



Maestría en Ciencia de Datos

**Analisis y toma de decisiones ante problemas propuestos**

Dirigido a nuestro cliente: Ana Delia Olvera Cervantes  
Desarrollado por: Melchor Nolasco Cosijoeza

31 de enero de 2025

# cosijoeza-melchor-nolasco

January 31, 2025

## 1 Examen 2

### 1.1 Problema 1

*Se planea en un restaurante eliminar del menú el pollo frito. Se afirma que las ventas habían descendido por debajo de la media histórica de 4500.*

*¿Parece una decisión adecuada si en una muestra de  $n = 144$  registros se observa que  $\bar{X} = 4477$ ,  $s = 1128$  con  $\alpha = 0.02$ ?*

Para poder eliminar el pollo frito del menú, primero debemos comprobar que realmente las ventas han bajado.

Dado que no podemos revisar las ventas de pollo frito desde que el restaurante abrió, vamos a tomar una muestra de **144 semanas** recientes y revisamos cuántas porciones de pollo frito se vendieron en promedio cada semana.

- Promedio de ventas por semana: 4477 porciones
- Antes, el promedio histórico era: 4500 porciones
- Las ventas varían mucho, con una desviación estándar de 1128

**¿Esta diferencia de 23 porciones por semana es realmente significativa o solo es una fluctuación normal?**

Hay una forma de comprobar si algo ha cambiado significativamente o si es solo una variación normal.

Usamos un test de hipótesis, que es como un “juicio estadístico” para decidir si las ventas han bajado lo suficiente como para eliminar el pollo frito del menú.

#### 1.1.1 Dos posibles decisiones, así que establecemos dos hipótesis

- Las ventas no han bajado significativamente (siguen en 4500 o más). No ha cambiado mucho entonces dejamos el pollo frito en el menú.
- Las ventas han bajado significativamente (son menores a 4500). Si ha bajado significativamente entonces quitamos el pollo frito.

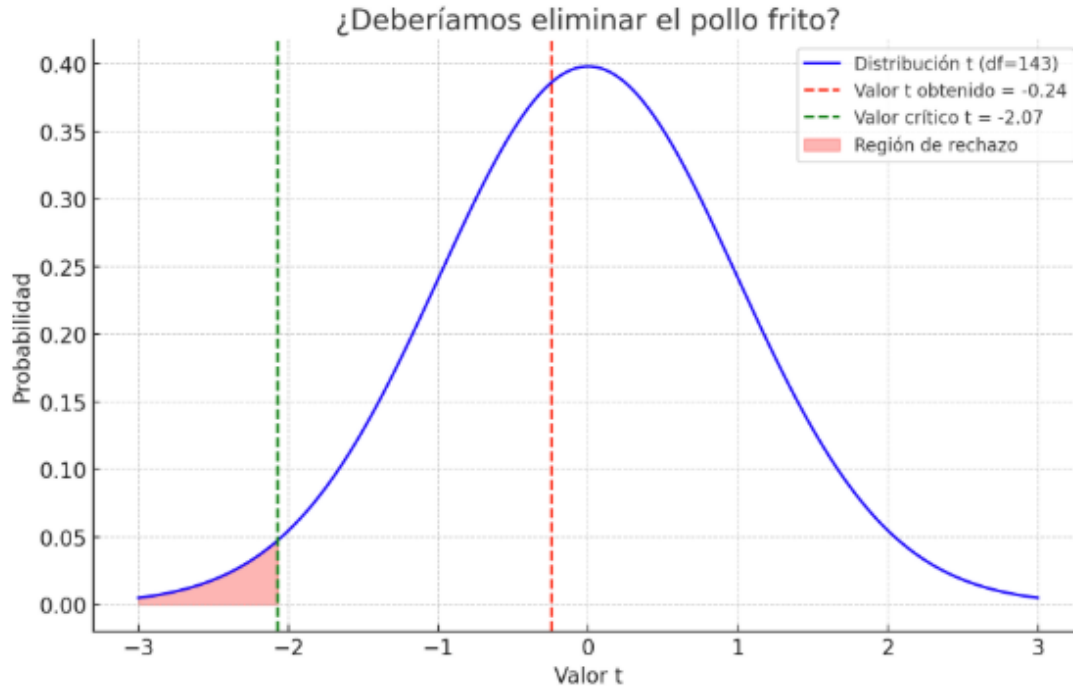
Usaremos un nivel de confianza del 98%, lo que significa que queremos estar 98% seguros de que la decisión que tomemos es correcta.

#### 1.1.2 Ayuda visual para interpretar los resultados

**La línea roja** Es el resultado de nuestro cálculo.

**La zona roja sombreada** Es la parte donde, si cae nuestra línea roja, significa que las ventas sí han bajado significativamente.

**Si la línea roja está dentro de la zona roja, eliminamos el pollo frito. Si no, lo dejamos en el menú.**



Dado que nuestra línea roja no está en la zona roja, quiere decir que no existe una diferencia considerable, esto nos permite concluir que lo mejor es **MANTENER EL POLLO FRITO EN EL MENÚ**

## 1.2 Problema 2

*Un estudio indicó que el 64% de los consumidores de supermercado creen en las marcas propias.*

*El fabricante de una salsa de tomate preguntó a 100 compradores donde 52 prefieren marca propia*

*Probar si el porcentaje de preferencias es menor al 64%, para un 4% de nivel de significancia.*

### 1.2.1 ¿Cómo podemos comprobarlo?

Para decidir si esta diferencia es importante o no, usamos una técnica llamada prueba de hipótesis.

