CENTRO DE ESTADÍSTICA

Universidad del Azuay

Cuenca,
Junio de 2024
No. 52

El Centro de Estadística de la Universidad del Azuay se establece como un espacio destinado a potenciar y respaldar la investigación mediante el apoyo y la ejecución de proyectos, garantizando un correcto y preciso manejo de los datos; así como, promueve instancias formativas para la comunidad universitaria, generando actividades de aprendizaje permanente vinculadas al empleo de datos, el uso de la estadística y sus múltiples técnicas.

Fernando Córdova León
Coordinador Centro de Estadística

Introducción
PÁG. 2

Supuestos
PÁG. 2

Procedimiento
PÁG. 2

Ejemplo
PÁG. 3

En esta edición:

Herramientas estadísticas: Prueba T de *Student*

Esta edición recopila los resultados obtenidos por estudiantes que, como parte de sus prácticas preprofesionales, realizan actividades relacionadas con el tema de la estadística. El trabajo analiza las pruebas y herramientas usadas para comparar dos muestras o grupos, enfocándose específicamente en la prueba paramétrica T de *Student*.

La motivación principal de la investigación ha sido dotar a las partes interesadas de información que aporte al conocimiento y apoye la toma de decisiones en un futuro.

Herramientas estadísticas: Prueba T de *Student*

Doménica Daniela Crespo
Estudiante de Ingeniería de la Producción

Introducción

La prueba T de *Student*, también conocida como *t-test*, es una herramienta estadística utilizada para determinar si existen diferencias significativas entre las medias de dos grupos. Esta prueba usa la distribución t de *Student*, que es apropiada para muestras pequeñas cuando se desconoce la desviación estándar de la población.

Tipos de prueba

1) **Prueba T para una muestra (*one-sample t-test*)**: Se utiliza para comparar la media de una sola muestra con una media conocida o hipotética de la población.
2) **Prueba T para dos muestras independientes (*independent two-sample t-test*)**: Se emplea para comparar las medias de dos grupos independientes, es decir, cuando los datos de los dos grupos no están relacionados entre sí.
3) **Prueba T para muestras emparejadas (*paired-sample t-test*)**: Se usa cuando los datos de los dos grupos están emparejados o relacionados, como en estudios antes y después o en estudios de casos y controles.

Supuestos

Para que los resultados de la prueba T de *Student* sean válidos, deben cumplirse ciertos supuestos:

- **Normalidad:** Los datos deben seguir una distribución aproximadamente normal. Esto es especialmente importante para muestras pequeñas. Se valida mediante las pruebas de *Shapiro Wilks* y *Kolmogorov Smirnov*, si el p valor es > 0.05 la distribución de la muestra es normal, si es ≤ 0.05 es no normal.
- **Independencia de las observaciones:** Las observaciones deben ser independientes entre sí. Se debe reconocer si los datos pertenecen a muestras relacionadas (las observaciones en la muestra están vinculadas o emparejadas con las observaciones en otra muestra) o son independientes (cada observación en una muestra es independiente de cualquier observación en la otra muestra).
- **Varianzas homogéneas (para la prueba T de dos muestras independientes):** Las varianzas de los dos grupos deben ser aproximadamente iguales. Existen versiones ajustadas de la prueba T (como la corrección de *Welch*) en caso de que este supuesto no se cumpla.

Procedimiento

En este caso se describe el procedimiento de una prueba T para dos muestras independientes.

1) **Establecer las Hipótesis**
   - Ho= Hipótesis nula "Signos de Igualdad" (= , ≥ , ≤). Establece que no hay diferencia significativa entre las medias.
   - Ha= Hipótesis alternativa "Signos de Desigualdad" (≠ , > , <). Plantea que sí hay una diferencia significativa entre las medias.
2) **Establecer el nivel de significancia**

   Valor alfa (α) que indica la probabilidad de rechazar la hipótesis nula cuando es verdadera (Error tipo 1). Puede tomar valores de 0.10, 0.05, o 0.0025. Comúnmente se utiliza 0.05 o 5%.
3) **Establecer los grados de libertad (gl)**

   Los grados de libertad representan el número de valores que son libres de variar en el cálculo de una estadística después de imponer restricciones. Para este caso:

$$gl = n_1 + n_2 - 2$$

   Donde:
   - gl = Grados de libertad

- n1 = Tamaño de la muestra 1
- n2 = Tamaño de la muestra 2

4) **Calcular el estadístico t**

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Donde:

- t = Valor t de *Student*.
- $\bar{x}_1$ = Media de la muestra 1.
- $\bar{x}_2$ = Media de la muestra 2.
- n_1 = Tamaño de la muestra 1.
- n_2 = Tamaño de la muestra 2.
- s_p = Desviación estándar combinada.

