
Clase 176 — Test t (una muestra, dos muestras, pareado)

Parte: 3 — Estadística Inferencial y Causal · Fuente: ISLP, cap. 13 + Bruce & Bruce, cap. 3
Statistical Experiments and Significance Testing. Duración estimada: 80 min.

Clase 176 — Test t (una muestra, dos muestras, pareado)

Parte: 3 — Estadística Inferencial y Causal · Fuente: ISLP, cap. 13 + Bruce & Bruce, cap. 3 Statistical Experiments and Significance Testing. Duración estimada: 80 min.

Objetivo

Que el alumno aplique correctamente las tres variantes del test t —una muestra, dos muestras independientes (Welch por default), pareado—, distinga hipótesis nula y alternativa, lea p-value e intervalo de confianza de la salida de `scipy.stats` y `pingouin`, y aprenda a reportar effect size (Cohen's d, Hedges' g) junto con el p-value para no caer en la trampa de "significativo pero irrelevante".

Resultados de aprendizaje

Al finalizar, el estudiante podrá:

- Formular H_0 y H_1 (bilateral / unilateral) para un problema concreto y elegir la variante correcta del test t.
- Ejecutar `scipy.stats.ttest_1samp`, `ttest_ind(equal_var=False)` y `ttest_rel`, interpretando el statistic, pvalue y el atributo `.confidence_interval()` (`scipy` ≥ 1.10).
- Verificar supuestos: normalidad por grupo (Shapiro / Q-Q plot) o invocar TCL si $n \geq 30$.
- Decidir entre test bilateral vs unilateral sin caer en p-hacking (la dirección debe estar fijada antes de mirar los datos).
- Reportar effect size: Cohen's d, Hedges' g corregido para muestras chicas, y su interpretación cualitativa (small/medium/large).

Temas

- H_0 / H_1 , errores tipo I (α) y tipo II (β).
- Test t de una muestra: $t = (\bar{x} - \mu) / (s/\sqrt{n})$, $gl = n - 1$.
- Test t de dos muestras independientes: Welch (varianzas distintas, default moderno) vs Student (varianzas iguales — supuesto fuerte, casi nunca correcto).
- Test t pareado (mismo sujeto antes/después).
- p-value: probabilidad bajo H_0 de observar algo al menos tan extremo. NO es $P(H_1 | \text{datos})$.
- Intervalo de confianza al 95 % como complemento del p-value.
- Complemento moderno: effect size (Cohen's d, Hedges' g, Cliff's δ) — la pregunta "¿es relevante?" que el p-value no responde.

Versión profundizada — 2026

El tema moderno que antes vivía como complemento dentro de esta clase ahora tiene su(s) clase(s) propia(s) con patrón completo, ejercicios y homework:

- Clase 147a — Effect size dedicado: Cohen's d, Hedges' g, Cliff's δ con `pingouin`

Definiciones y características

- Hipótesis nula H_0 : la afirmación "conservadora" — no hay diferencia, o la diferencia es 0. Se busca rechazarla, no probarla.
- Hipótesis alternativa H_1 : lo que querés demostrar. Bilateral (\neq) o unilateral ($>$, $<$).
- p-value: $P(\text{estadístico al menos tan extremo como el observado} \mid H \text{ verdadera})$. Si $p < \alpha$, rechazo H_0 . NO es la probabilidad de que H_0 sea verdadera.
- α (alpha): probabilidad máxima de error tipo I (rechazar H_0 siendo verdadera). Convencional: 0.05. Para problemas críticos: 0.01 o 0.001.
- β (beta): probabilidad de error tipo II (no rechazar H_0 siendo falsa). Poder = $1 - \beta$ (Clase 154).
- Welch's t-test: variante que no asume varianzas iguales. Es el default razonable. `equal_var=True` (Student clásico) es legacy.
- Test pareado: cuando cada observación de un grupo tiene su "par" en el otro (mismo paciente antes/después, misma página web con/sin cambio). Reduce varianza al diferenciar dentro del par.
- Cohen's d: magnitud estandarizada de la diferencia. $d = 0.2/0.5/0.8 \rightarrow \text{small/medium/large}$.
- `alternative='greater'`: unilateral derecho. Solo legítimo si la dirección la fijaste antes de ver los datos.

Dataset / recursos

- `seaborn.load_dataset('tips')`: comparar tip entre `sex='Male'` vs `'Female'`, o `time='Lunch'` vs `'Dinner'`.
- Para pareado: simular un dataset antes/después con `numpy.random.default_rng` (presión arterial pre/post fármaco).
- Librerías: `scipy.stats`, `pingouin` (pip install pingouin), `seaborn`.

