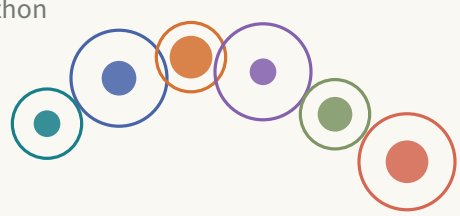
MANUAL VISUAL · DESCARGA GRATUITA

12

Estadística aplicada

12 pruebas para analizar, comparar y comprender datos con criterio.

Fundamentos · aplicación técnica · ejemplos reproducibles en Python



COMPARAR · RELACIONAR · EVALUAR · INTERPRETAR

Una guía para elegir la prueba según la pregunta, el tipo de datos y el diseño del estudio.

DANTE AGUIRRE BRAVO

Material de estudio abierto para quienes exploran la ciencia de datos y la estadística aplicada.

Una guía rápida de selección

Empieza por la pregunta y el diseño. El tipo de variable, la relación entre observaciones y el número de grupos ayudan a identificar qué contraste puede responderla.

01 ¿Qué observas?

Medias, tasas, categorías, rangos o relaciones.

02 ¿Qué diseño tienes?

Grupos independientes, pares o mediciones repetidas.

03 ¿Qué quieres saber?

Diferencia, asociación, distribución o dispersión.

Encuentra tu pregunta

CAPÍTULO Y PÁGINAS DEL MANUAL

COMPARAR MEDIAS

01 t de Welch — 03-05
Dos grupos independientes; compara medias.

02 t pareada — 06-08
Cambio medio en las mismas unidades.

03 ANOVA de un factor — 09-11
Medias de tres o más grupos.

CATEGORÍAS Y PROPORCIONES

04 Chi cuadrado — 12-14
Asociación entre variables categóricas.

05 Z de dos proporciones — 15-17
Compara dos tasas binarias independientes.

ASOCIACIÓN ENTRE VARIABLES

06 Pearson y Spearman — 18-20
Relación lineal o monótona entre pares.

DISTRIBUCIONES Y DISPERSIÓN

07 Shapiro-Wilk — 21-23
Compatibilidad de una muestra con normalidad.

08 Kolmogorov-Smirnov — 24-26
Compara dos distribuciones continuas.

09 Levene — 27-29
Compara la variabilidad entre grupos.

COMPARAR POR RANGOS

10 Mann-Whitney U — 30-32
Dos grupos independientes.

11 Wilcoxon de rangos con signo — 33-35
Cambios pareados respecto de cero.

12 Kruskal-Wallis — 36-38
Tres o más grupos independientes.

TRES CLAVES PARA INTERPRETAR

Diseño primero

Define independencia o pareamiento antes de elegir el contraste.

Magnitud y precisión

Acompaña el valor p con diferencia o efecto e intervalo de confianza.

Evidencia con contexto

Un p alto no demuestra igualdad; considera el diseño y la relevancia práctica.

Comparar medias

Prueba t de Welch

01 Fundamentos 02 Aplicación técnica 03 Ejemplo y Python

¿Difieren los promedios de dos grupos independientes?

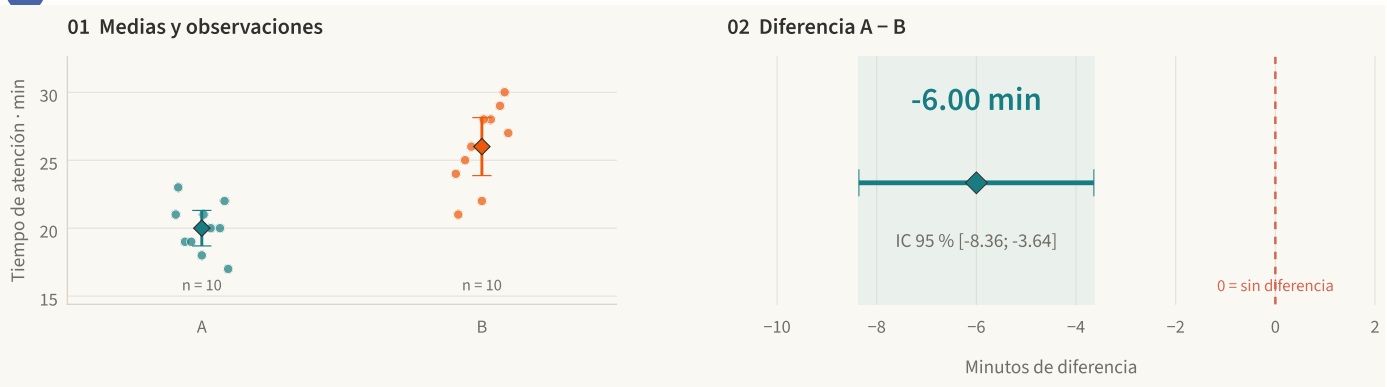
Compara una diferencia de medias con la incertidumbre de estimarla. Permite que los grupos tengan varianzas y tamaños distintos.

LA IDEA DETRÁS DE LA PRUEBA

La media resume el nivel esperado de una variable numérica. Una diferencia observada puede surgir por variación del muestreo aunque las poblaciones tengan la misma media. Welch estandariza esa diferencia mediante un error estándar que estima la variabilidad de cada grupo por separado.

Su distribución de referencia es una t con grados de libertad ajustados. Un valor absoluto grande de t indica una separación difícil de explicar bajo la hipótesis de igualdad. El signo depende del orden A menos B; el contraste bilateral detecta diferencias en ambas direcciones.

1 Prueba t de Welch · lectura visual



Cada punto es un trámite; rombos y barras muestran medias e IC del 95 %. A la derecha, el IC de la diferencia queda por debajo de cero.

EL ESTADÍSTICO

$$t = \frac{\bar{x}_A - \bar{x}_B}{\sqrt{s_A^2/n_A + s_B^2/n_B}}$$

s^2 es la varianza muestral; n es el tamaño del grupo. Los grados de libertad se aproximan con Welch-Satterthwaite.

HIPÓTESIS NULA H0

$$\mu_A = \mu_B$$

Las medias poblacionales son iguales: μA = μB.

HIPÓTESIS ALTERNATIVA H1

$$\mu_A \neq \mu_B$$

Las medias poblacionales difieren: μA ≠ μB.

Prueba t de Welch

Comparar medias

Aplicar con criterio.

CUÁNDO USARLA

Dos grupos distintos, una respuesta cuantitativa y observaciones independientes. Es una opción habitual para comparar medias sin imponer igualdad de varianzas.

LÍMITE PRINCIPAL

Welch no corrige sesgos de selección, dependencia entre registros ni valores extremos. Con muestras pequeñas y colas muy pesadas, la inferencia sobre medias puede ser frágil.

CUATRO PASOS PARA APLICARLA

01

Definir la comparación

Fija la unidad de análisis y la diferencia A menos B antes de mirar el resultado. Establece una diferencia mínima que tenga utilidad práctica, por ejemplo tres minutos de atención.

02

Preparar los grupos

Conserva una medición por unidad independiente. Describe n, media, desviación estándar y posibles atípicos. Si hay varios registros por persona, no los trates como personas distintas.

03

Examinar la distribución

Usa puntos y gráficos Q-Q por grupo. Con pocos casos, evalúa asimetría y atípicos con cuidado. No es necesario que Levene resulte significativo para elegir Welch.

04

Estimar y comunicar

Ejecuta ttest_ind con equal_var=False. Presenta diferencia, intervalo de confianza, t, grados de libertad y p. Contrasta el intervalo con el umbral de utilidad acordado.

CONDICIONES DE VALIDEZ

- Independencia dentro y entre grupos, sustentada por el diseño.
- Respuesta numérica; medias y varianzas finitas.
- Normalidad aproximada por grupo para inferencia exacta con pocos casos; tamaños mayores pueden dar robustez, sin eliminar todo problema.

ALTERNATIVAS SEGÚN EL OBJETIVO

Si son las mismas personas en dos momentos, usa t pareada. Si interesan órdenes o distribuciones, considera Mann-Whitney. Para medias con colas extremas, estudia métodos robustos o remuestreo apropiado al diseño.

PLANTILLA DE REPORTE

La diferencia media A − B fue Δ unidades (IC del 95 % [L, U]); t de Welch(gl) = t, p = p. Se informa además n y dispersión por grupo.

CÓMO LEER EL VALOR P

Suponiendo H0 y el modelo, p es la probabilidad de un resultado tan extremo o más que el observado. No es la probabilidad de que H0 sea verdadera. El umbral α = 0,05 es una referencia que se fija antes del análisis.

Prueba t de Welch

Comparar medias

01 Fundamentos 02 Aplicación técnica 03 Ejemplo y Python

CASO APLICABLE

Una unidad municipal compara el tiempo de atención de dos módulos. Se observan diez trámites distintos por módulo, sin usuarios repetidos. Los datos son minutos y sirven para practicar; no representan una evaluación real.

PYTHON · EJEMPLO REPRODUCIBLE

```python
import numpy as np
from scipy import stats

a = np.array([18,20,19,22,21,17,23,20,19,21])
b = np.array([22,28,25,30,24,27,21,29,26,28])
r = stats.ttest_ind(a, b, equal_var=False)
ic = r.confidence_interval(confidence_level=0.95)
delta = a.mean() - b.mean()
print(f"Medias A/B: {a.mean():.2f} / {b.mean():.2f}")
print(f"Delta A-B: {delta:.2f} min")
print(f"t={r.statistic:.3f}; gl={r.df:.2f}")
print(f"p={r.pvalue:.6f}")
print(f"IC95%: [{ic.low:.2f}, {ic.high:.2f}]")
```

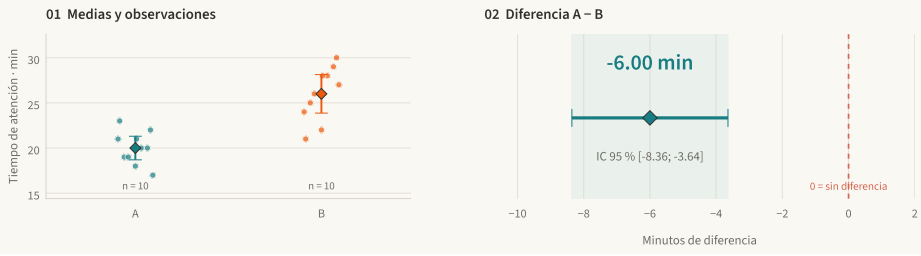
```
$ python ejemplo_01.py
Medias A/B: 20.00 / 26.00
Delta A-B: -6.00 min
t=-5.427; gl=14.92
p=0.000071
IC95%: [-8.36, -3.64]
✓ salida verificada
```

Δ A−B	T
-6.00 min	-5.427
GL WELCH	**P**
14.92	0.000071

LECTURA

El módulo A tiene una media de 20 minutos y B de 26.

2 Cómo se organiza la evidencia del ejemplo



INTERPRETACIÓN

El módulo A tiene una media de 20 minutos y B de 26. La diferencia estimada es −6 minutos. Si el intervalo excluye cero, hay evidencia de una diferencia de medias bajo los supuestos del diseño. El intervalo expresa también la precisión de la estimación.

DECISIÓN Y ALCANCE

Investiga carga y complejidad de los trámites antes de atribuir la diferencia al módulo. Una comparación observacional no identifica por sí sola la causa del menor tiempo.

DOCUMENTACIÓN SciPy · ttest_ind

Comparar cambios

Prueba t pareada

01 Fundamentos 02 Aplicación técnica 03 Ejemplo y Python

¿Existe un cambio medio en las mismas unidades?

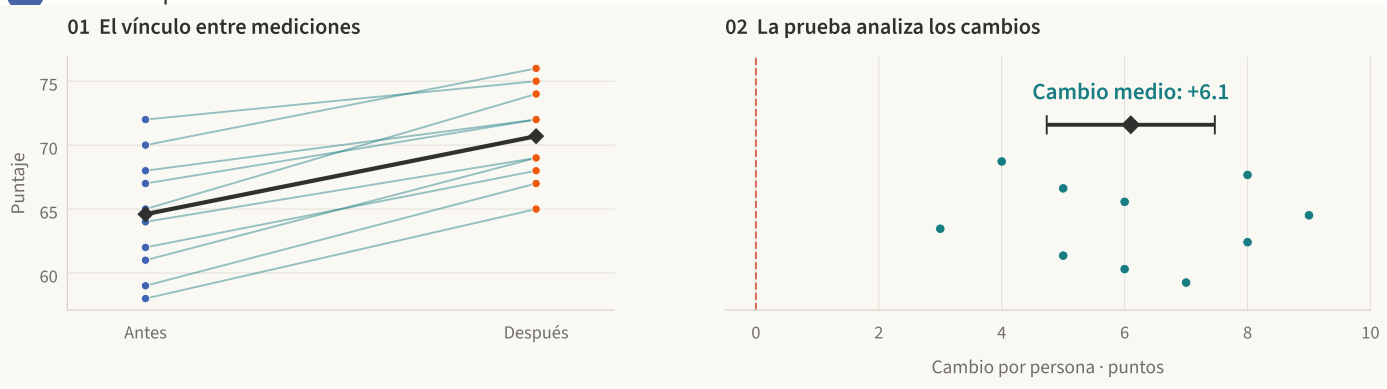
Convierte dos mediciones relacionadas en una sola serie de diferencias. La pregunta se responde sobre el cambio dentro de cada par.

