
Clase 180 — Tests no paramétricos: Mann-Whitney, Wilcoxon, Kruskal-Wallis

Parte: 3 — Estadística Inferencial y Causal · Fuente: Bruce & Bruce, cap. 3 Resampling and Non-parametric Tests + Conover, Practical Nonparametric Statistics. Duración estimada: 70 min.

Clase 180 — Tests no paramétricos: Mann-Whitney, Wilcoxon, Kruskal-Wallis

Parte: 3 — Estadística Inferencial y Causal · Fuente: Bruce & Bruce, cap. 3 Resampling and Non-parametric Tests + Conover, Practical Nonparametric Statistics. Duración estimada: 70 min.

Objetivo

Aplicar las tres alternativas no paramétricas más usadas: Mann-Whitney U (= dos muestras independientes, análogo a Welch's t), Wilcoxon signed-rank (= pareado, análogo a `ttest_rel`) y Kruskal-Wallis (= ≥ 3 grupos, análogo a ANOVA one-way). Saber cuándo elegirlos sobre los paramétricos: muestras chicas con datos visiblemente asimétricos, datos ordinales (Likert, ranks), o presencia de outliers extremos.

Resultados de aprendizaje

Al finalizar, el estudiante podrá:

- Reconocer las 3 situaciones en que un test no paramétrico es preferible al paramétrico (n chico + asimetría, ordinal, outliers).
- Aplicar `scipy.stats.mannwhitneyu`, `wilcoxon`, `kruskal` con los argumentos correctos (`alternative`, `method='exact'` vs `'asymptotic'`).
- Interpretar que los no paramétricos testean distribuciones (estocásticamente iguales) o medianas, no medias.
- Reportar effect size no paramétrico: rank-biserial correlation (Mann-Whitney) o ϵ^2 / η^2_H (Kruskal-Wallis).
- Hacer post-hoc no paramétrico tras Kruskal con Dunn's test (`scikit-posthocs`) y corrección por múltiples comparaciones.

Temas

#	Test	Reemplaza a	Para qué
1	Mann-Whitney U (Wilcoxon rank-sum)	<code>ttest_ind</code> Welch	2 grupos independientes
2	Wilcoxon signed-rank	<code>ttest_rel</code>	Pareado / 1 muestra contra mediana
3	Kruskal-Wallis H	ANOVA one-way	≥ 3 grupos independientes
4	Dunn's test (post-hoc)	Tukey HSD	Pares post Kruskal
5	Cliff's δ / rank-biserial r	Cohen's d	Effect size no paramétrico

Definiciones y características

- Test no paramétrico: no asume forma específica de la distribución (no requiere normalidad). Trabaja sobre rangos de los datos.
- Mann-Whitney U: para cada par (x_i, y_j) cuenta cuántas veces $x_i > y_j$. Bajo H de distribuciones iguales, U tiene distribución conocida. H: $P(X > Y) = P(X < Y) = 0.5$.
- Wilcoxon signed-rank: para datos pareados o una muestra contra mediana hipotética. Calcula la

diferencia d_i , las rankea por $|d_i|$, suma rangos con signo. Asume simetría de la distribución de diferencias.

- Kruskal-Wallis H: extiende Mann-Whitney a k grupos. $H = (12 / (n(n+1))) \cdot \sum R_i^2/n_i - 3(n+1)$. Bajo H (todas las distribuciones iguales), $H \sim \chi^2(k-1)$ asintóticamente.
- Rank-biserial correlation $r_{rb} = 1 - 2U / (n \cdot n)$: effect size para Mann-Whitney. Va de -1 a 1.
- Cliff's δ : equivalente, $\delta = (\#(x_i > y_j) - \#(x_i < y_j)) / (n \cdot n)$. Interpretación: < 0.147 small, < 0.33 medium, ≥ 0.474 large (Romano et al. 2006).
- Dunn's test: comparaciones pareadas no paramétricas tras Kruskal-Wallis, basadas en la diferencia promedio de rangos. Se ajusta por múltiples tests (Bonferroni, BH).

Dataset / recursos

- `seaborn.load_dataset('tips')`: tip por sex o day.
- Datos con outliers: precios de Airbnb (Kaggle) — cola larga a la derecha por mansiones.
- Likert ordinal: simular respuestas 1–5 con `rng.choice([1,2,3,4,5], p=...)`.
- Librerías: `scipy.stats`, `pingouin`, `scikit-posthocs` (pip install `scikit-posthocs`).

