Métodos Estadísticos y Simulación

Unidad 7: Inferencia Estadística: Pruebas de hipótesis no paramétricas

Mg. J. Eduardo Gamboa U.

2025-07-07

Introducción

¿Por qué importa la distribución de los datos?

En las pruebas de hipótesis paramétricas hemos asumido supuestos acerca de la distribución de los datos y/o el tamaño de muestra. Sin embargo, no habíamos realizado la verificación de dichos supuestos.
Cuando no se cumplen estos supuestos, se debe optar por el uso de pruebas no paramétricas, que se caracterizan por su robustez ante el incumplimiento de supuestos.
No obstante, usadas incorrectamente, las pruebas no paramétricas pueden perder potencia y descartar información al usar rangos en vez de los datos originales.

Pruebas de bondad de ajuste

Una prueba de bondad de ajuste compara las frecuencias observadas o la distribución empírica de los datos con las frecuencias esperadas bajo la hipótesis nula de que los datos provienen de una distribución específica H_0.

Pruebas de Kolmogorov-Smirnov

- La prueba de K-S fue propuesta por Andréi Kolmogórov (1903 - 1987) y Nikolái Smirnov (1900 - 1966)
- Compara la distribución de distribución acumulada empírica $F_n(x)$ con la teórica $F(x)$ mediante el estadístico de prueba:

$$D = max|F_n(x) - F(x)|$$

- Se debe conocer los parámetros de la distribución en cuestión.
- Se sugiere su uso para datos continuos.

Ejemplo

¿Los siguientes datos se ajustan a una distribución uniforme en el intervalo [0,1]?
Considerar $\alpha = 0.10$

`0.2876,0.7883,0.4090,0.8830,0.9405,0.0456,0.5281,0.8924,0.5514,0.4566`

H_0 : Los datos se ajustan a la distribución $U(0,1)$

H_1 : Los datos no se ajustan a la distribución $U(0,1)$

$\alpha = 0.10$

```r
datos <- c(0.2876,0.7883,0.4090,0.8830,0.9405,0.0456,0.5281,0.8924,
           0.5514,0.4566)
datos |> ks.test("punif", min = 0, max = 1)
```

```

	Exact one-sample Kolmogorov-Smirnov test

data:  datos
D = 0.209, p-value = 0.7016
alternative hypothesis: two-sided
```

Con un n.s. del 10%, no se rechaza H_0, por lo tanto los datos sí se ajustan a una distribución uniforme en el intervalo [0,1].

Ejemplo

Una municipalidad quiere saber si el tiempo de atención (en minutos) de sus ventanillas sigue una distribución exponencial con tiempo medio de 5 minutos, porque esperan que los tiempos entre atenciones sean aleatorios y con una tasa constante.
Los tiempos registrados son los siguientes:

`4.22,2.88,6.65,0.16,0.28,1.58,12.57,0.73,13.63`

H_0 : Los datos se ajustan a la distribución $Exp(\lambda = 0.2)$

H_1 : Los datos no se ajustan a la distribución $Exp(\lambda = 0.2)$

$\alpha = 0.10$

```r
datos <- c(4.22,2.88,6.65,0.16,0.28,1.58,12.57,0.73,13.63)
datos |> ks.test("pexp", rate = 1/5)
```

```

	Exact one-sample Kolmogorov-Smirnov test

data:  datos
D = 0.19749, p-value = 0.8102
alternative hypothesis: two-sided
```

Con un n.s. del 10%, no se rechaza H_0, por lo tanto los datos sí se ajustan a una distribución exponencial con media de 5 minutos.

Ejemplo

Un psicólogo registró los tiempos de reacción (en segundos) de estudiantes ante un estímulo visual. Queremos verificar si estos tiempos siguen una distribución normal con media $\mu = 1.5$ y desviación estándar $\sigma = 0.5$.

Los tiempos registrados son los siguientes:

`2.2,2.4,3.3,2.5,2.6,3.4,2.7,1.9`

H_0 : Los datos se ajustan a la distribución $N(\mu = 1.5, \sigma^2 = 0.5^2)$

H_1 : Los datos no se ajustan a la distribución $N(\mu = 1.5, \sigma^2 = 0.5^2)$

$\alpha = 0.10$

```
tiempos <- c(1.2,1.4,2.3,1.5,1.6,2.4,1.7,0.9)
tiempos |> ks.test("pnorm", mean = 1.5, sd = 1)
```

```
    Exact one-sample Kolmogorov-Smirnov test

data:  tiempos
D = 0.27425, p-value = 0.5005
alternative hypothesis: two-sided
```

Con un n.s. del 10%, no se rechaza H_0, por lo tanto los datos sí se ajustan a una distribución normal con media de 1.5 segundos y desviación estándar de 0.5 segundo.