LA IDEA DETRÁS DE LA PRUEBA

Cuando se mide dos veces a una persona, sus valores comparten características. La t pareada aprovecha esa relación y compara el cambio medio con el error estándar de los cambios. Analizar ambos momentos como grupos independientes perdería la estructura del estudio.

La normalidad pertinente es la de las diferencias, no la de cada medición por separado. El contraste se interpreta en la escala original: minutos ganados, puntos de aprendizaje o unidades de producción. Un cambio consistente entre pares suele estimarse con más precisión.

1 Prueba t pareada · lectura visual



Líneas finas: pares completos. Línea oscura: medias. A la derecha, cada punto es un cambio; el rombo resume su media y su IC del 95 %.

EL ESTADÍSTICO

$$d_i = x_{i,\,\text{después}} - x_{i,\,\text{antes}}, \qquad t = \frac{\bar{d}}{s_d/\sqrt{n}}, \qquad \nu = n - 1$$

n es el número de pares completos, sd la desviación estándar de sus diferencias. Los grados de libertad son n − 1.

HIPÓTESIS NULA H0

$$\mu_d = 0$$

El cambio medio poblacional es cero: μd = 0.

HIPÓTESIS ALTERNATIVA H1

$$\mu_d \neq 0$$

El cambio medio poblacional es distinto de cero: μd ≠ 0.

Prueba t pareada

Comparar cambios

Aplicar con criterio.

CUÁNDO USARLA

Antes y después en las mismas personas, expedientes emparejados o unidades vinculadas de forma justificada. Cada par debe corresponder a una unidad independiente de los demás pares.

LÍMITE PRINCIPAL

Una mejora antes y después puede coincidir con aprendizaje, estacionalidad u otros cambios. El contraste detecta cambio medio; no demuestra que una intervención lo haya causado.

CUATRO PASOS PARA APLICARLA

01

Construir los pares

Une los momentos mediante un identificador estable. Comprueba que no existan duplicados ni cruces de personas. Si falta un momento, explica el tratamiento de ese par.

02

Fijar el sentido

Define d = después menos antes y mantén ese orden en el código y el relato. Para tiempos, un cambio negativo puede representar mejora; para puntajes, uno positivo.

03

Revisar las diferencias

Grafica líneas por persona y un Q-Q de d. Busca diferencias atípicas y revisa su origen. La independencia se exige entre pares, aunque dentro del par exista dependencia.

04

Cuantificar el cambio

Aplica ttest_rel(después, antes). Acompaña p con cambio medio e IC del 95 %. Si informas dz = media(d)/sd, aclara que está estandarizado por la variabilidad de los cambios.

CONDICIONES DE VALIDEZ

- Pares correctamente vinculados y comparables en su medición.
- Independencia entre pares y diferencias aproximadamente normales con n pequeño.
- Escala numérica y ausencia de diferencias extremadamente influyentes.

ALTERNATIVAS SEGÚN EL OBJETIVO

Para cambios simétricos sin normalidad, considera Wilcoxon de rangos con signo. Si la simetría no es defendible, evalúa una prueba de signos. Con tres o más momentos, usa modelos de medidas repetidas.

PLANTILLA DE REPORTE

En n pares, el cambio medio después − antes fue Δ (IC del 95 % [L, U]); t(n−1) = t, p = p. Explica qué significa el signo del cambio.

CÓMO LEER EL VALOR P

Suponiendo H0 y el modelo, p es la probabilidad de un resultado tan extremo o más que el observado. No es la probabilidad de que H0 sea verdadera. El umbral α = 0,05 es una referencia que se fija antes del análisis.

Prueba t pareada

Comparar cambios

01 Fundamentos 02 Aplicación técnica 03 Ejemplo y Python

CASO APLICABLE

Diez funcionarios completan una prueba de conocimientos antes y después de una capacitación. Los puntajes son comparables y están ordenados por la misma persona. Se estima el cambio en puntos y no solo si el valor p cruza un umbral.

PYTHON · EJEMPLO REPRODUCIBLE

```python
import numpy as np
from scipy import stats

antes = np.array([58,62,67,61,72,65,70,64,59,68])
despues = np.array([65,68,72,69,75,74,76,69,67,72])
d = despues - antes
r = stats.ttest_rel(despues, antes)
ic = r.confidence_interval(confidence_level=0.95)
dz = d.mean() / d.std(ddof=1)
print(f"Pares: {len(d)}; cambio: {d.mean():.2f}")
print(f"t={r.statistic:.3f}; gl={r.df:.0f}")
print(f"p={r.pvalue:.6f}; dz={dz:.3f}")
print(f"IC95%: [{ic.low:.2f}, {ic.high:.2f}]")
```

```
$ python ejemplo_02.py

Pares: 10; cambio: 6.10
t=10.089; gl=9
p=0.000003; dz=3.190
IC95%: [4.73, 7.47]

✓ salida verificada
```

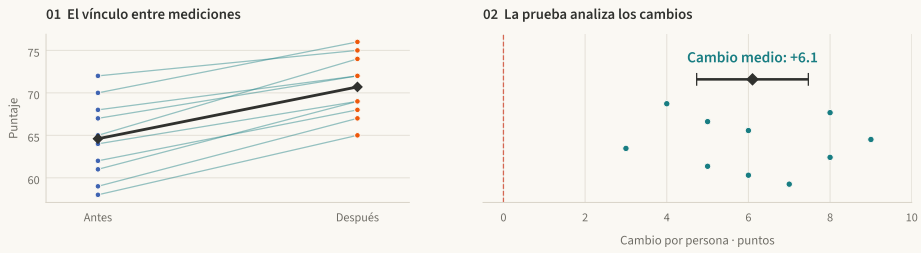
CAMBIO MEDIO	T
6.10 pt	10.089

P	DZ
0.000003	3.190

LECTURA

El cambio medio es positivo: los puntajes suben después de la capacitación.

2 Cómo se organiza la evidencia del ejemplo



INTERPRETACIÓN

El cambio medio es positivo: los puntajes suben después de la capacitación. El intervalo permite juzgar si el aumento podría ser pequeño o relevante. El valor dz describe el cambio relativo a su dispersión entre funcionarios, no un porcentaje de mejora.

DECISIÓN Y ALCANCE

Para evaluar causalidad, incorpora un grupo de comparación y un diseño que controle cambios concurrentes. Revisa además si el nuevo aprendizaje se sostiene y se traduce en desempeño.

DOCUMENTACIÓN SciPy · ttest_rel

Comparar medias

ANOVA de un factor

¿Son iguales las medias de tres o más grupos?

Contrasta variación entre grupos con variación dentro de ellos. El resultado global indica si alguna media difiere, sin señalar todavía cuál.

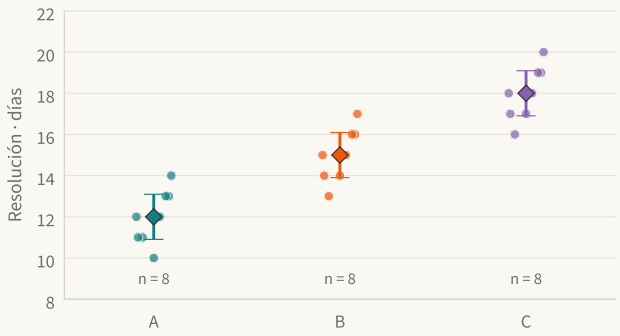
LA IDEA DETRÁS DE LA PRUEBA

ANOVA representa cada observación como media general, efecto del grupo y error individual. Si las medias poblacionales fueran iguales, la variación entre medias muestrales debería ser compatible con la variación residual. El estadístico F mide esa relación.

Un F grande aporta evidencia contra la igualdad global. No implica que todos los pares difieran. El análisis de un factor estudia una clasificación, como tipo de procedimiento; no ajusta automáticamente por edad, carga de trabajo ni otras variables.

1 ANOVA de un factor · lectura visual

01 Tres medias en una comparación



02 Partición de la variación observada

$\eta^2 = 0{,}80$



El resto corresponde a variación dentro de los grupos.

Izquierda: puntos, medias e IC por grupo. Derecha: SC entre / SC total. El 80 % describe asociación con el factor; no una proporción causal.

EL ESTADÍSTICO

$$F = \frac{MS_{\text{entre}}}{MS_{\text{dentro}}} = \frac{SC_{\text{entre}}/(k-1)}{SC_{\text{dentro}}/(N-k)}$$

SC es suma de cuadrados; k es el número de grupos; N es el total de observaciones. Los grados de libertad son k − 1 y N − k.

HIPÓTESIS NULA H0

$$\mu_1 = \cdots = \mu_k$$

Todas las medias poblacionales son iguales: μ1 = … = μk.

HIPÓTESIS ALTERNATIVA H1

$$\exists\, i, j : \ \mu_i \neq \mu_j$$

Al menos una media poblacional difiere de otra.

ANOVA de un factor

Comparar medias

Aplicar con criterio.

CUÁNDO USARLA

Respuesta numérica y un factor categórico con grupos independientes. Es especialmente útil cuando la pregunta científica se refiere a medias en la escala original.

LÍMITE PRINCIPAL

La versión clásica supone varianzas iguales. Comparar cada par con pruebas sin ajuste aumenta la probabilidad de falsos positivos en la familia de comparaciones.

CUATRO PASOS PARA APLICARLA

01

Precisar el factor

Define niveles mutuamente excluyentes y una respuesta medida igual en todos. Inspecciona tamaños y evita grupos definidos después de explorar qué separación produce menor p.

02

Evaluar los residuos

Examina variación dentro de los grupos, gráficos Q-Q y atípicos. La distribución conjunta puede ser multimodal aun si el modelo por grupo es apropiado.

03

Ejecutar el contraste global

Aplica f_oneway para el ANOVA clásico. Informa F, ambos grados de libertad y p. Calcula eta cuadrado como SC entre dividida por SC total para describir variación asociada al factor.

04

Localizar diferencias

Si procede, aplica Tukey HSD para pares en el modelo homocedástico. Si usas ANOVA de Welch, elige comparaciones compatibles, como Games-Howell, y comunica diferencias e intervalos ajustados.

CONDICIONES DE VALIDEZ

- Independencia sustentada por muestreo o asignación.
- Errores aproximadamente normales dentro de cada grupo para la inferencia clásica.
- Varianzas poblacionales iguales en el ANOVA clásico; revisar especialmente diseños desequilibrados.

ALTERNATIVAS SEGÚN EL OBJETIVO

Ante varianzas desiguales y una pregunta sobre medias, considera ANOVA de Welch (equal_var=False desde SciPy 1.16). Para rangos, Kruskal-Wallis cambia el objetivo del contraste. Para ajustar covariables, usa regresión o ANCOVA.

PLANTILLA DE REPORTE

Las medias difirieron globalmente: $F(k-1, N-k) = F$, $p = p$, $\eta^2 =$ valor. Identifica los pares con diferencias mediante un procedimiento posterior con ajuste.

CÓMO LEER EL VALOR P

Suponiendo H0 y el modelo, p es la probabilidad de un resultado tan extremo o más que el observado. No es la probabilidad de que H0 sea verdadera. El umbral $\alpha = 0{,}05$ es una referencia que se fija antes del análisis.

ANOVA de un factor

Comparar medias

01 Fundamentos 02 Aplicación técnica 03 Ejemplo y Python

CASO APLICABLE

Tres modalidades de tramitación se comparan en ocho expedientes independientes por modalidad. La respuesta es el tiempo en días. El ejemplo usa una dispersión similar por grupo para ilustrar el ANOVA clásico y las comparaciones de Tukey.

PYTHON · EJEMPLO REPRODUCIBLE

```python
import numpy as np
from scipy import stats

a = np.array([10,12,11,13,12,14,11,13])
b = np.array([14,15,13,16,15,17,14,16])
c = np.array([17,18,16,19,18,20,17,19])
grupos = [a, b, c]
r = stats.f_oneway(*grupos)
todo = np.concatenate(grupos)
sc_entre = sum(len(g)*(g.mean()-todo.mean())**2
               for g in grupos)
eta2 = sc_entre / sum((todo-todo.mean())**2)
print(f"F(2,21)={r.statistic:.3f}; p={r.pvalue:.8f}")
print(f"Eta cuadrado={eta2:.3f}")
tukey = stats.tukey_hsd(*grupos)
print("p ajustados A-B/A-C/B-C:")
print([f"{p:.3g}" for p in
       tukey.pvalue[np.triu_indices(3,1)]])
```

```
$ python ejemplo_03.py

F(2,21)=42.000; p=0.00000005
Eta cuadrado=0.800
p ajustados A-B/A-C/B-C:
['0.000456', '2.56e-08',
'0.000456']

✓ salida verificada
```

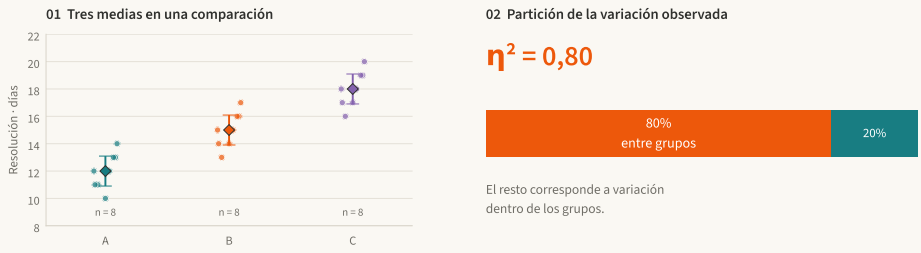
F(2, 21)	P
42.000	4.6e-08

⊠²	TUKEY MÍN.
0.800	2.6e-08

LECTURA

Las medias de A, B y C son 12, 15 y 18 días.