Ejercicios

1. Una muestra: con `tips.total_bill`, testá $H_0: \mu = 20$ vs $H_1: \mu \neq 20$ con `scipy.stats.ttest_1samp(tips.total_bill, popmean=20)`. Reportá t, p y el IC95 % (`.confidence_interval()`).
2. Dos muestras (Welch): testá si tip difiere entre `sex='Male'` y `sex='Female'` con `ttest_ind(equal_var=False)`. Calculá Cohen's d a mano y verificá contra `pingouin.ttest`.
3. Pareado: simulá presión arterial antes/después de un fármaco con `rng = np.random.default_rng(0)`: `antes = rng.normal(140, 12, 30)`, `despues = antes - rng.normal(5, 3, 30)`. Aplicá `ttest_rel(antes, despues)` y comparalo contra hacer `ttest_ind` mal (verás cómo el pareado tiene mucho más poder).
4. Bilateral vs unilateral: para el ejercicio 2, repetí con `alternative='greater'` y `'less'`. Observá cómo el p-value se divide ≈ 2 .
5. Significativo vs relevante: generá `grupo_a = rng.normal(100, 15, 10_000)` y `grupo_b = rng.normal(100.5, 15, 10_000)`. El test va a dar $p < 0.001$; calculá Cohen's d y discutí en 2 líneas por qué el resultado "no importa".

Homework verifiable

Sobre tips:

1. Hipótesis: la propina promedio es distinta para `time='Lunch'` y `time='Dinner'`.
2. Verificar normalidad de cada grupo con `pingouin.normality` (Shapiro).
3. Ejecutar `pingouin.ttest` y reportar: T, gl, p-value, IC95 %, Cohen's d, power.
4. Una conclusión de 3 líneas que mencione (a) si rechazás H_0 , (b) la magnitud del efecto en palabras (small/medium/large), (c) si recomendarías ese hallazgo a un dueño de restaurante.

Criterio de aceptación: el reporte debe incluir effect size y la conclusión no puede limitarse a " $p < 0.05$ ". Si Cohen's d es < 0.2 , la respuesta a (c) debería ser "no" aunque el p sea bajo.

Errores comunes

Síntoma / mensaje	Causa y cómo arreglar
Aplico ttest_ind con equal_var=True por co	Student clásico con varianzas desiguales y
$p < 0.05$ con $n = 10^6$ y declaro un hallazgo	"Significancia estadística" con muestra gi
Cambio alternative='greater' después de ve	Eso es p-hacking. La dirección se fija con
Aplico ttest_ind a observaciones pareadas	Pierde poder enormemente y puede dar $p > 0$
El test asume normalidad y mis datos son m	Welch es robusto pero no mágico. Fix: scip

Preguntas frecuentes

¿Welch o Student por default?

Welch siempre. No cuesta nada en poder cuando las varianzas son iguales, y protege contra el caso (muy común) de varianzas distintas. La literatura moderna (Delacre, Lakens & Leys 2017) recomienda abandonar el Student t-test.

¿Tengo que testear normalidad antes del t-test?

No con $n \geq 30$ por grupo (TCL). Con n chico y datos visiblemente asimétricos, sí — y si Shapiro rechaza, mejor pasar a bootstrap o Mann-Whitney. Cuidado: con n enorme, Shapiro rechaza siempre por desviaciones triviales; mirá Q-Q plot como complemento.

¿Qué es BF10 que reporta pinguin?

Factor de Bayes contra H. $BF_{10} > 3$ es evidencia moderada a favor de H; > 10 fuerte. Es la versión bayesiana del p-value y no tiene los problemas del NHST (la testeás directamente en la Clase 158).

¿Por qué el IC95 % de la diferencia y el p-value siempre concuerdan?

Porque son la misma información presentada de otra forma. Si el IC95 % de $\mu_a - \mu_b$ no incluye 0, entonces $p < 0.05$ (bilateral). Reportar el IC es más informativo: muestra magnitud y dirección.

Cohen's d me da 0.3, ¿cómo lo explico al cliente?

"Un efecto pequeño-mediano". Operacionalmente: si dibujás las dos distribuciones, el ≈ 55 % de los individuos del grupo tratamiento superan a la mediana del grupo control (vs 50 % bajo H). Para una conversión, tradúcelo a "incremento absoluto de X puntos porcentuales".

Referencias

- ISLP, cap. 13 — Multiple Testing, intro al p-value.
- Bruce & Bruce, cap. 3 — Statistical Experiments and Significance Testing.
- Delacre, Lakens & Leys (2017), Why Psychologists Should by Default Use Welch's t-test, IRSP.
- Cohen, J. (1988), Statistical Power Analysis for the Behavioral Sciences — referencia canónica de effect size.
- pingouin docs — Vallat, R. (2018), Pingouin: statistics in Python, JOSS.
- scipy.stats.ttest_ind — atención a equal_var y alternative.

Siguiente clase

Clase 177 — Effect size dedicado: Cohen's d, Hedges' g, Cliff's δ con pingouin

Apéndice: notebook (primer bloque)

Primera celda ejecutable del notebook de la clase.

```
# Imports y configuración inicial
```

Archivos complementarios

- notebook.ipynb