Prueba de Lilliefors

- Propuesta por Hubert Lilliefors (1928 - 2008).
- Se utiliza cuando se desea probar normalidad en un conjunto de datos sin asumir valores específicos para los parámetros, por lo que deben ser estimados.

Ejemplo

Un psicólogo registró los tiempos de reacción (en segundos) de estudiantes ante un estímulo visual. Queremos verificar si estos tiempos siguen una distribución normal. Los tiempos registrados son los siguientes:

`2.2,2.4,3.3,2.5,2.6,3.4,2.7,1.9`

H_0 : Los datos se ajustan a la distribución normal

H_1 : Los datos no se ajustan a la distribución normal

$\alpha = 0.10$

```r
tiempos <- c(1.2,1.4,2.3,1.5,1.6,2.4,1.7,1.9)
library(nortest)
tiempos |> lillie.test()
```

```

	Lilliefors (Kolmogorov-Smirnov) normality test

data:  tiempos
D = 0.17191, p-value = 0.6941
```

```r
tiempos |> mean()
```

```
[1] 1.75
```

```r
tiempos |> sd()
```

```
[1] 0.4242641
```

Con un n.s. del 10%, se rechaza H_0, por lo tanto los datos no se ajustan a una distribución normal con media $\mu = 1.75$ y desviación estándar $\sigma = 0.424$.

Prueba de Shapiro Wilk

- Está diseñada específicamente para probar normalidad.
- Es más potente, incluso para muestras pequeñas.
- Si la muestra es grande (mayor a 2000), casi siempre rechazará la hipótesis nula de normalidad. En ese caso, complementar con gráficos (histograma, boxplot, qqplot).
- Es sensible a outliers.
- No se invalida la prueba si se estima μ y σ^2 a partir de la muestra.

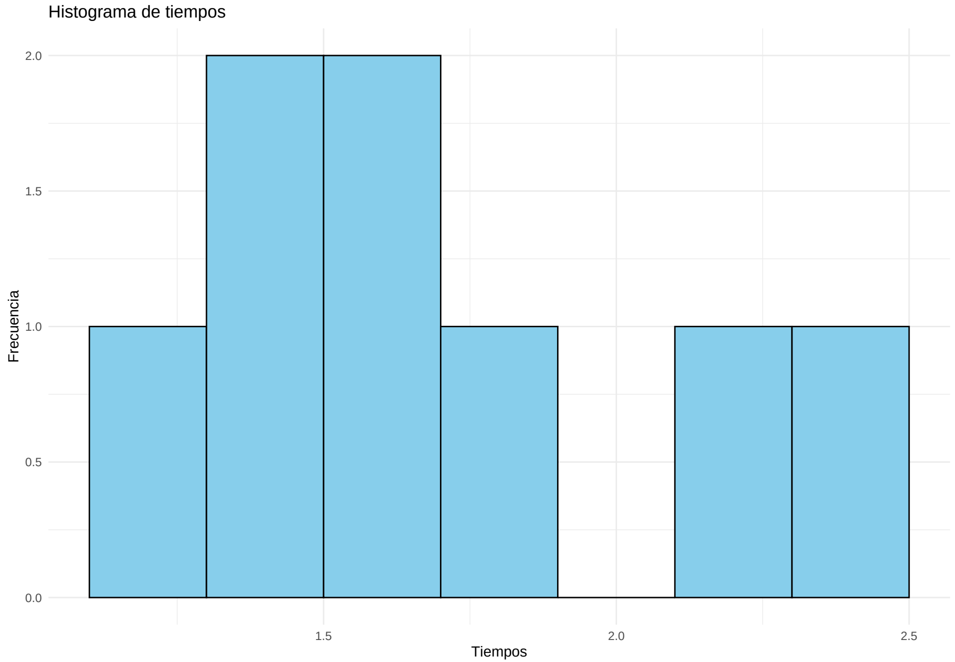
Ejemplo (cont.)

```r
tiempos <- c(1.2,1.4,2.3,1.5,1.6,2.4,1.7,1.9)
tiempos |> shapiro.test()
```