Ejercicios

1. Mann-Whitney: comparar tip entre sex con `scipy.stats.mannwhitneyu(a, b, alternative='two-sided')`. Compará el p con el del t-test del ejercicio 2 de la Clase 147. Calculá rank-biserial r.
2. Wilcoxon signed-rank: con el dataset simulado de presión arterial antes/después de la Clase 147, aplicá `scipy.stats.wilcoxon(antes, despues)`. Verificá supuesto de simetría con un histograma de las diferencias.
3. Outliers: a un dataset normal `rng.normal(50, 5, 100)` agregale 3 outliers de valor 200. Compará Welch's t-test vs Mann-Whitney contra otro grupo normal — el Mann-Whitney es mucho más robusto.
4. Kruskal-Wallis: aplicalo a `body_mass_g` por species en penguins. Comparalo con el ANOVA de la Clase 149.
5. Post-hoc Dunn: con `scikit_posthocs.posthoc_dunn(penguins, val_col='body_mass_g', group_col='species', p_adjust='holm')` identificá qué pares difieren.

Homework verifiable

Tomar el dataset de Airbnb por neighborhood (o un sintético equivalente con cola larga):

1. Verificar normalidad por grupo (Shapiro o KS). Mostrar que se rechaza.
2. Comparar precio entre 4 vecindarios con Kruskal-Wallis.
3. Post-hoc Dunn con corrección Holm.
4. Reportar mediana \pm IQR por grupo (no mean \pm SD, que es engañoso con asimetría).
5. Comparar conclusiones con las que daría un ANOVA clásico ingenuo.

Criterio de aceptación: el reporte debe usar mediana/IQR (no media/SD), identificar al menos un par significativo tras Dunn-Holm, y explicar en 2 líneas por qué ANOVA sería sospechoso aquí (cola larga inflando la varianza del grupo con outliers).

Errores comunes

Síntoma / mensaje	Causa y cómo arreglar
-------------------	-----------------------

Aplico Mann-Whitney y reporto "la media di	El test no es sobre medias; es sobre la pr
Aplico Wilcoxon a diferencias muy asimétri	El signed-rank asume simetría de las difer
Uso Mann-Whitney con $n=10^6$ y se vuelve len	Es $O(n \log n)$ por el ranking, pero scipy l
Reporto Kruskal-Wallis sin post-hoc y digo	Kruskal solo te dice que al menos uno difi
Aplico no paramétrico "para ir a la segura	El t-test tiene más poder cuando sus supue

Preguntas frecuentes

¿Mann-Whitney testea medianas?

Solo si las distribuciones tienen la misma forma (mismo shape, distinto location). En general testea $P(X > Y) \neq 0.5$, que es una afirmación sobre superioridad estocástica, no sobre la mediana per se.

¿Wilcoxon o test de signos?

Wilcoxon usa magnitud de las diferencias (rangos) → más poderoso. Test de signos solo usa la dirección (positivo/negativo) → más robusto pero menos poderoso. Si dudás de la simetría, signos.

¿Cuánto poder pierdo usando no paramétrico cuando los supuestos se cumplen?

Para Mann-Whitney vs t-test con datos normales, la eficiencia asintótica relativa es $3/\pi \approx 0.955$ — perdés $\approx 5\%$ de poder. Si los datos son no normales, podés ganar mucho. Por eso "no paramétrico por default" no es una mala estrategia para n chico.

¿alternative='greater' significa lo mismo que en t-test?

Sí, pero referido a la dirección de la dominancia estocástica (no a la media). alternative='greater' en Mann-Whitney ↔ "la distribución de X tiende a producir valores mayores que la de Y".

¿Qué hago con datos ordinales (Likert 1–5)?

No paramétrico siempre. Ranks son la operación natural sobre escalas ordinales. Mann-Whitney para 2 grupos, Kruskal-Wallis para ≥ 3 .

Referencias

- Bruce & Bruce, cap. 3 — Resampling and Non-parametric Tests.
- Conover, W.J. (1999), Practical Nonparametric Statistics (3rd ed.) — referencia canónica.
- Romano et al. (2006) — interpretación de Cliff's δ .
- scipy.stats.mannwhitneyu, wilcoxon, kruskal.
- scikit-posthocs — Dunn, Conover, Nemenyi.

Siguiente clase

Clase 181 — Corrección de comparaciones múltiples (Bonferroni, FDR)

Apéndice: notebook (primer bloque)

Primera celda ejecutable del notebook de la clase.

```
# Imports y configuración inicial
```

Archivos complementarios

- notebook.ipynb