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

Donde:

- S_1 = Desv. estándar de la muestra 1.
- S_2 = Desv. estándar de la muestra 2.

5) **Comparar el estadístico t calculado con el valor crítico de t o valor p**

Utilizar la tabla de distribución t para encontrar el valor crítico basado en α y los grados de libertad. Verificar si la prueba es unilateral (una cola) o bilateral (dos colas).

Tabla 1

Extracto de la distribución t de Student

ν grados de libertad							α/2
	0,0005	0,001	0,005	0,01	0,025	0,05	0,1
1	636,619	318,309	63,657	31,821	12,706	6,314	3,078
2	31,599	22,327	9,925	6,965	4,303	2,920	1,886
3	12,924	10,215	5,841	4,541	3,182	2,353	1,638
4	8,610	7,173	4,604	3,747	2,776	2,132	1,533
5	6,869	5,893	4,032	3,365	2,571	2,015	1,476
6	5,959	5,208	3,707	3,143	2,447	1,943	1,440
7	5,408	4,785	3,499	2,998	2,365	1,895	1,415
8	5,041	4,501	3,355	2,896	2,306	1,860	1,397
9	4,781	4,297	3,250	2,821	2,262	1,833	1,383
10	4,587	4,144	3,169	2,764	2,228	1,812	1,372
11	4,437	4,025	3,106	2,718	2,201	1,796	1,363
12	4,318	3,930	3,055	2,681	2,179	1,782	1,356
13	4,221	3,852	3,012	2,650	2,160	1,771	1,350
14	4,140	3,787	2,977	2,624	2,145	1,761	1,345
15	4,073	3,733	2,947	2,602	2,131	1,753	1,341
16	4,015	3,686	2,921	2,583	2,120	1,746	1,337
17	3,965	3,646	2,898	2,567	2,110	1,740	1,333
18	3,922	3,610	2,878	2,552	2,101	1,734	1,330
19	3,883	3,579	2,861	2,539	2,093	1,729	1,328
20	3,850	3,552	2,845	2,528	2,086	1,725	1,325
21	3,819	3,527	2,831	2,518	2,080	1,721	1,323
22	3,792	3,505	2,819	2,508	2,074	1,717	1,321
23	3,768	3,485	2,807	2,500	2,069	1,714	1,319

Nota. Obtenido de https://estdg.blogs.upv.es.

- Comparación con el valor crítico: Si el valor absoluto del estadístico t es mayor que el valor crítico, se rechaza la hipótesis nula.
- Comparación con el valor p: Si el valor p es menor que el nivel de significancia (α), se rechaza la hipótesis nula.

Ejemplo

Carlos cultiva tomates en dos campos separados. Cuando los tomates están listos para ser cosechados, le interesa saber si las alturas de las plantas difieren entre los dos campos.

	Campo A	Campo B
Muestra	22	24
Media	1,3m	1,6m
Desviación	0,5m	0,3m

- Ho: Altura de las plantas del Campo A = Altura de las plantas del Campo B.
- Ha: Altura de las plantas del Campo A ≠ Altura de las plantas del Campo B.
- Nivel de significancia = 0,05.
- Grados de libertad: 22 + 24 – 2 = 44.

Desviación estándar combinada:

$$s_p = \sqrt{\frac{(22 - 1) * 0{,}5^2 + (24 - 1) * 0{,}3^2}{22 + 24 - 2}}$$

$$s_p = 0{,}4079$$

Valor t:

$$t = \frac{1{,}3 - 1{,}6}{0{,}4079 * \sqrt{\frac{1}{22} + \frac{1}{24}}}$$

$$t = -2{,}4919$$

- Valor crítico de la distribución t para dos colas = 2,015.
- Valor p de la distribución t para dos colas es aproximadamente de 0,02.

Interpretación:

Dado que este valor p es menor que 0.05 y t es mayor que el valor crítico, se rechaza la hipótesis nula (Ho) a un nivel de significancia de 0.05, indicando que la diferencia entre las medias de altura de las plantas de tomate es estadísticamente significativa.