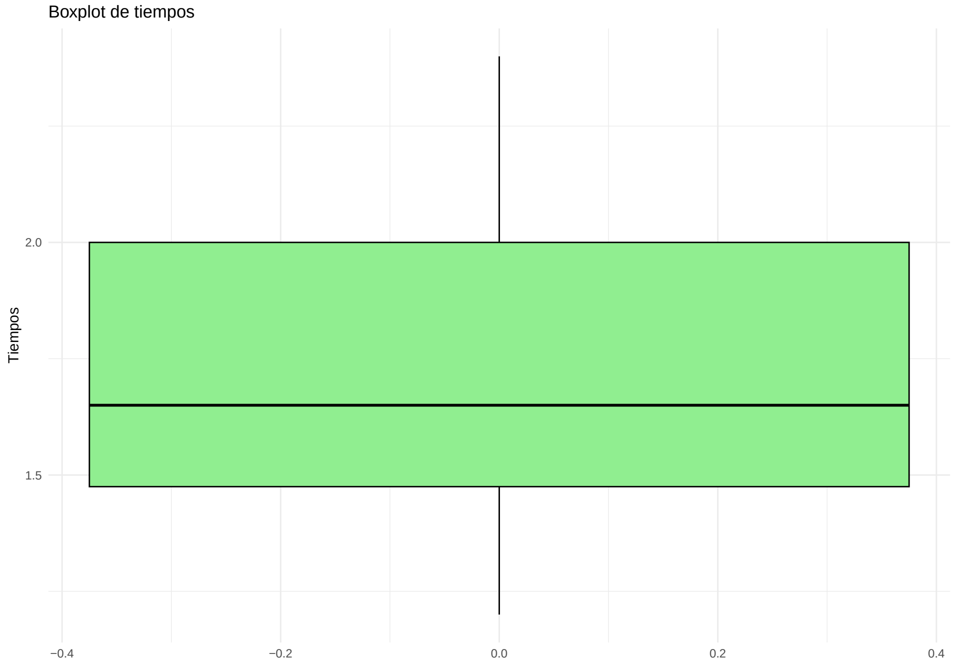
```
    Shapiro-Wilk normality test

data:  tiempos
W = 0.93678, p-value = 0.5797
```

Con un n.s. del 10%, se rechaza H_0, por lo tanto los datos no se ajustan a una distribución normal con media $\mu = 1.75$ y desviación estándar $\sigma = 0.424$.

```r
library(ggplot2)
tiempos |>
  data.frame() |>
  ggplot(aes(x = tiempos)) +
  geom_histogram(binwidth = 0.2, fill = "skyblue", color = "black") +
  labs(title = "Histograma de tiempos", x = "Tiempos", y = "Frecuencia") +
  theme_minimal()
```

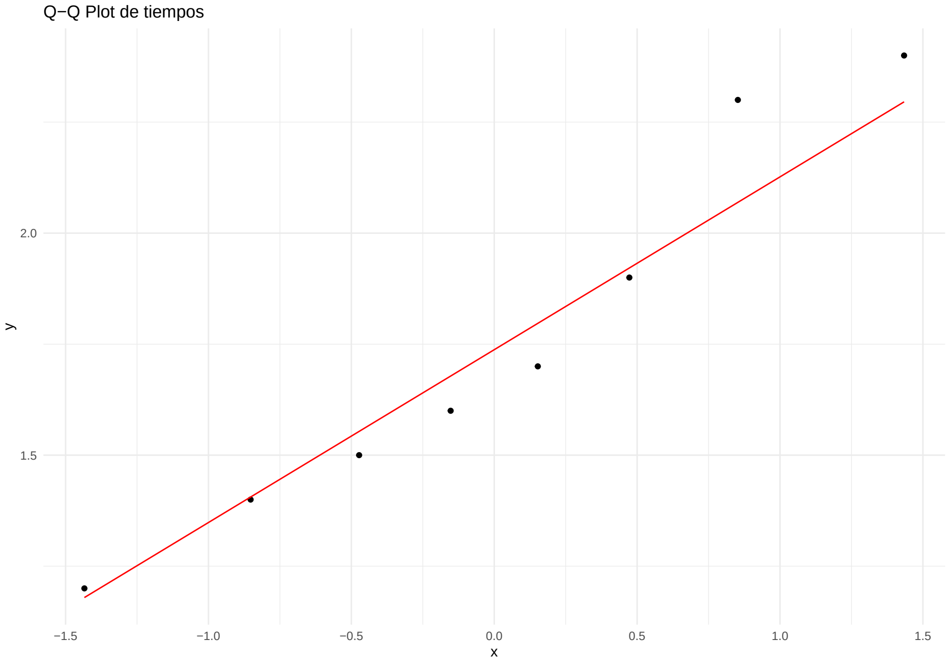

Histograma de tiempos
2.0
1.5
1.0
0.5
0.0
Frecuencia
1.5
2.0
2.5
Tiempos

```r
tiempos |>
  data.frame() |>
  ggplot(aes(y = tiempos)) +
  geom_boxplot(fill = "lightgreen", color = "black") +
  labs(title = "Boxplot de tiempos", y = "Tiempos") +
  theme_minimal()
```