2 Cómo se organiza la evidencia del ejemplo



INTERPRETACIÓN

Las medias de A, B y C son 12, 15 y 18 días. El contraste global evalúa su igualdad; Tukey examina los tres pares con control del error familiar. Eta cuadrado es una proporción de variabilidad observada asociada al factor, no una fracción causal de tiempo ahorrado.

DECISIÓN Y ALCANCE

Antes de escoger una modalidad, compara complejidad de los expedientes y costos de implementación. Una diferencia estadística entre grupos no establece equivalencia en condiciones de operación.

DOCUMENTACIÓN SciPy · f_oneway SciPy · tukey_hsd

Relacionar categorías

Chi cuadrado de independencia

01 Fundamentos 02 Aplicación técnica 03 Ejemplo y Python

¿Están asociadas dos variables categóricas?

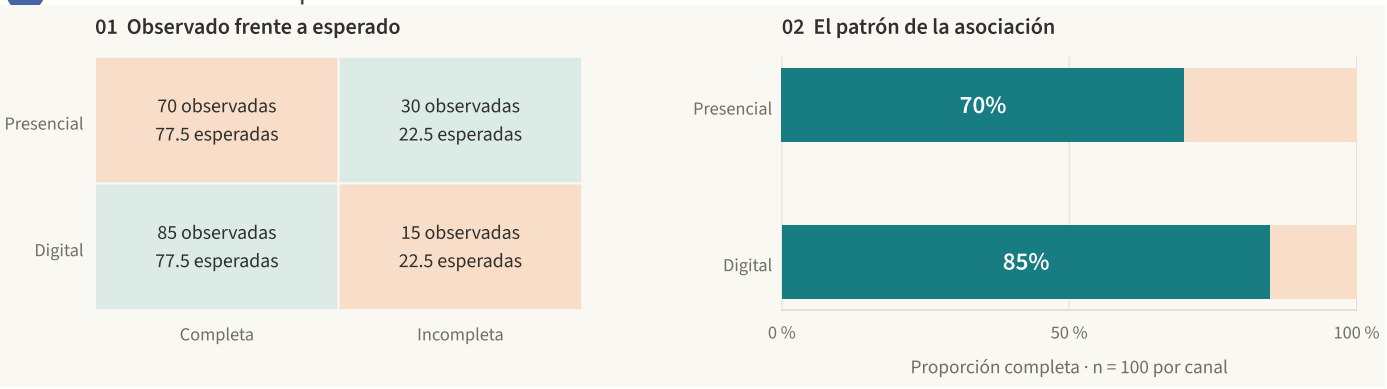
Compara los conteos observados de una tabla con los que esperaríamos si las dos clasificaciones fueran independientes.

LA IDEA DETRÁS DE LA PRUEBA

La independencia significa que la distribución de una categoría no cambia según la otra variable. El contraste compara toda la tabla; no prueba solamente una celda. Su estadístico acumula discrepancias ponderadas por la frecuencia esperada de cada combinación.

Un valor grande de chi cuadrado sugiere una asociación. La magnitud depende también de N: por eso conviene añadir porcentajes condicionados y V de Cramér. Para una tabla 2 × 2, V coincide con el valor absoluto de phi.

1 Chi cuadrado de independencia · lectura visual



La tabla muestra conteos observados y esperados bajo independencia. Las barras usan el mismo denominador y evidencian una diferencia de 15 puntos porcentuales.

EL ESTADÍSTICO

$$\chi^2 = \sum_{i,j} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}, \qquad E_{ij} = \frac{n_{i\cdot}\, n_{\cdot j}}{N}$$

Esperado de una celda = total de fila × total de columna / N. En una tabla R × C, gl = (R − 1)(C − 1).

HIPÓTESIS NULA H0

$$P(X, Y) = P(X)P(Y)$$

Las variables categóricas son independientes en la población.

HIPÓTESIS ALTERNATIVA H1

$$\exists\, x, y : \; P(x, y) \neq P(x)P(y)$$

Existe asociación entre las variables en la población.

Chi cuadrado de independencia

Relacionar categorías

Aplicar con criterio.

CUÁNDO USARLA

Registros clasificados una sola vez en dos variables categóricas, con categorías excluyentes. La entrada al cálculo son conteos, no porcentajes redondeados.

LÍMITE PRINCIPAL

El contraste no demuestra causalidad ni identifica por sí solo qué celdas explican la asociación. Dependencia entre registros y esperados escasos afectan su validez.

CUATRO PASOS PARA APLICARLA

01

Construir la tabla

Cuenta unidades independientes y define claramente filas y columnas. Comprueba que las sumas recuperen N y que cada unidad aparezca en una sola celda.

02

Verificar los esperados

Calcula la tabla esperada bajo independencia. Como regla conservadora para este manual, exige esperados de al menos cinco para usar la aproximación; no es una ley universal.

03

Contrastar la asociación

Ejecuta chi2_contingency. Indica si se aplica corrección de Yates en 2 × 2. El ejemplo la desactiva de forma explícita para reproducir el chi cuadrado de Pearson.

04

Explicar el patrón

Compara porcentajes por fila o columna según la pregunta. Calcula V de Cramér. Si exploras celdas o muchos pares, planifica el ajuste por comparaciones múltiples.

CONDICIONES DE VALIDEZ

- Unidades independientes y categorías exhaustivas y excluyentes.
- Tabla de frecuencias absolutas sin duplicar personas.
- Esperados suficientes para la aproximación chi cuadrado y ausencia de filas o columnas con total cero.

ALTERNATIVAS SEGÚN EL OBJETIVO

En tablas 2 × 2 con conteos pequeños, considera Fisher exacto. Para respuestas binarias pareadas, usa McNemar. En tablas mayores escasas, estudia métodos exactos o Monte Carlo adecuados.

PLANTILLA DE REPORTE

Se observó asociación entre X e Y: $\chi^2(gl, N = n) = valor$, $p = p$, $V = valor$. Añade porcentajes por categoría y especifica la corrección utilizada.

CÓMO LEER EL VALOR P

Suponiendo H0 y el modelo, p es la probabilidad de un resultado tan extremo o más que el observado. No es la probabilidad de que H0 sea verdadera. El umbral $\alpha = 0{,}05$ es una referencia que se fija antes del análisis.

Chi cuadrado de independencia

Relacionar categorías

CASO APLICABLE

Se estudia si el canal de atención se asocia con completar una solicitud. Las filas son presencial y digital; las columnas, completa e incompleta. Hay 100 solicitudes independientes por canal: 70/30 y 85/15.

PYTHON · EJEMPLO REPRODUCIBLE

```python
import numpy as np
from scipy import stats

tabla = np.array([[70, 30], [85, 15]])
chi2, p, gl, esperados = stats.chi2_contingency(
    tabla, correction=False)
n = tabla.sum()
v = np.sqrt(chi2 / (n * (min(tabla.shape)-1)))
print(f"Chi2={chi2:.3f}; gl={gl}; p={p:.6f}")
print(f"N={n}; V de Cramer={v:.3f}")
print("Esperados:", esperados.tolist())
tasas = tabla[:, 0] / tabla.sum(axis=1)
print("Completas por canal:", tasas)
```

```
$ python ejemplo_04.py

Chi2=6.452; gl=1; p=0.011085
N=200; V de Cramer=0.180
Esperados: [[77.5, 22.5], [77.5,
22.5]]
Completas por canal: [0.7  0.85]

✓ salida verificada
```

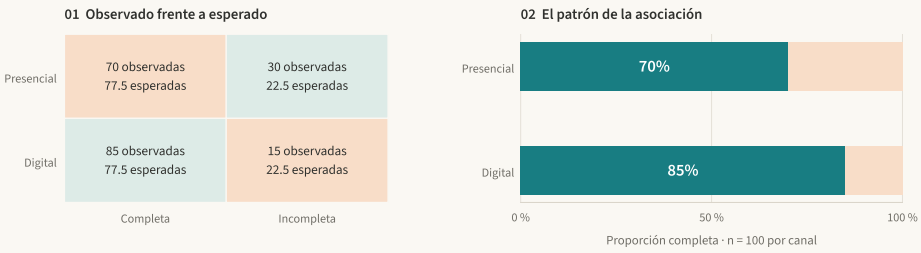
χ^2	GL
6.452	1

P	V CRAMÉR
0.011085	0.180

LECTURA

La proporción de solicitudes completas es 70 % presencial y 85 % digital.

2 Cómo se organiza la evidencia del ejemplo



INTERPRETACIÓN

La proporción de solicitudes completas es 70 % presencial y 85 % digital. El contraste evalúa la asociación global y V describe su intensidad. Los esperados son suficientemente grandes para la aproximación usada en este ejemplo.

DECISIÓN Y ALCANCE

Estudia si los usuarios y los tipos de trámite difieren entre canales. La asociación puede orientar una mejora de formularios, pero no demuestra que trasladar un usuario al canal digital cause la mejora observada.

DOCUMENTACIÓN SciPy · chi2_contingency

Comparar proporciones

Prueba Z de dos proporciones

01 Fundamentos 02 Aplicación técnica 03 Ejemplo y Python

¿Difieren las tasas de éxito de dos grupos?

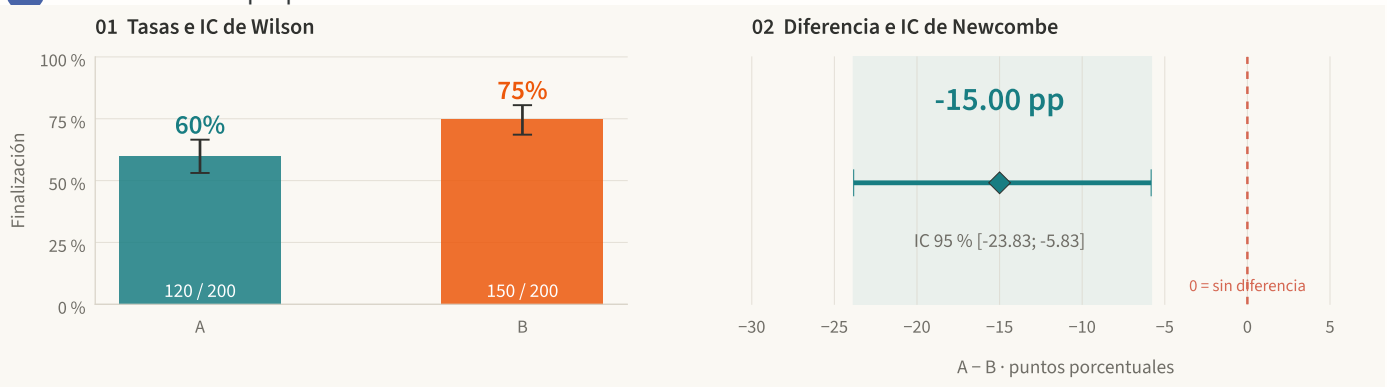
Estima la diferencia entre dos proporciones independientes y la compara con su variación esperada bajo igualdad de tasas.

LA IDEA DETRÁS DE LA PRUEBA

Cada unidad aporta un resultado binario: éxito o fracaso. Al crecer los conteos, la diferencia de proporciones tiene una aproximación normal. La prueba Z usa una estimación combinada de la variabilidad porque la hipótesis nula impone una tasa común.

La diferencia absoluta se expresa en puntos porcentuales. La diferencia relativa divide por una tasa de referencia y responde otra pregunta. Para un intervalo de la diferencia pueden usarse métodos distintos al error combinado del contraste; el ejemplo emplea Newcombe.

1 Prueba Z de dos proporciones · lectura visual



Los intervalos individuales describen cada tasa. El intervalo de la diferencia responde directamente a la comparación: A está 15 puntos porcentuales por debajo de B.

EL ESTADÍSTICO

$$Z = \frac{\widehat{p}_A - \widehat{p}_B}{\sqrt{\widehat{p}(1-\widehat{p})(1/n_A + 1/n_B)}}, \qquad \widehat{p} = \frac{x_A + x_B}{n_A + n_B}$$

p̂A = xA/nA y p̂B = xB/nB. Bajo H0 se usa la proporción combinada p̂ = (xA + xB)/(nA + nB).

