Baja California Norte, Coahuila, Chihuahua, Nuevo León, Sonora y Tamaulipas. Para la zona centro se consideran también seis entidades federativas: Ciudad de México, Guanajuato, Estado de México, Michoacán, Querétaro y Tlaxcala.

Como primer paso se revisa si ambas muestras cumplen con la normalidad y de acuerdo a los  $p$ -valores obtenidos, se concluye que una de ellas no lo cumple. Es por esto que se utiliza la prueba de suma de rangos de *Wilcoxon*.

Shapiro-Wilk normality test

```
data:  norte
W = 0.96785, p-value = 0.1548
-----
```

Shapiro-Wilk normality test

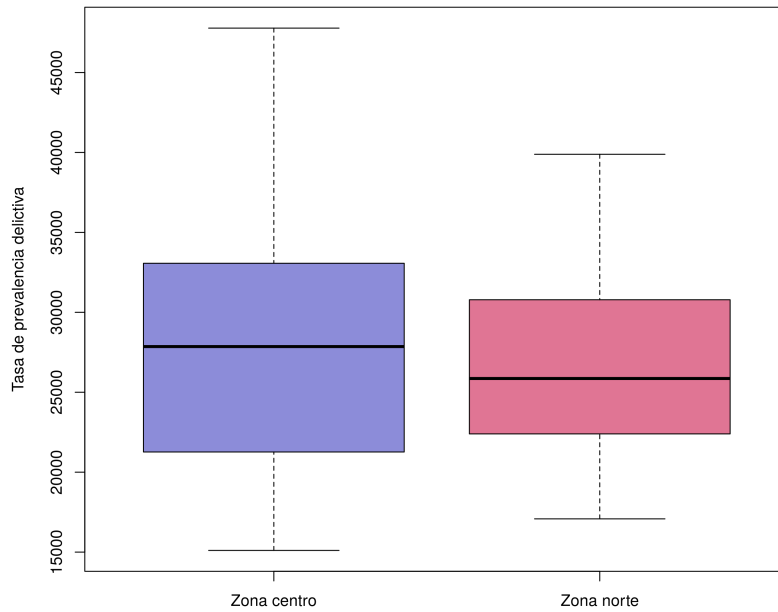


Figura 3: Diagrama caja-bigote de la variación de la tasa de prevalencia delictiva en distintas zonas del país.

```
data: centro
W = 0.92417, p-value = 0.002158
```

El  $p$ -valor obtenido de la prueba de suma de rangos de *Wilcoxon* es mayor que el valor de  $\alpha$  por lo que se concluye que no existe suficiente evidencia estadística para rechazar dicha hipótesis nula.

Wilcoxon rank sum test with continuity correction

```
data: norte and centro
W = 1298, p-value = 0.3271
alternative hypothesis: true location shift is not equal to 0
```

Sin embargo, si se grafica un diagrama de caja-bigote de ambas muestras tal y como se observa en la figura 3, es claro que las muestras no tienen la misma media.

## 2.5. Prueba de Kolmogorov-Smirnov

Esta prueba es utilizada para determinar si dos muestras siguen la misma distribución. La hipótesis nula  $H_0$  de esta prueba es precisamente que ambas muestras provienen de la misma distribución.

## 2.6. Prueba de Fisher

Es usado para determinar si dos muestras tiene la misma varianza. Para un ejemplo de su uso, se retoman las muestras utilizadas en la subsección 2.4. El resultado obtenido de aplicar la prueba es el siguiente:

F test to compare two variances

```
data: centro and norte
F = 2.1669, num df = 53, denom df = 53, p-value = 0.00564
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
1.257380 3.734393
sample estimates:
ratio of variances
2.166922
```

Vea que se cumple que  $p\text{-valor} < \alpha$ , por lo que se rechaza la hipótesis nula y se acepta la hipótesis alterna que las varianzas son distintas.

## 2.7. Prueba Chi cuadrada

Esta prueba se usa para determinar si dos características son independientes o tienen alguna asociación.

Para ejemplificar esta prueba se toman los resultados de 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 tres tipos de autoridad: policía federal, estatal y municipal. Los datos se muestran en el cuadro 1.

Con estos datos, se desea saber 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. Los resultados obtenidos se muestran a continuación.

Pearson's Chi-squared test

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

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

El  $p$ -valor es mayor que el nivel de significación, lo que significa que no se puede rechazar la hipótesis nula. Para 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. Observe que *chi cuadrada* ¡3.841, de acuerdo a esto, la hipótesis nula no se rechaza.

## Referencias

- [1] Instituto Nacional de Estadística y Geografía. Seguridad pública y justicia. <https://www.inegi.org.mx/datos/?t=0230>.
- [2] Selva Prabhakaran. R-statistics. <http://r-statistics.co/Statistical-Tests-in-R.html>.