Boxplot de tiempos
Tiempos
2.0
1.5
−0.4
−0.2
0.0
0.2
0.4

```r
tiempos |>
  data.frame() |>
  ggplot(aes(sample = tiempos)) +
  stat_qq() +
  stat_qq_line(color = "red") +
  labs(title = "Q-Q Plot de tiempos") +
  theme_minimal()
```


Q−Q Plot de tiempos
2.0
1.5
y
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
x

Otras pruebas de normalidad

- Anderson Darling: enfocada principalmente en evaluar las colas de la distibución, sensible a valores extremos
- Jarque-Bera: utiliza el tercer y cuarto momento para decidir acerca de la normalidad.

Pruebas no paramétricas en una población

Usarlas cuando los datos no son normales (verificado con Shapiro-Wilk o Lilliefors) y la muestra es pequeña (menor a 30); o cuando la variable está medida en escala ordinal.

Se tienen las siguientes pruebas:

- Prueba del signo
- Prueba de Wilcoxon

Prueba del signo

- Verifica si la mediana de la población es igual a un valor hipotético Me_0 contrastando las hipótesis:

$$H_0 : Me = Me_0 \qquad H_1 : Me \neq Me_0$$

- Usa solo los signos ($+$ o -) de las diferencias entre los datos y la mediana hipotética Me_0.
- Puede evaluarse también de manera unilateral, a la derecha o a la izquierda.

Prueba de Wilcoxon (o de rango con signo)

- Verifica si la mediana de la población es igual a un valor hipotético Me_0, contrastando las hipótesis:

$$H_0 : Me = Me_0 \qquad H_1 : Me \neq Me_0$$

- Considera tanto el signo como el tamaño de las diferencias (rango con signo).
- Es más potente que la prueba de signos, pero requiere que las diferencias sean simétricas alrededor de la mediana.
- Puede evaluarse también de manera unilateral, a la derecha o a la izquierda.

Ejemplo

Un investigador quiere saber si la mediana del tiempo que las personas pasan en una página web es 10 minutos, y para ello toma una muestra aleatoria de 7 usuarios, registrando los siguientes tiempos:

`12,8,15,13,7,43,5`

Verificamos normalidad:

H_0: El tiempo de permanencia en la página web sigue una distribución Normal

H_1: El tiempo de permanencia en la página web no sigue una distribución Normal

$\alpha = 0.10$

```r
tiempos = c(12,8,15,13,7,43,5)
tiempos |> lillie.test()
```

```

	Lilliefors (Kolmogorov-Smirnov) normality test

data:  tiempos
D = 0.34836, p-value = 0.01033
```

```r
tiempos |> shapiro.test()
```

```
	Shapiro-Wilk normality test

data:  tiempos
W = 0.71195, p-value = 0.004998
```

No se cumple el supuesto de normalidad, y el tamaño de muestra es pequeño.

$$H_0 : Me = 10 \qquad H_1 : Me \neq 10 \qquad \alpha = 0.10$$

```r
(tiempos - 10) |> mean()
```

```
[1] 4.714286
```

```r
(tiempos - 10) |> median()
```

```
[1] 2
```

Se aprecia que las diferencias no son simétricas, por lo que es más conveniente usar la prueba del signo.

```r
library(BSDA)
tiempos |> SIGN.test(md = 10, alternative = "two.sided")
```

```

	One-sample Sign-Test

data:  tiempos
s = 4, p-value = 1
alternative hypothesis: true median is not equal to 10
95 percent confidence interval:
  5.628571 34.200000
sample estimates:
median of x
         12

Achieved and Interpolated Confidence Intervals:

                  Conf.Level L.E.pt U.E.pt
Lower Achieved CI     0.8750 7.0000   15.0
Interpolated CI       0.9500 5.6286   34.2
Upper Achieved CI     0.9844 5.0000   43.0
```

Se puede afirmar que el tiempo mediano de permanencia en la página web es de 10 minutos.