HIPÓTESIS NULA H0

$$p_A = p_B$$

Las proporciones poblacionales son iguales: pA = pB.

HIPÓTESIS ALTERNATIVA H1

$$p_A \neq p_B$$

Las proporciones poblacionales son diferentes: pA ≠ pB.

Prueba Z de dos proporciones

Comparar proporciones

Aplicar con criterio.

CUÁNDO USARLA

Dos muestras independientes, una respuesta binaria y denominadores definidos. En un experimento A/B, la unidad de asignación debe coincidir con la unidad del análisis.

LÍMITE PRINCIPAL

Muchos eventos de una misma persona no equivalen a muchos participantes independientes. La aproximación es poco fiable cuando los conteos de éxitos o fracasos son escasos.

CUATRO PASOS PARA APLICARLA

01

Fijar éxito y denominador

Define elegibilidad, exposición y ventana de seguimiento antes del análisis. Cuenta a las personas o unidades asignadas, evitando cambiar el denominador después de observar resultados.

02

Comprobar el diseño

Revisa exclusión entre grupos y comparabilidad. Para la aproximación normal, comprueba que los éxitos y fracasos esperados bajo la tasa combinada sean suficientes; usa diez como guía conservadora.

03

Ejecutar una prueba prevista

Usa proportions_ztest con los éxitos y tamaños en el mismo orden. Mantén bilateral salvo que una alternativa direccional haya sido justificada de antemano.

04

Cuantificar la mejora

Informa tasas, diferencia en puntos porcentuales e intervalo. Si monitoreas repetidamente un A/B para detenerlo al primer p pequeño, necesitas un diseño secuencial que controle el error.

CONDICIONES DE VALIDEZ

- Resultados binarios y unidades independientes.
- Denominadores válidos; una unidad no aparece en ambos grupos.
- Conteos suficientes para usar la aproximación normal.

ALTERNATIVAS SEGÚN EL OBJETIVO

Con conteos pequeños en una tabla 2 × 2, considera Fisher. Para resultados pareados, McNemar. Para ajustar covariables o conglomerados, usa un modelo binomial con una estructura de dependencia apropiada.

PLANTILLA DE REPORTE

A tuvo xA/nA éxitos y B xB/nB. La diferencia A − B fue Δ puntos porcentuales (IC del 95 % [L, U]); Z = valor, p = p.

CÓMO LEER EL VALOR P

Suponiendo H0 y el modelo, p es la probabilidad de un resultado tan extremo o más que el observado. No es la probabilidad de que H0 sea verdadera. El umbral α = 0,05 es una referencia que se fija antes del análisis.

Prueba Z de dos proporciones

Comparar proporciones

01 Fundamentos 02 Aplicación técnica 03 Ejemplo y Python

CASO APLICABLE

Dos mensajes de recordatorio se asignan al azar a personas distintas. En A completan su trámite 120 de 200; en B, 150 de 200. El objetivo es estimar la diferencia de finalización en una misma ventana de seguimiento.

PYTHON · EJEMPLO REPRODUCIBLE

```python
import numpy as np
from statsmodels.stats.proportion import (
    proportions_ztest, confint_proportions_2indep)

x = np.array([120, 150])
n = np.array([200, 200])
z, p = proportions_ztest(x, n, alternative="two-sided")
lo, hi = confint_proportions_2indep(
    x[0], n[0], x[1], n[1],
    method="newcomb", compare="diff", alpha=0.05)
delta = x[0]/n[0] - x[1]/n[1]
print("Tasas:", x/n)
print(f"Z={z:.3f}; p={p:.6f}")
print(f"Delta A-B={100*delta:.2f} pp")
print(f"IC95%: [{100*lo:.2f}, {100*hi:.2f}] pp")
```

```
$ python ejemplo_05.py

Tasas: [0.6  0.75]
Z=-3.203; p=0.001362
Delta A-B=-15.00 pp
IC95%: [-23.83, -5.83] pp

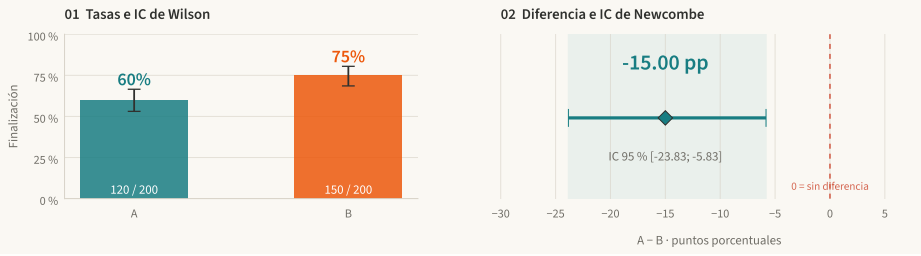
✓ salida verificada
```

A − B	Z
-15.0 pp	-3.203
P	**IC 95 %**
0.001362	-23.8 a -5.8 pp

LECTURA

Las tasas son 60 % y 75 %.

2 Cómo se organiza la evidencia del ejemplo



INTERPRETACIÓN

Las tasas son 60 % y 75 %. A menos B equivale a −15 puntos porcentuales; con B como mejora respecto de A, el aumento relativo es 25 %. Los dos números son correctos, pero describen escalas distintas.

DECISIÓN Y ALCANCE

Si la asignación y el seguimiento se respetaron, el experimento permite una interpretación causal más sólida. Revisa costos, efectos adversos y la precisión del intervalo antes de implementar el mensaje B.

DOCUMENTACIÓN statsmodels · proportions_ztest statsmodels · confint_proportions_2indep

Relacionar variables

Correlación de Pearson y Spearman

01 Fundamentos 02 Aplicación técnica 03 Ejemplo y Python

¿Cómo se relacionan dos variables medidas en las mismas unidades?

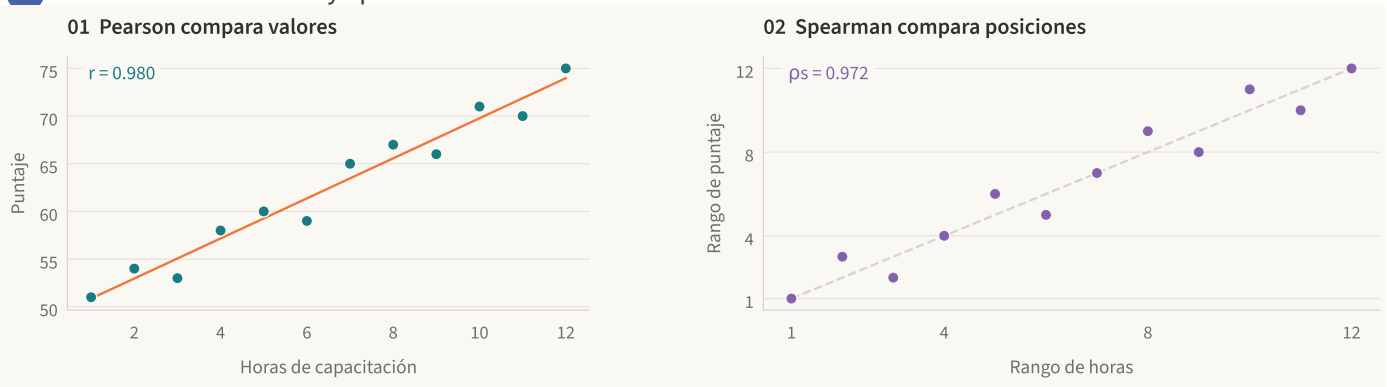
Pearson mide asociación lineal entre valores; Spearman mide asociación monótona entre sus rangos. Elegirlos depende de la relación que se quiere describir.

LA IDEA DETRÁS DE LA PRUEBA

Pearson resume cuánto se alinean los puntos alrededor de una recta; puede ser sensible a valores influyentes. Spearman reemplaza valores por posiciones y detecta relaciones crecientes o decrecientes aunque sean curvas. No elimina todos los efectos de los atípicos.

Cero no significa ausencia de cualquier relación: una forma de U puede dar coeficientes cercanos a cero. La inferencia usual de Pearson se apoya en supuestos de normalidad bivariada; un coeficiente puede calcularse sin ellos, pero su prueba paramétrica requiere justificación.

1 Correlación de Pearson y Spearman · lectura visual



Son los mismos doce pares. La recta izquierda es una guía lineal; la diagonal derecha representa orden idéntico. La concordancia del caso no implica causalidad.

EL ESTADÍSTICO

$$r = \frac{\operatorname{cov}(X, Y)}{s_X s_Y}, \qquad \rho_s = \operatorname{corr}(R_X, R_Y)$$

Ambos coeficientes van de −1 a 1. Con empates se asignan rangos medios; la fórmula simplificada de Spearman sin empates no siempre corresponde.

HIPÓTESIS NULA H0

$$\rho = 0 \quad \text{o} \quad \rho_s = 0$$

Pearson: ρ lineal = 0. Spearman: ρ de rangos = 0.

HIPÓTESIS ALTERNATIVA H1

$$\rho \neq 0 \quad \text{o} \quad \rho_s \neq 0$$

El coeficiente poblacional elegido es distinto de cero.

Correlación de Pearson y Spearman

Relacionar variables

06

Aplicar con criterio.

CUÁNDO USARLA

Pares X,Y por unidad y unidades independientes entre sí. Pearson necesita variables cuantitativas con variación; Spearman admite escalas ordinales cuando los rangos tienen sentido.

LÍMITE PRINCIPAL

Correlación no es causalidad. Un tercer factor, una tendencia temporal o mezclar subgrupos puede crear una asociación aparente o invertir su dirección.

CUATRO PASOS PARA APLICARLA

01

Graficar primero

Haz un diagrama de dispersión. Identifica linealidad, monotonía, subgrupos y observaciones influyentes. Si una variable es constante, el coeficiente no está definido.

02

Elegir el objetivo

Usa Pearson para una relación lineal de valores y Spearman para orden o monotonía. No elijas el método únicamente por cuál produce un p menor.

03

Definir la inferencia

Para Pearson, examina los supuestos de su prueba e intervalo. Para Spearman con pocas observaciones, considera permutar los emparejamientos; el p asintótico puede ser impreciso.

04

Comunicar sin causalidad

Informa coeficiente, n, p, método de inferencia y un gráfico. Añade IC cuando sea posible y distingue asociación del rendimiento de un modelo de predicción.

CONDICIONES DE VALIDEZ

- Cada X se vincula con su Y; pares independientes entre sí.
- Relación relevante para el coeficiente y ausencia de valores constantes.
- Para permutar emparejamientos bajo independencia, las unidades deben ser intercambiables; no usar libremente en series temporales.

ALTERNATIVAS SEGÚN EL OBJETIVO

Si hay una relación de U, usa gráficos y modelos no lineales. Para ajustar confundentes, considera regresión. En datos longitudinales o por centros, modela la dependencia.

PLANTILLA DE REPORTE

Se estimó una asociación r o ρs = valor con n pares; p = p, mediante método especificado. Para Pearson, agrega el IC y describe la forma observada de la relación.

CÓMO LEER EL VALOR P

Suponiendo H0 y el modelo, p es la probabilidad de un resultado tan extremo o más que el observado. No es la probabilidad de que H0 sea verdadera. El umbral α = 0,05 es una referencia que se fija antes del análisis.

Correlación de Pearson y Spearman

Relacionar variables

CASO APLICABLE

Se relacionan horas de capacitación con una evaluación posterior en doce funcionarios. El código calcula ambos coeficientes. Pearson incorpora un IC; Spearman usa una prueba de permutación de emparejamientos con semilla fija.

PYTHON · EJEMPLO REPRODUCIBLE

```python
import numpy as np
from scipy import stats

x = np.arange(1, 13)
y = np.array([51,54,53,58,60,59,65,67,66,71,70,75])
pr = stats.pearsonr(x, y)
ic = pr.confidence_interval()
rho = stats.spearmanr(x, y).statistic
def estadistico(xp):
    return stats.spearmanr(xp, y).statistic
perm = stats.permutation_test(
    (x,), estadistico, permutation_type="pairings",
    alternative="two-sided", n_resamples=9999,
    random_state=42)
print(f"Pearson r={pr.statistic:.3f}; p={pr.pvalue:.8f}")
print(f"IC95% r: [{ic.low:.3f}, {ic.high:.3f}]")
print(f"Spearman rho={rho:.3f}; p perm={perm.pvalue:.4f}")
```

```
$ python ejemplo_06.py

Pearson r=0.980; p=0.00000002
IC95% r: [0.930, 0.995]
Spearman rho=0.972; p perm=0.0002

✓ salida verificada
```

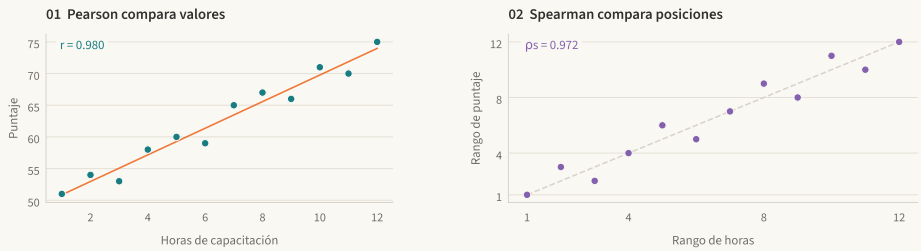
PEARSON R	SPEARMAN ⊠
0.980	0.972

P PEARSON	P PERM.
2.2e-08	0.0002

LECTURA

Ambos coeficientes son positivos y altos en estos datos didácticos.

