



# Inferencia Estadística

## Ejercicio: Contraste no paramétrico

DSLlab

noviembre, 2024

### Ejercicio 1

**Pregunta:** Un estudio quiere analizar si hay diferencias significativas en la distribución de la duración del sueño entre dos grupos de personas: estudiantes universitarios y trabajadores a tiempo completo. Para ello, se recolectaron dos muestras independientes de tamaño  $n_1 = 20$  para estudiantes y  $n_2 = 30$  para trabajadores. Los datos representan las horas de sueño por noche durante una semana para cada participante.

A partir de los datos recopilados, se desea determinar si hay evidencia estadística para afirmar que las distribuciones de horas de sueño difieren entre los dos grupos.

```
# Muestras de horas de sueño para estudiantes universitarios (grupo 1)
muestras_estudiantes <- c(7.2, 7.0, 6.8, 7.5, 7.1, 7.3, 7.2, 7.4, 7.0,
                           7.6, 7.5, 6.9, 7.2, 7.3, 7.1, 7.4, 7.5, 7.0,
                           7.2, 7.3)

# Muestras de horas de sueño para trabajadores a tiempo completo (grupo 2)
muestras_trabajadores <- c(6.5, 6.3, 6.7, 6.4, 6.6, 6.8, 6.7, 6.9, 6.5,
                           6.2, 6.8, 6.7, 6.3, 6.6, 6.9, 6.5, 6.4, 6.7,
                           6.6, 6.8, 6.7, 6.9, 6.1, 6.3, 6.6, 6.1, 6.9,
                           7.1, 6.3, 6.6)
```

### Ejercicio 2

**Pregunta:** Un estudio realizado en una comunidad recopiló información sobre el género (hombre, mujer) y las preferencias de deportes (fútbol, baloncesto, tenis) de un grupo de

personas. Los datos se muestran en la siguiente tabla de contingencia:

	Fútbol	Baloncesto	Tenis
Hombre	30	20	10
Mujer	20	25	15

Se desea determinar si hay una asociación significativa entre el género y las preferencias de deportes en la población utilizando la prueba de chi-cuadrado de independencia.

### Ejercicio 3

#### Pregunta:

Se llevó a cabo un estudio para comparar los tiempos de respuesta entre dos grupos de personas expuestas a diferentes condiciones de entrenamiento. Para ello, se registraron los tiempos de respuesta (en milisegundos) de 15 personas en el Grupo A y 18 personas en el Grupo B. Los datos obtenidos se muestran a continuación:

Grupo A: 45.1, 38.0, 52.4, 40.1, 47.5, 41.2, 43.5, 39.7, 36.1, 44.2, 50.5, 48.1, 42.4, 46.1, 49.5

Grupo B: 55.3, 60.4, 58.9, 62.0, 54.1, 57.8, 59.3, 63.6, 56.4, 61.7, 58, 64.3, 53.7, 66.1, 67.1, 65.2, 68.3, 50.1

Determine si hay una diferencia significativa en los tiempos de respuesta entre los dos grupos. Para determinar si hay una diferencia significativa en los tiempos de respuesta entre los dos grupos, puedes utilizar la prueba de Mann-Whitney U. Esta prueba no paramétrica se utiliza para comparar las distribuciones de dos muestras independientes y determinar si una muestra tiene valores significativamente mayores o menores que la otra.

### Ejercicio 4

**Pregunta:** Supongamos que estamos trabajando con un conjunto de datos de un modelo de clasificación que predice si los clientes de una tienda en línea harán una compra (Sí o No) basado en varias características. Queremos investigar si existe una asociación entre dos variables categóricas en nuestro conjunto de datos: "Género" (Hombre, Mujer) y "Compra" (Sí, No).

Hemos entrenado un modelo de clasificación y obtenido las siguientes predicciones y datos observados para una muestra de 200 clientes:

Género	Compra: Sí	Compra: No
Hombre	50	30
Mujer	70	50

### Ejercicio 5

**Pregunta:** Supongamos que hemos desarrollado un modelo de clasificación para predecir el riesgo crediticio de los clientes de un banco, clasificándolos como “Alto Riesgo” o “Bajo Riesgo”. Queremos evaluar si existe una asociación significativa entre las predicciones del modelo y los valores reales observados en un conjunto de prueba.

Se toma una muestra de 300 clientes, y se obtiene la siguiente tabla de contingencia que muestra la distribución de las predicciones del modelo y los valores reales observados:

	Real: Alto Riesgo	Real: Bajo Riesgo	Total
Predicción: Alto Riesgo	80	30	110
Predicción: Bajo Riesgo	50	140	190
Total	130	170	300

### Ejercicio 6

**Pregunta:** Un investigador desea evaluar si un nuevo método de enseñanza mejora las calificaciones de los estudiantes. Para ello, selecciona a 10 estudiantes y mide sus calificaciones en una prueba antes y después de aplicar el nuevo método durante un semestre. Los datos obtenidos son los siguientes:

Estudiante	Calificación Antes (X)	Calificación Después (Y)
1	65	75
2	70	78
3	60	65
4	72	80
5	68	74
6	75	85
7	63	68
8	70	77
9	67	73
10	66	69

Para determinar si hay una diferencia significativa en las calificaciones antes y después del nuevo método de enseñanza, se realizará un contraste de hipótesis utilizando la prueba de Wilcoxon para muestras pareadas.

### Ejercicio 7

**Pregunta:** Un nutricionista quiere evaluar el efecto de una nueva dieta en la presión arterial de los pacientes. Se mide la presión arterial sistólica de 12 pacientes antes y después de seguir la dieta durante un mes. Los datos son los siguientes:

Paciente	Antes (X)	Después (Y)
1	135	130
2	140	135
3	150	145
4	145	142
5	160	155
6	155	150
7	148	150
8	160	158
9	142	140
10	138	136
11	150	149
12	145	146

Determinar si hay una diferencia significativa en la presión arterial sistólica antes y después de seguir la dieta utilizando la prueba del signo para observaciones por pares.

### Ejercicio 8

**Pregunta:** Un investigador desea determinar si existe una relación entre el nivel de educación y la preferencia por el tipo de música. Se realiza una encuesta a 200 personas y se obtienen los siguientes datos:

	Rock	Pop	Clásica	Total
Secundaria	30	40	10	80
Universitaria	20	50	30	100
Posgrado	10	5	5	20
<b>Total</b>	60	95	45	200

Determinar si hay una relación significativa entre el nivel de educación y la preferencia por el tipo de música utilizando la prueba chi-cuadrado de independencia.

**Ejercicio 9**

**Pregunta:** Una compañía de supermercados está interesada en la preferencia del consumidor con respecto a dos marcas (A, y B) de refresco que compiten entre sí. Se seleccionan, de modo aleatorio, 15 personas y se les pide que clasifiquen las bebidas mediante una escala del 1 (poca aceptación) al 5 (muchacha aceptación). El orden en la selección de la bebida fue aleatorio. Se obtiene la siguiente información:

ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
A	3	2	5	2	4	5	2	1	3	1	4	3	3	5	2
B	1	1	4	3	4	1	2	1	2	1	2	4	3	2	2

Mediante el uso de la prueba del signo. ¿Se tiene alguna razón para creer que existe una diferencia en la preferencia para estas dos marcas de refrescos? Supón  $\alpha = 0.1$

**Ejercicio 10**

**Pregunta:** Para los datos del ejercicio anterior, emplea la prueba de rangos de signos de Wilcoxon ¿Se obtienen las mismas conclusiones?