Ejemplo

Un investigador mide los tiempos (en segundos) que toma a 9 estudiantes resolver un problema matemático. Quiere saber si la mediana poblacional del tiempo es mayor a 18 segundos. Los tiempos registrados son:

`4,20,21,29,10,20,2,66,9`

Verificamos normalidad:

H_0: El tiempo de resolución del problema matemático sigue una distribución Normal

H_1: El tiempo de resolución del problema matemático no sigue una distribución Normal

$\alpha = 0.10$

```r
tiempos = c(4,20,21,29,10,20,2,66,9)
tiempos |> lillie.test()
```

```

	Lilliefors (Kolmogorov-Smirnov) normality test

data:  tiempos
D = 0.25947, p-value = 0.08174
```

```r
tiempos |> shapiro.test()
```

```

	Shapiro-Wilk normality test

data:  tiempos
W = 0.80425, p-value = 0.02283
```

No se cumple el supuesto de normalidad, y el tamaño de muestra es pequeño.

$$H_0 : Me \leq 18 \qquad H_1 : Me > 18 \qquad \alpha = 0.10$$

```r
(tiempos - 18) |> mean()
```

```
[1] 2.111111
```

```r
(tiempos - 18) |> median()
```

```
[1] 2
```

Se aprecia simetría de las diferencias, por lo que es conveniente utilizar la prueba de Wilcoxon.

```r
tiempos |> wilcox.test(mu = 18, alternative = "greater")
```

```
	Wilcoxon signed rank test with continuity correction

data:  tiempos
V = 21, p-value = 0.5937
alternative hypothesis: true location is greater than 18
```

No se rechaza H_0, por lo tanto no hay evidencia estadística para afirmar que la mediana poblacional del tiempo es mayor a 18 segundos.

Pruebas no paramétricas en dos poblaciones

- Si las muestras provienen de poblaciones independientes: Prueba U de Mann Whitney
- Si las muestras provienen de poblaciones relacionadas: Prueba de Wilcoxon para datos pareados

Ejemplo

Una ONG evalúa el impacto de dos campañas ambientales sobre la actitud hacia el reciclaje en dos comunidades distintas:

- Comunidad A: Vio la campaña en redes sociales.
- Comunidad B: Participó en talleres presenciales.

Se preguntó: "En una escala del 1 al 20, ¿cuánto te comprometes a separar residuos?, donde 1 significa Nada comprometido y 20 Muy comprometido.

Se recolectaron los siguientes datos:

- Comunidad A: `1,8,2`
- Comunidad B: `11,15,7,18`

¿Existe diferencia significativa en el nivel de compromiso entre ambas comunidades?

H_0: No hay diferencia entre las distribuciones de compromiso.

H_1: Las distribuciones de compromiso difieren

$\alpha = 0.10$

```r
A <- c(1,8,2)
B <- c(11,15,7,18)
A |> wilcox.test(B, alternative = "two.sided", paired = FALSE)
```

```

	Wilcoxon rank sum exact test

data:  A and B
W = 1, p-value = 0.1143
alternative hypothesis: true location shift is not equal to 0
```

Con un nivel de significancia del 10%, no existe evidencia estadística para afirmar que las distribuciones de compromiso difieren entre las comunidades A y B.

Ejemplo

Un médico quiere evaluar si un nuevo medicamento reduce la presión arterial sistólica. Mide la presión en 10 pacientes antes y después del tratamiento:

- Antes: 145,150,138,142,148,135,140, 155,149,151
- Después: 140,144,137,140,142,130,136,150,143,147

H_0: La mediana de las diferencias antes-después es mayor o igual a 0

H_1: La mediana de las diferencias antes-después es menor a 0

$\alpha = 0.10$

```r
antes = c(145,150,138,142,148,135,140, 155,149,151)
despues = c(140,144,137,140,142,130,136,150,143,147)
antes |> wilcox.test(despues, alternative = "greater", paired = TRUE)
```

```
	Wilcoxon signed rank test with continuity correction

data:  antes and despues
V = 55, p-value = 0.002817
alternative hypothesis: true location shift is greater than 0
```

```r
andes = antes - despues
andes |> wilcox.test(alternative = "greater")
```

```
	Wilcoxon signed rank test with continuity correction

data:  andes
V = 55, p-value = 0.002817
alternative hypothesis: true location is greater than 0
```

```r
andes |> mean()
```

```
[1] 4.4
```

Con un nivel de significancia del 10%, existe evidencia estadística para afirmar que el nuevo medicamento sí reduce la presión arterial sistólica.