2 Cómo se organiza la evidencia del ejemplo



INTERPRETACIÓN

Ambos coeficientes son positivos y altos en estos datos didácticos. La concordancia es esperable porque el patrón es creciente y aproximadamente lineal. El p de permutación es una estimación reproducible con resolución limitada por el número de remuestreos.

DECISIÓN Y ALCANCE

No concluyas que aumentar horas causa el incremento del puntaje: motivación y experiencia pueden influir en ambas variables. Una correlación alta tampoco garantiza predicciones precisas fuera del rango observado.

DOCUMENTACIÓN SciPy · pearsonr SciPy · spearmanr SciPy · permutation_test

Diagnosticar distribuciones

Prueba de Shapiro Wilk

01 Fundamentos 02 Aplicación técnica 03 Ejemplo y Python

¿Son los datos compatibles con una distribución normal?

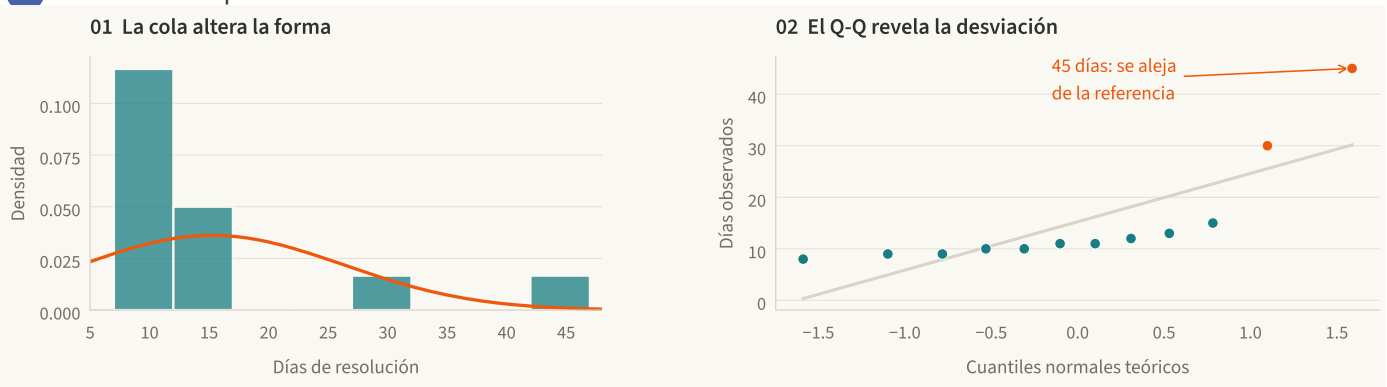
Examina si el patrón de los valores ordenados se parece al esperado bajo normalidad. Es un diagnóstico que debe acompañarse con gráficos y contexto.

LA IDEA DETRÁS DE LA PRUEBA

La normalidad describe una forma de distribución, no la calidad de los datos. Shapiro-Wilk compara una combinación ponderada de valores ordenados con la variación total. Valores bajos de W aportan evidencia de desviaciones como asimetría o colas incompatibles con normalidad.

El tamaño muestral altera la capacidad de detectar desviaciones. Con pocos casos, p alto puede reflejar poca potencia; con muchísimos, una desviación mínima puede ser detectable sin afectar de manera importante el análisis posterior.

1 Prueba de Shapiro Wilk · lectura visual



La curva normal del histograma está ajustada solo como guía visual. En el Q-Q, la cola superior se aparta de la recta; Shapiro-Wilk produce W = 0,641.

EL ESTADÍSTICO

$$W = \frac{\left[\sum_{i=1}^{n} a_i x_{(i)}\right]^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

x(i) son los datos ordenados. Los pesos ai se obtienen de propiedades de estadísticos de orden de una normal. W cercano a 1 indica mayor alineación.

HIPÓTESIS NULA H0

$$X \sim \mathcal{N}(\mu, \sigma^2)$$

La población de la que procede la muestra tiene distribución normal.

HIPÓTESIS ALTERNATIVA H1

$$X \not\sim \mathcal{N}(\mu, \sigma^2)$$

La distribución poblacional no es normal.

Prueba de Shapiro Wilk

Diagnosticar distribuciones

07

Aplicar con criterio.

CUÁNDO USARLA

Una serie de observaciones independientes, o un diagnóstico exploratorio de la variable pertinente. En t pareada interesan las diferencias; en ANOVA, la distribución de los errores por grupo.

LÍMITE PRINCIPAL

p mayor que 0,05 no confirma normalidad. Una prueba de normalidad tampoco verifica independencia, ausencia de sesgo, igualdad de varianzas ni pertinencia de un modelo.

CUATRO PASOS PARA APLICARLA

01

Elegir qué evaluar

Identifica la variable cuyo supuesto requiere normalidad. No juntes grupos con medias distintas para decidir si los errores del ANOVA son normales.

02

Revisar los registros

Examina faltantes, redondeos, truncamiento y atípicos. No elimines observaciones válidas solo para aumentar el valor p. El mínimo computacional de SciPy es de tres datos.

03

Contrastar con un Q-Q

Ejecuta shapiro y compara cuantiles observados con normales. Curvatura sostenida sugiere asimetría; desviaciones en extremos pueden indicar colas distintas.

04

Decidir por impacto

Considera tamaño, equilibrio del diseño y sensibilidad del método posterior. Para más de 5.000 observaciones, SciPy advierte que el p de Shapiro puede ser inexacto.

CONDICIONES DE VALIDEZ

- Muestra apropiada para estudiar una distribución continua.
- Independencia para interpretar el contraste en su forma habitual.
- Evaluación del componente correcto del modelo, junto con inspección gráfica.

ALTERNATIVAS SEGÚN EL OBJETIVO

Considera Q-Q, Anderson-Darling o diagnósticos específicos del modelo. Si transformas una variable, explica cómo cambia el objetivo: analizar logaritmos no equivale a comparar medias aritméticas originales.

PLANTILLA DE REPORTE

Shapiro-Wilk produjo W = valor y p = p, con n observaciones. Describe el patrón del Q-Q y justifica si la desviación es relevante para el método que se utilizará.

CÓMO LEER EL VALOR P

Suponiendo H0 y el modelo, p es la probabilidad de un resultado tan extremo o más que el observado. No es la probabilidad de que H0 sea verdadera. El umbral α = 0,05 es una referencia que se fija antes del análisis.

Prueba de Shapiro Wilk

Diagnosticar distribuciones

07

CASO APLICABLE

Se analizan tiempos de resolución de doce expedientes. La mayor parte tarda entre 8 y 15 días, con dos demoras largas de 30 y 45 días. Se evalúa la compatibilidad con normalidad sin borrar las demoras válidas.

PYTHON · EJEMPLO REPRODUCIBLE

```python
import numpy as np
from scipy import stats

x = np.array([8,9,9,10,10,11,11,12,13,15,30,45])
r = stats.shapiro(x)
print(f"n={len(x)}; W={r.statistic:.3f}")
print(f"p={r.pvalue:.6f}")
print(f"Media={x.mean():.2f}; mediana={np.median(x):.2f}")
# Coordenadas para inspeccionar o graficar un Q-Q
(q, ordenados), ajuste = stats.probplot(x, dist="norm")
print("Cuantiles normales:", np.round(q, 2))
print("Valores ordenados:", ordenados.astype(int))
```

```
$ python ejemplo_07.py
n=12; W=0.641
p=0.000240
Media=15.25; mediana=11.00
Cuantiles normales: [-1.59 -1.1
-0.78 -0.53 -0.31 -0.1   0.1

✓ salida verificada
```

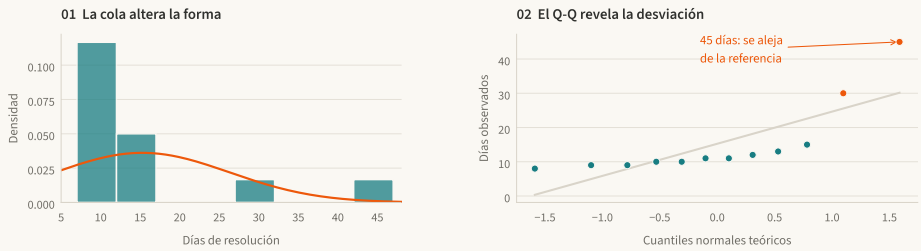
SHAPIRO W	P
0.641	0.000240

N	MEDIANA
12	11.00

LECTURA

Las demoras altas generan una cola derecha visible en el Q-Q y separan media de mediana.

2 Cómo se organiza la evidencia del ejemplo



INTERPRETACIÓN

Las demoras altas generan una cola derecha visible en el Q-Q y separan media de mediana. Un p pequeño apoya la incompatibilidad con normalidad en este ejemplo; no indica por sí solo qué prueba debe sustituir al análisis previsto.

DECISIÓN Y ALCANCE

Investiga por qué algunos expedientes demoran más. Si el objetivo es describir servicio, reporta mediana y percentiles junto con la media; si es inferir, elige una técnica adecuada al parámetro y al diseño.

DOCUMENTACIÓN SciPy · shapiro

Comparar distribuciones

Prueba de Kolmogorov Smirnov

01 Fundamentos 02 Aplicación técnica 03 Ejemplo y Python

¿Difieren dos distribuciones continuas?

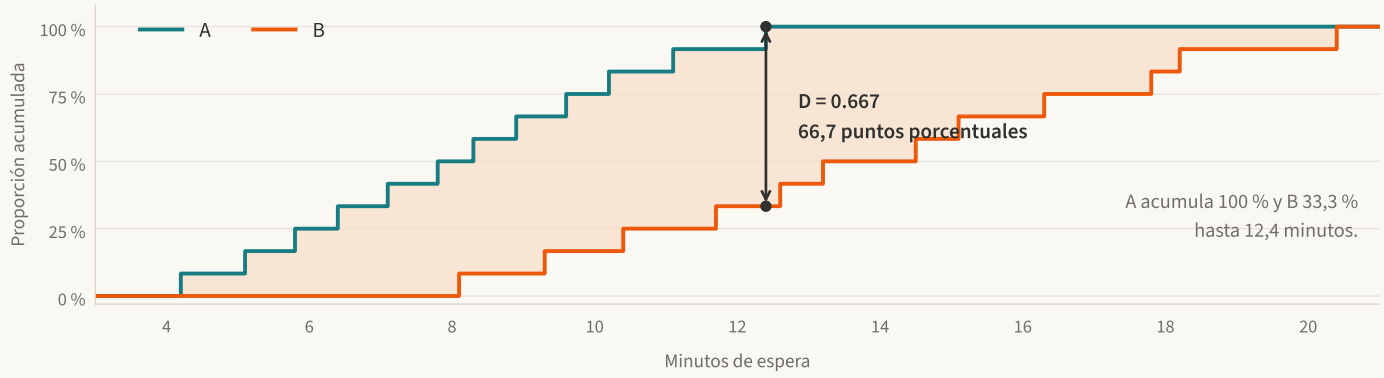
Mide la mayor distancia vertical entre funciones de distribución acumulada. Puede detectar cambios de posición, dispersión o forma.

LA IDEA DETRÁS DE LA PRUEBA

Una distribución acumulada indica la proporción de valores menores o iguales que un umbral. Kolmogorov-Smirnov busca la separación máxima de esas curvas; D varía entre 0 y 1. El contraste bilateral es sensible a diferencias que no se reducen a una media.

En la versión de una muestra, la distribución teórica y sus parámetros deben estar fijados sin estimarlos de esa misma muestra para usar el p estándar. Ajustar media y desviación y luego aplicar KS normal sin corrección altera la calibración; requiere otro procedimiento.

1 Prueba de Kolmogorov Smirnov · lectura visual



D es una distancia vertical entre proporciones acumuladas. La flecha marca su máximo en 12,4 minutos; no representa una diferencia de tiempo.

EL ESTADÍSTICO

$$D_{n,m} = \sup_x |F_n(x) - G_m(x)|$$

Fn y Gm son proporciones acumuladas observadas. En una muestra se reemplaza Gm por una distribución teórica F0 totalmente especificada.

HIPÓTESIS NULA H0

$$F(x) = G(x) \quad \forall x$$

Dos muestras: F(x) = G(x) para todo x. Una muestra: F(x) = F0(x).

HIPÓTESIS ALTERNATIVA H1

$$\exists\, x : \; F(x) \neq G(x)$$

Las funciones de distribución difieren en algún valor de x.

Prueba de Kolmogorov Smirnov

Comparar distribuciones

Aplicar con criterio.

CUÁNDO USARLA

Dos muestras independientes de variables continuas; o una muestra frente a una referencia continua especificada. Es útil para explorar cambios globales en una distribución de tiempos.

LÍMITE PRINCIPAL

No dice dónde está toda la diferencia ni por qué ocurrió. Empates abundantes, datos discretos o dependencia temporal invalidan la calibración continua habitual.

CUATRO PASOS PARA APLICARLA

01

Elegir la versión

Distingue referencia teórica de comparación entre muestras. Para dos períodos, verifica que no sean las mismas unidades medidas dos veces y que no exista dependencia relevante.

02

Graficar acumuladas

Calcula ECDF para ambas muestras y marca el punto de mayor separación. Mantén la misma escala y revisa si el cambio ocurre en la zona central o en las colas.

03

Contrastar

Usa ks_2samp para dos muestras. Indica si el método de p es automático, exacto o asintótico. Para una referencia normal ajustada en los datos, considera Lilliefors o un bootstrap correctamente calibrado.

04

Medir utilidad práctica

Acompaña D y p con medianas, percentiles y diferencias relevantes para el servicio. Si se evalúan muchas variables para detectar cambios, ajusta el error o define prioridades.

CONDICIONES DE VALIDEZ

- Muestras independientes y observaciones independientes en cada una.
- Distribuciones continuas para el p estándar.
- En una muestra, referencia fijada de antemano; parámetros ajustados requieren calibración específica.

ALTERNATIVAS SEGÚN EL OBJETIVO

Para categorías o variables discretas, usa métodos adecuados a conteos o permutaciones válidas. Si el interés son colas extremas, complementa con percentiles o pruebas que ponderen más las colas.

PLANTILLA DE REPORTE

Las distribuciones se compararon con KS bilateral: D = valor, p = p; nA = n y nB = n. Explica la mayor separación acumulada y el método de cálculo del p.

CÓMO LEER EL VALOR P

Suponiendo H0 y el modelo, p es la probabilidad de un resultado tan extremo o más que el observado. No es la probabilidad de que H0 sea verdadera. El umbral α = 0,05 es una referencia que se fija antes del análisis.

Prueba de Kolmogorov Smirnov

Comparar distribuciones

CASO APLICABLE

Se comparan minutos de espera de usuarios diferentes en dos períodos. El registro conserva decimales para evitar empates por redondeo. Los valores de B están más desplazados hacia esperas largas; el código calcula el contraste bilateral.

PYTHON · EJEMPLO REPRODUCIBLE

```python
import numpy as np
from scipy import stats

a = np.array([4.2,5.1,5.8,6.4,7.1,7.8,
              8.3,8.9,9.6,10.2,11.1,12.4])
b = np.array([8.1,9.3,10.4,11.7,12.6,13.2,
              14.5,15.1,16.3,17.8,18.2,20.4])
r = stats.ks_2samp(a, b, alternative="two-sided",
                   method="exact")
print(f"D={r.statistic:.3f}; p={r.pvalue:.6f}")
print(f"Mayor distancia en x={r.statistic_location:.2f}")
print(f"Medianas A/B={np.median(a):.2f}/{np.median(b):.2f}")
```

```
$ python ejemplo_08.py
D=0.667; p=0.007859
Mayor distancia en x=12.40
Medianas A/B=8.05/13.85

✓ salida verificada
```

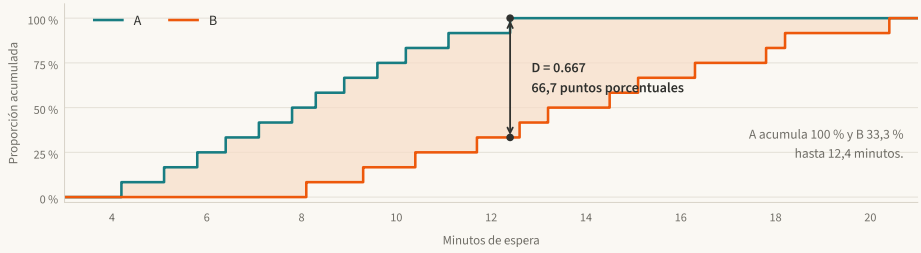
D	P
0.667	0.007859

X*	MEDIANAS
12.40	8.1/13.8

LECTURA

D resume la mayor diferencia en proporción acumulada entre A y B.

2 Cómo se organiza la evidencia del ejemplo



INTERPRETACIÓN

D resume la mayor diferencia en proporción acumulada entre A y B. No es una diferencia de minutos ni un porcentaje de usuarios que empeoró. El gráfico permite ubicar la separación; las medianas aportan una medida en unidades de tiempo.

DECISIÓN Y ALCANCE

Antes de atribuir el cambio al período, compara demanda, mezcla de trámites y capacidad de atención. Si los datos son una serie con autocorrelación, rediseña el análisis para conservar esa estructura.

DOCUMENTACIÓN SciPy · ks_2samp SciPy · kstest

Comparar dispersión

Prueba de Levene

01 Fundamentos 02 Aplicación técnica 03 Ejemplo y Python

¿Tienen los grupos la misma variabilidad?

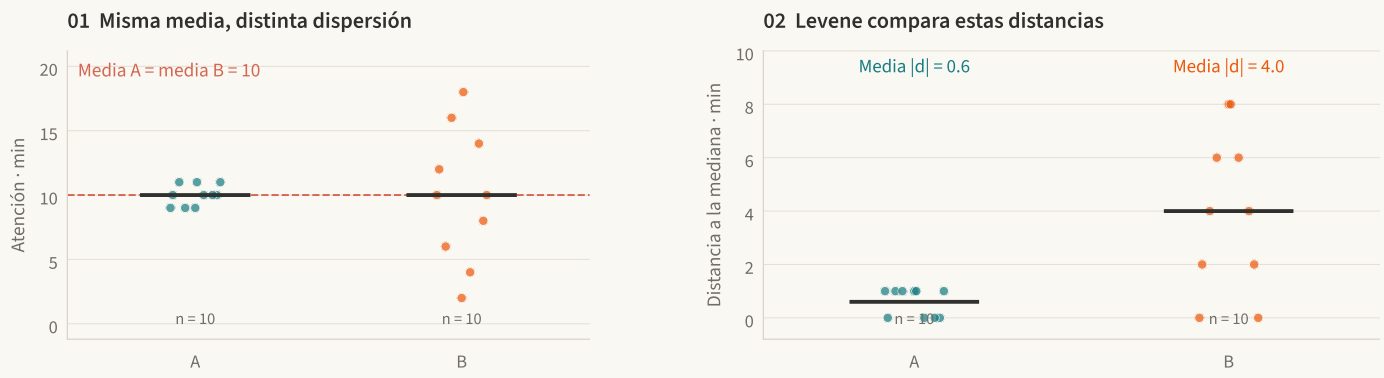
Analiza distancias absolutas al centro de cada grupo. Su objetivo es la dispersión, aunque las medias puedan ser parecidas.

LA IDEA DETRÁS DE LA PRUEBA

Dos servicios pueden tener el mismo promedio y una regularidad muy distinta. Levene transforma cada observación en una distancia absoluta respecto del centro de su grupo y aplica una comparación entre esas distancias. La referencia habitual es una distribución F.

Con centro en la mediana, el cálculo reduce la influencia del centro estimado en distribuciones asimétricas. El contraste sigue siendo una prueba sobre igualdad de varianzas bajo sus condiciones; no mide directamente cuántas veces una varianza supera a otra.

1 Prueba de Levene · lectura visual



La línea discontinua muestra la media común. Levene centrado en la mediana transforma cada valor en su distancia absoluta al centro y compara los grupos transformados.

EL ESTADÍSTICO

$$z_{ij} = |x_{ij} - \widetilde{x}_j|\,, \qquad W = \frac{MS_{\text{entre}}(z)}{MS_{\text{dentro}}(z)}$$

La versión original usa la media. Con centro en la mediana se obtiene la variante de Brown-Forsythe, más robusta frente a asimetría.

HIPÓTESIS NULA H0

$$\sigma_1^2 = \cdots = \sigma_k^2$$

Las varianzas poblacionales son iguales: $\sigma 1^2 = \ldots = \sigma k^2$.

HIPÓTESIS ALTERNATIVA H1

$$\exists\, i, j:\ \sigma_i^2 \neq \sigma_j^2$$

Al menos una varianza poblacional difiere.

Prueba de Levene

Comparar dispersión

Aplicar con criterio.

CUÁNDO USARLA

Dos o más grupos independientes con una respuesta numérica. Es útil cuando la estabilidad del proceso importa o cuando se diagnostican supuestos de un modelo.

LÍMITE PRINCIPAL

No rechazar igualdad no demuestra homocedasticidad. Usar Levene como filtro automático para escoger entre Student y Welch puede introducir una selección innecesaria del análisis.

CUATRO PASOS PARA APLICARLA

01

Distinguir media y variación

Define si la pregunta se refiere al tiempo promedio o a su consistencia. Usa desviaciones estándar, rangos intercuartílicos y puntos individuales para mostrar la dispersión.

02

Seleccionar el centro

Especifica center="median" para Brown-Forsythe o center="mean" para Levene original. No cambies el centro después de observar cuál produce la conclusión preferida.

03

Revisar el tamaño y la forma

Comprueba independencia, colas y datos extremos. En muestras pequeñas, la aproximación puede tener poca potencia o mala calibración; considera remuestreo apropiado al diseño.

04

Separar diagnóstico y elección

Informa W, grados de libertad y p, además de dispersión por grupo. Si el objetivo final es comparar medias con varianzas posiblemente distintas, Welch puede elegirse directamente.

CONDICIONES DE VALIDEZ

- Grupos y observaciones independientes.
- Respuesta cuantitativa; varianzas finitas para interpretar el objetivo.
- Muestreo y tamaños que hagan razonable la aproximación F usada.

ALTERNATIVAS SEGÚN EL OBJETIVO

Bartlett es sensible a la falta de normalidad. Fligner-Killeen ofrece un diagnóstico basado en rangos. Para comparar medias con heterocedasticidad, considera Welch o modelos con errores apropiados.

PLANTILLA DE REPORTE

La variante de Levene centrada en la mediana produjo $W(k-1, N-k)$ = valor, p = p. Añade desviaciones estándar e intervalos o una descripción de su incertidumbre.

CÓMO LEER EL VALOR P

Suponiendo H0 y el modelo, p es la probabilidad de un resultado tan extremo o más que el observado. No es la probabilidad de que H0 sea verdadera. El umbral $\alpha = 0{,}05$ es una referencia que se fija antes del análisis.

Prueba de Levene

Comparar dispersión

01 Fundamentos 02 Aplicación técnica 03 Ejemplo y Python

CASO APLICABLE

Dos ventanillas tienen tiempos cercanos a diez minutos en promedio. Una es estable y la otra presenta tiempos muy cortos y muy largos. Se compara la dispersión mediante la variante centrada en la mediana.

PYTHON · EJEMPLO REPRODUCIBLE

```python
import numpy as np
from scipy import stats

a = np.array([9,10,11,10,9,11,10,10,9,11])
b = np.array([2,4,6,8,10,10,12,14,16,18])
r = stats.levene(a, b, center="median")
print(f"W(1,18)={r.statistic:.3f}; p={r.pvalue:.6f}")
print(f"Medias A/B: {a.mean():.2f}/{b.mean():.2f}")
print(f"DE A/B: {a.std(ddof=1):.2f}/{b.std(ddof=1):.2f}")
razon = b.var(ddof=1) / a.var(ddof=1)
print(f"Razon de varianzas B/A: {razon:.2f}")
```

```
$ python ejemplo_09.py
W(1,18)=12.626; p=0.002271
Medias A/B: 10.00/10.00
DE A/B: 0.82/5.16
Razon de varianzas B/A: 40.00

✓ salida verificada
```

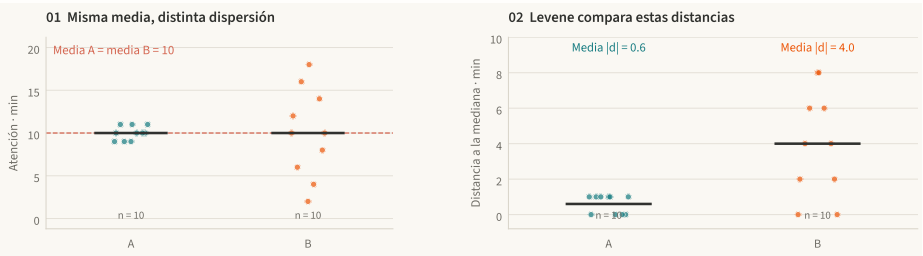
LEVENE W	P
12.626	0.002271

DE A/B	RAZÓN VAR.
0.82/5.16	40.0

LECTURA

Las medias coinciden, pero las desviaciones estándar son muy distintas.

2 Cómo se organiza la evidencia del ejemplo



INTERPRETACIÓN

Las medias coinciden, pero las desviaciones estándar son muy distintas. El contraste se refiere a esa variabilidad. La razón muestral de varianzas aporta contexto descriptivo, aunque con pocos datos también tiene incertidumbre considerable.

DECISIÓN Y ALCANCE

Una mayor dispersión puede afectar la experiencia del usuario aun con la misma media. Examina asignación de trámites, interrupciones y capacidad antes de intervenir sobre el proceso.

DOCUMENTACIÓN SciPy · levene

Comparar rangos

Prueba U de Mann Whitney

01 Fundamentos 02 Aplicación técnica 03 Ejemplo y Python

¿Tienden los valores de un grupo a superar a los del otro?

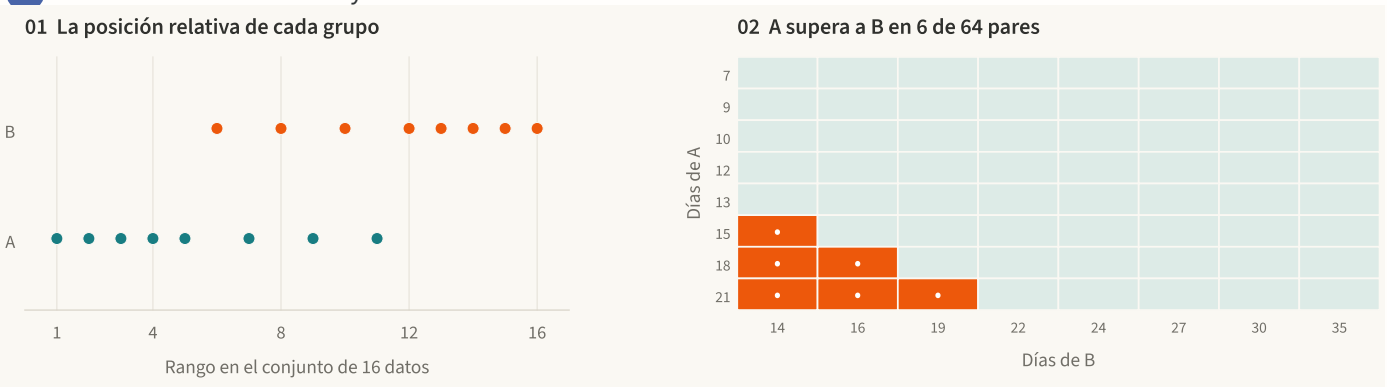
Compara posiciones relativas al ordenar conjuntamente dos grupos independientes. No requiere normalidad, pero tampoco es una prueba universal de medianas.

LA IDEA DETRÁS DE LA PRUEBA

El procedimiento combina las observaciones, asigna rangos medios a los empates y compara la suma de posiciones de cada grupo. U también se entiende como un recuento de victorias por pares entre A y B, con medio punto por empate.

Bajo un modelo de desplazamiento con formas y dispersiones similares, una diferencia de rangos puede interpretarse como diferencia de localización o medianas. Si las formas difieren, esa lectura deja de ser suficiente. Además, no detecta toda diferencia posible de distribución.

1 Prueba U de Mann Whitney · lectura visual



Cada casilla cruza un dato de A con uno de B; naranja significa A > B. U de A = 6 y $\hat{\theta}$ = 6/64 = 0,094: A tiende a registrar menos días.

EL ESTADÍSTICO

$$U_A = R_A - \frac{n_A(n_A+1)}{2}, \qquad \widehat{\theta} = \frac{U_A}{n_A n_B}$$

RA es la suma de rangos de A. $\hat{\theta}$ estima P(A > B) + ½P(A = B). El sentido del efecto cambia al intercambiar el orden de los grupos.

HIPÓTESIS NULA H0

$$F_A = F_B$$

En el contraste habitual, ambas distribuciones poblacionales son iguales.

HIPÓTESIS ALTERNATIVA H1

Diferencia de orden

Hay una diferencia de orden estocástico en la dirección evaluada.

Prueba U de Mann Whitney

Comparar rangos

10

Aplicar con criterio.

CUÁNDO USARLA

Dos grupos independientes; valores numéricos u ordinales que se pueden ordenar de manera coherente. Es adecuado cuando el interés es la posición relativa.

LÍMITE PRINCIPAL

No es una sustitución automática de Welch al fallar Shapiro. Una prueba sobre rangos y una prueba sobre medias responden preguntas distintas.

CUATRO PASOS PARA APLICARLA

01

Definir qué se compara

Especifica si interesa una tendencia a valores mayores o una diferencia de medias. Si buscas medianas, revisa si la suposición de formas similares es defendible.

02

Examinar empates y tamaños

Grafica los valores por grupo. Los empates son frecuentes en escalas ordinales. La opción exacta clásica de SciPy no ajusta por empates; no la fuerces en ese caso.

03

Elegir el cálculo de p

Para muestras pequeñas sin empates puede usarse method="exact". Con empates, considera aproximación asintótica para tamaños adecuados o una permutación válida en muestras pequeñas.

04

Añadir magnitud

Reporta U, p, n y resúmenes por grupo. Calcula $\hat{\theta}$ o correlación biserial de rangos $2\hat{\theta}-1$, indicando el orden A menos B y explicando su interpretación.

CONDICIONES DE VALIDEZ

- Unidades independientes dentro y entre grupos.
- Escala al menos ordinal y medición comparable.
- Para interpretar localización o medianas, distribuciones de forma y dispersión similares.

ALTERNATIVAS SEGÚN EL OBJETIVO

Para las mismas personas, usa Wilcoxon pareado. Para medias, Welch conserva ese objetivo. Si interesa comparar probabilidad de superioridad con heterogeneidad más general, considera Brunner-Munzel.

PLANTILLA DE REPORTE

U de A = valor, p = p (método), nA y nB; $\hat{\theta}$ = valor. Añade medianas y dispersión, evitando afirmar igualdad de medianas como hipótesis universal.

CÓMO LEER EL VALOR P

Suponiendo H0 y el modelo, p es la probabilidad de un resultado tan extremo o más que el observado. No es la probabilidad de que H0 sea verdadera. El umbral $\alpha = 0{,}05$ es una referencia que se fija antes del análisis.

Prueba U de Mann Whitney

Comparar rangos

10

CASO APLICABLE

Se comparan días de tramitación de dos procedimientos en ocho expedientes distintos por grupo. Los valores no tienen empates para ilustrar el cálculo exacto. El procedimiento A suele registrar menos días.

PYTHON · EJEMPLO REPRODUCIBLE

```python
import numpy as np
from scipy import stats

a = np.array([7,9,10,12,13,15,18,21])
b = np.array([14,16,19,22,24,27,30,35])
r = stats.mannwhitneyu(a, b, alternative="two-sided",
                       method="exact")
theta = r.statistic / (len(a)*len(b))
rb = 2*theta - 1
print(f"U de A={r.statistic:.1f}; p={r.pvalue:.6f}")
print(f"Theta P(A>B)+0.5P(empate)={theta:.3f}")
print(f"Biserial de rangos A-B={rb:.3f}")
print(f"Medianas: {np.median(a):.1f}/{np.median(b):.1f}")
```

```
$ python ejemplo_10.py

U de A=6.0; p=0.004662
Theta P(A>B)+0.5P(empate)=0.094
Biserial de rangos A-B=-0.812
Medianas: 12.5/23.0

✓ salida verificada
```

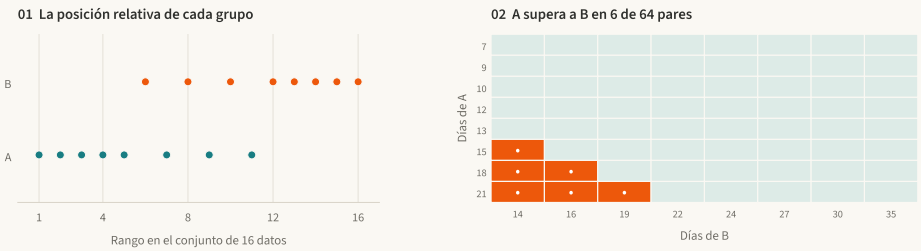
U DE A	P
6	0.004662

θ	MEDIANAS
0.094	12.5/23.0

LECTURA

Un θ menor que 0,5 indica que un expediente de A tiende a registrar menos días que uno de B.

2 Cómo se organiza la evidencia del ejemplo



INTERPRETACIÓN

Un θ menor que 0,5 indica que un expediente de A tiende a registrar menos días que uno de B. El coeficiente biserial negativo conserva esa dirección. El p bilateral evalúa la evidencia de separación bajo la hipótesis y condiciones del método.

DECISIÓN Y ALCANCE

Comunica la probabilidad de superioridad junto con días y dispersión. Para afirmar una diferencia de medianas, justifica adicionalmente las formas; para atribuirla al procedimiento, revisa el diseño y la mezcla de expedientes.

DOCUMENTACIÓN SciPy · mannwhitneyu

Comparar cambios

Wilcoxon de rangos con signo

¿Se desplazan los cambios pareados respecto de cero?

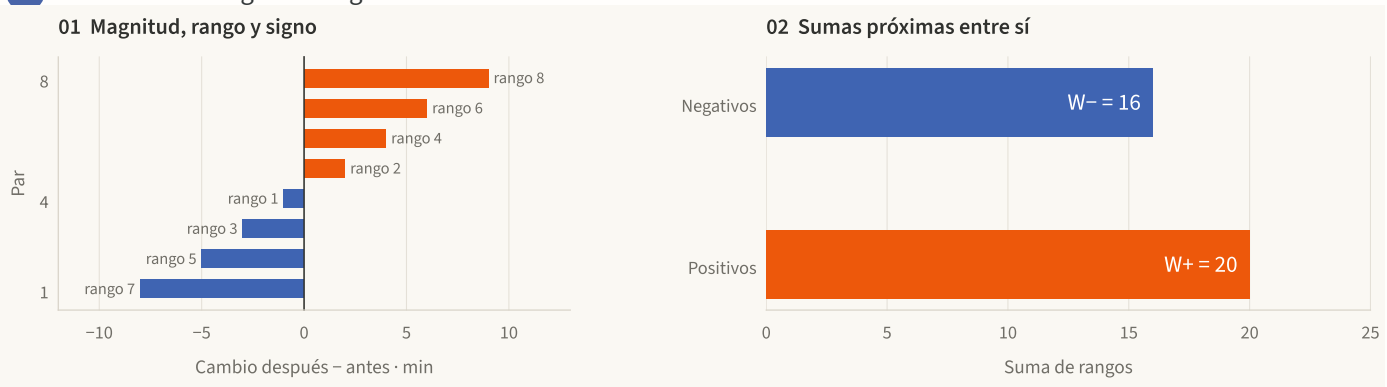
Ordena las magnitudes de las diferencias y conserva sus signos. Aprovecha cuánto cambia cada par sin exigir normalidad de los cambios.

LA IDEA DETRÁS DE LA PRUEBA

Wilcoxon de rangos con signo combina dirección y magnitud ordinal de los cambios. Primero calcula una diferencia por par, después asigna rangos a sus valores absolutos y compara la suma positiva con la negativa. Un predominio de rangos en un sentido sugiere desplazamiento.

Para interpretar el contraste como una prueba de localización o mediana de cambios se necesita simetría. No basta con que la variable sea no normal. Si los cambios son muy asimétricos, una prueba de signos responde a una pregunta más simple sobre su dirección.

1 Wilcoxon de rangos con signo · lectura visual



Las diferencias absolutas reciben rangos de 1 a 8. Wilcoxon suma esos rangos por signo: W = min(20, 16) = 16; p = 0,84375.

EL ESTADÍSTICO

$$W^+ = \sum_{d_i>0} R_i, \quad W^- = \sum_{d_i<0} R_i, \qquad W = \min\left(W^+, W^-\right)$$

Se ordena |d|. En el contraste bilateral, SciPy informa min(W+, W−). Los ceros y empates requieren un tratamiento explícito.

HIPÓTESIS NULA H0

$$D \overset{d}{=} -D$$

La distribución de las diferencias es simétrica alrededor de cero.

HIPÓTESIS ALTERNATIVA H1

Desplazamiento respecto de cero

Las diferencias no están centradas en cero bajo el modelo simétrico.

Wilcoxon de rangos con signo

Comparar cambios

Aplicar con criterio.

CUÁNDO USARLA

Dos mediciones relacionadas o diferencias respecto de un valor fijo. Los pares son independientes entre sí y la magnitud de sus diferencias se puede ordenar con sentido.

LÍMITE PRINCIPAL

Escalas ordinales sin distancias interpretables pueden no justificar restar y ordenar magnitudes. Los empates y cambios exactamente cero alteran la distribución del estadístico.

CUATRO PASOS PARA APLICARLA

01

Vincular y restar

Ordena ambos momentos por identificador y calcula después menos antes. Documenta pérdidas de seguimiento. Si trabajas con decimales, redondea solo según la precisión real de la medición.

02

Examinar simetría

Inspecciona los cambios, no solo cada momento. Busca asimetría pronunciada y muchos ceros. El supuesto de simetría es relevante para interpretar una diferencia de localización.

03

Declarar ceros y método

zero_method="wilcox" excluye ceros. Otros convenios producen resultados diferentes. El método exacto clásico es apropiado sin ceros ni empates de magnitudes; en otros casos considera permutación o aproximación.

04

Reportar cambio y evidencia

Informa estadístico, p, número de pares analizados, ceros y método. Añade mediana de cambios y un efecto pareado, como la diferencia de sumas de rangos dividida por su total.

CONDICIONES DE VALIDEZ

- Pares correctamente vinculados; independencia entre pares.
- Diferencias con magnitudes ordenables; simetría para la lectura de localización.
- Tratamiento de ceros, empates y precisión numérica coherente con el cálculo de p.

ALTERNATIVAS SEGÚN EL OBJETIVO

Usa t pareada si el objetivo es el cambio medio y la inferencia es adecuada. Si no se sostiene simetría y solo interesa el signo, considera una prueba binomial de signos. Para más momentos, estudia Friedman o modelos longitudinales.

PLANTILLA DE REPORTE

Wilcoxon bilateral de rangos con signo: W = valor, p = p, n pares no nulos, método especificado. Reporta mediana de cambios y el sentido después − antes.

CÓMO LEER EL VALOR P

Suponiendo H0 y el modelo, p es la probabilidad de un resultado tan extremo o más que el observado. No es la probabilidad de que H0 sea verdadera. El umbral α = 0,05 es una referencia que se fija antes del análisis.

Wilcoxon de rangos con signo

Comparar cambios

CASO APLICABLE

Se miden minutos para completar una tarea en ocho funcionarios antes y después de una mejora. Los cambios son −8, −5, −3, −1, 2, 4, 6 y 9: no hay ceros ni empates en valor absoluto. El ejemplo permite un cálculo exacto transparente.

PYTHON · EJEMPLO REPRODUCIBLE

```python
import numpy as np
from scipy import stats

antes = np.array([40,42,38,45,41,43,39,44])
despues = np.array([32,37,35,44,43,47,45,53])
d = despues - antes
r = stats.wilcoxon(d, alternative="two-sided",
                   zero_method="wilcox", method="exact")
rangos = stats.rankdata(np.abs(d))
wp = rangos[d > 0].sum()
wm = rangos[d < 0].sum()
rb = (wp-wm)/(wp+wm)
print(f"W={r.statistic:.1f}; p={r.pvalue:.6f}")
print(f"Mediana de cambios={np.median(d):.2f} min")
print(f"W+={wp:.1f}; W-={wm:.1f}; rb={rb:.3f}")
```

```
$ python ejemplo_11.py
W=16.0; p=0.843750
Mediana de cambios=0.50 min
W+=20.0; W-=16.0; rb=0.111

✓ salida verificada
```

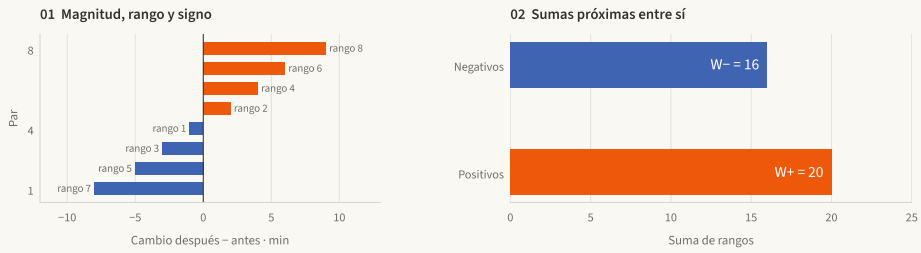
W	P
16	0.843750

MED. CAMBIO	R BISERIAL
0.50	0.111

LECTURA

Los cambios mezclan reducciones y aumentos; sus rangos positivos y negativos son parecidos.

2 Cómo se organiza la evidencia del ejemplo



INTERPRETACIÓN

Los cambios mezclan reducciones y aumentos; sus rangos positivos y negativos son parecidos. Un p alto no prueba ausencia de cambio ni equivalencia: estos ocho pares permiten una conclusión limitada sobre un desplazamiento sistemático.

DECISIÓN Y ALCANCE

Si necesitas demostrar que el cambio es suficientemente pequeño, define un margen de equivalencia y usa un procedimiento adecuado. Aumenta la muestra según precisión o potencia prevista, sin añadir casos hasta obtener el p deseado.

DOCUMENTACIÓN SciPy · wilcoxon

Comparar rangos

Prueba de Kruskal Wallis

01 Fundamentos 02 Aplicación técnica 03 Ejemplo y Python

¿Difieren las posiciones de tres o más grupos independientes?

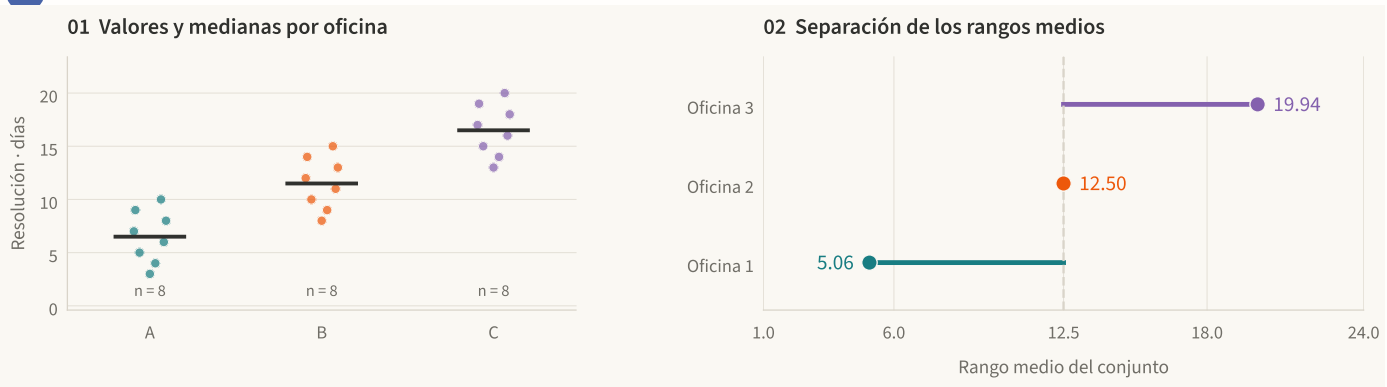
Extiende la comparación de rangos a varios grupos. Evalúa un contraste global y requiere análisis posteriores para localizar diferencias.

LA IDEA DETRÁS DE LA PRUEBA

Kruskal-Wallis ordena todos los valores juntos y examina la separación entre rangos medios de los grupos. No impone normalidad y admite tamaños distintos. La inferencia habitual aproxima la distribución de H por una chi cuadrado.

Con formas y dispersiones similares, el contraste puede leerse como comparación de localizaciones o medianas. Si las formas difieren, puede reflejar otros cambios de distribución. Un resultado global significativo no identifica qué pares difieren ni implica que todos sean distintos.

1 Prueba de Kruskal Wallis · lectura visual



Las barras oscuras son medianas. A la derecha se muestran rangos medios con corrección de empates en el cálculo de H; la línea marca el rango medio global, 12,5.

EL ESTADÍSTICO

$$H = \frac{12}{N(N+1)} \sum_{j=1}^{k} \frac{R_j^2}{n_j} - 3(N+1)$$

Rj es suma de rangos del grupo j. Se corrige H por empates. La referencia asintótica es chi cuadrado con k − 1 grados de libertad.

HIPÓTESIS NULA H0

$$F_1 = \cdots = F_k$$

Las distribuciones de los grupos son iguales; bajo formas comunes, sus localizaciones son iguales.

HIPÓTESIS ALTERNATIVA H1

Patrones de rangos distintos

Algún grupo presenta un patrón de rangos distinto.

Prueba de Kruskal Wallis

Comparar rangos

12

Aplicar con criterio.

CUÁNDO USARLA

Tres o más grupos independientes con una variable al menos ordinal. El objetivo es una diferencia de rangos o, con condiciones adicionales, de localización.

LÍMITE PRINCIPAL

No corrige dependencia ni representa automáticamente una prueba de medianas. Si todos los grupos contienen las mismas personas en distintos momentos, el diseño exige otro análisis.

CUATRO PASOS PARA APLICARLA

01

Definir grupos y unidad

Comprueba que cada unidad aparece una vez. Describe tamaño, mediana y dispersión por grupo y conserva la escala original para interpretar la utilidad de las diferencias.

02

Revisar formas y empates

Inspecciona gráficos por grupo. Una regla común es disponer de al menos cinco observaciones por grupo para la aproximación, sin tomar ese número como garantía universal.

03

Ejecutar el contraste global

Usa kruskal y reporta H, gl y p. Si hay grupos muy pequeños o empates extremos, evalúa una calibración por permutación apropiada. No reduzcas el problema a un umbral automático.

04

Comparar después con ajuste

Planifica Dunn con corrección o pruebas por pares apropiadas con ajuste Holm. El ejemplo usa Mann-Whitney por pares y Holm; sus rangos se calculan por par y no es la prueba de Dunn.

CONDICIONES DE VALIDEZ

- Independencia dentro y entre grupos.
- Mediciones comparables y escala ordenable.
- Formas similares para hablar de medianas; tamaños suficientes para la aproximación del p global.

ALTERNATIVAS SEGÚN EL OBJETIVO

Para medias con varianzas distintas, ANOVA de Welch mantiene ese objetivo. Para mediciones repetidas no paramétricas, considera Friedman. Para ajustar variables de contexto, usa un modelo acorde a la respuesta.

PLANTILLA DE REPORTE

Kruskal-Wallis: $H(k-1) = \text{valor}$, $p = p$, con tamaños por grupo. Informa el método posterior y sus p ajustados, junto con resúmenes en unidades originales.

CÓMO LEER EL VALOR P

Suponiendo H0 y el modelo, p es la probabilidad de un resultado tan extremo o más que el observado. No es la probabilidad de que H0 sea verdadera. El umbral $\alpha = 0{,}05$ es una referencia que se fija antes del análisis.

Prueba de Kruskal Wallis

Comparar rangos

CASO APLICABLE

Tres oficinas registran días de resolución en ocho expedientes independientes cada una. Se analiza si las posiciones de sus tiempos difieren. El ejemplo ejecuta el contraste global y tres comparaciones por pares ajustadas con Holm.

PYTHON · EJEMPLO REPRODUCIBLE

```python
import numpy as np
from scipy import stats
from itertools import combinations
from statsmodels.stats.multitest import multipletests

g = [np.array([3,4,5,6,7,8,9,10]),
     np.array([8,9,10,11,12,13,14,15]),
     np.array([13,14,15,16,17,18,19,20])]
r = stats.kruskal(*g)
pares = list(combinations(range(3), 2))
p = [stats.mannwhitneyu(g[i], g[j],
     alternative="two-sided", method="asymptotic").pvalue
     for i, j in pares]
p_holm = multipletests(p, method="holm")[1]
print(f"H(2)={r.statistic:.3f}; p={r.pvalue:.6f}")
print("Medianas:", [float(np.median(x)) for x in g])
for (i, j), q in zip(pares, p_holm):
    print(f"Oficinas {i+1}-{j+1}: p Holm={q:.6f}")
```

```
$ python ejemplo_12.py

H(2)=17.748; p=0.000140
Medianas: [6.5, 11.5, 16.5]
Oficinas 1-2: p Holm=0.008971
Oficinas 1-3: p Holm=0.002817
Oficinas 2-3: p Holm=0.008971

✓ salida verificada
```

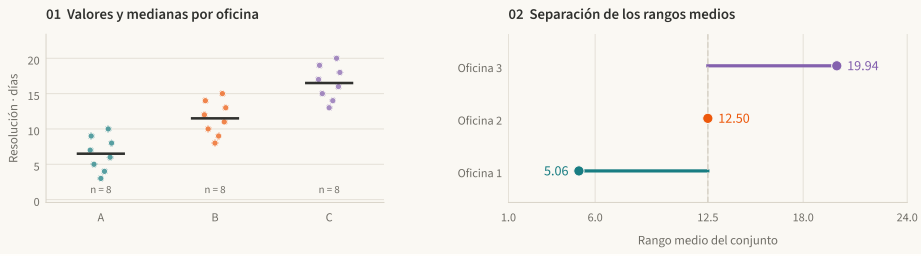
KRUSKAL H	P
17.748	0.000140

MEDIANAS	HOLM MÍN.
6.5 / 11.5 / 16.5	0.0028

LECTURA

Las medianas de las oficinas son 6,5, 11,5 y 16,5 días, con formas similares en estos datos.

2 Cómo se organiza la evidencia del ejemplo



INTERPRETACIÓN

Las medianas de las oficinas son 6,5, 11,5 y 16,5 días, con formas similares en estos datos. El contraste global evalúa el patrón de rangos; las comparaciones ajustadas permiten describir qué pares aportan evidencia sin ignorar la multiplicidad.

DECISIÓN Y ALCANCE

No conviertas automáticamente una oficina en referencia de desempeño: revisa complejidad y recursos. Para una decisión, añade diferencias en días e incertidumbre y define qué magnitud sería operativamente relevante.

DOCUMENTACIÓN SciPy · kruskal SciPy · mannwhitneyu statsmodels · multipletests