La idea es que asumimos que el estudio original sigue siendo correcto (64% de preferencia) y tratamos de ver si nuestro nuevo resultado (52%) es suficientemente diferente como para dudar de eso.

### 1.2.2 ¿Cómo tomamos la decisión?

Usamos una herramienta matemática llamada prueba Z, que nos ayuda a medir si la diferencia de 12% es grande o solo una fluctuación normal.

- Si la diferencia entre 52% y 64% es pequeña, diremos que es normal.
- Si es muy grande, entonces concluiremos que la preferencia realmente ha bajado.

### 1.2.3 Decision y nivel de confianza

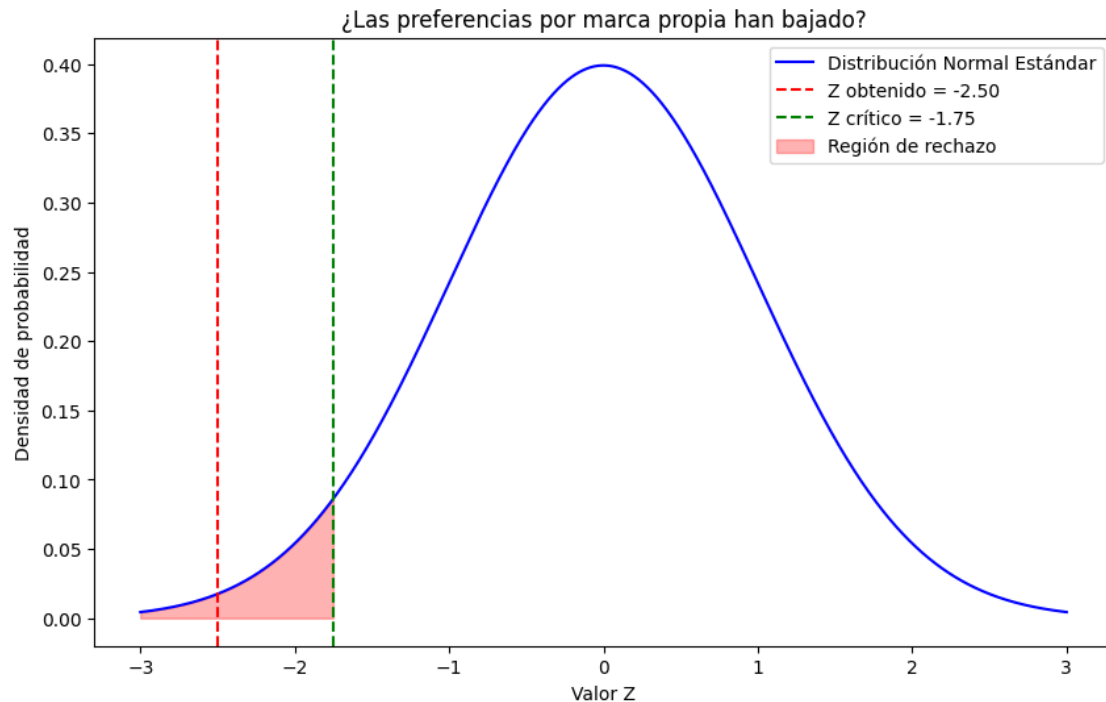
Dado el nivel de significancia de 4%, nosotros queremos estar el 96% seguros antes de decir que la preferencia ha cambiado.

### 1.2.4 Grafica para ilustrar mejor los resultados

- El centro representa la proporción esperada del 64% (o sea, lo que se creía antes).
- Nuestro resultado (52%) se representa con una línea roja en la gráfica.
- La región de rechazo está sombreada en rojo a la izquierda.

Si la línea roja cae **dentro de la zona roja**, significa que la diferencia es demasiado grande como para ser solo una casualidad, y podemos decir que las preferencias realmente han bajado.

Si la línea roja queda **fuera de la zona roja**, entonces la diferencia no es suficiente para concluir que las preferencias cambiaron, y seguimos creyendo en el 64%.



### 1.2.5 Conclusión

Dado que la línea roja está dentro de la zona roja, entonces podemos concluir que la diferencia es demasiado grande para ser solo una casualidad y podemos concluir que **LAS PREFERENCIAS POR MARCA PROPIA REALMENTE HAN BAJADO**

### 1.3 Problema 3

*Las ganancias por acción son de más de 3.5 dolares, para probar esta afirmacion se tomo una muestra de datos como sigue*

1.92 2.16 3.63 3.16 4.02 3.14 2.2 2.34 3.05 2.38

*¿Podriamos decir que las ganancias en promedio supere 3.5 dólares? Calcule es IC a 95% de confianza para el problema. ¿Que podria concluir de todo su analisis?*

Con la muestra de datos que se nos ha proporcionado, vamos a investigar si las ganancias promedio por acción son mayores a \$3.5

**¿Esta muestra nos da suficiente evidencia para decir que, en promedio, las ganancias superan los \$3.5?**

Para poder resolver esta pregunta, vamos a utilizar una herramienta matemática llamada **prueba de hipótesis**. Esta nos dice lo siguiente:

- Partimos de la suposición de que las ganancias promedio son \$3.5 o menos.
- Comparamos la muestra con ese valor y vemos si la diferencia es grande o solo una casualidad.
- Si la diferencia es lo suficientemente grande, concluimos que las ganancias sí son mayores.

Con los datos que tenemos vamos a tratar de demostrar uno de los dos posibles resultados

- Las ganancias son \$3.5 o menos
- Las ganancias son mayores a \$3.5

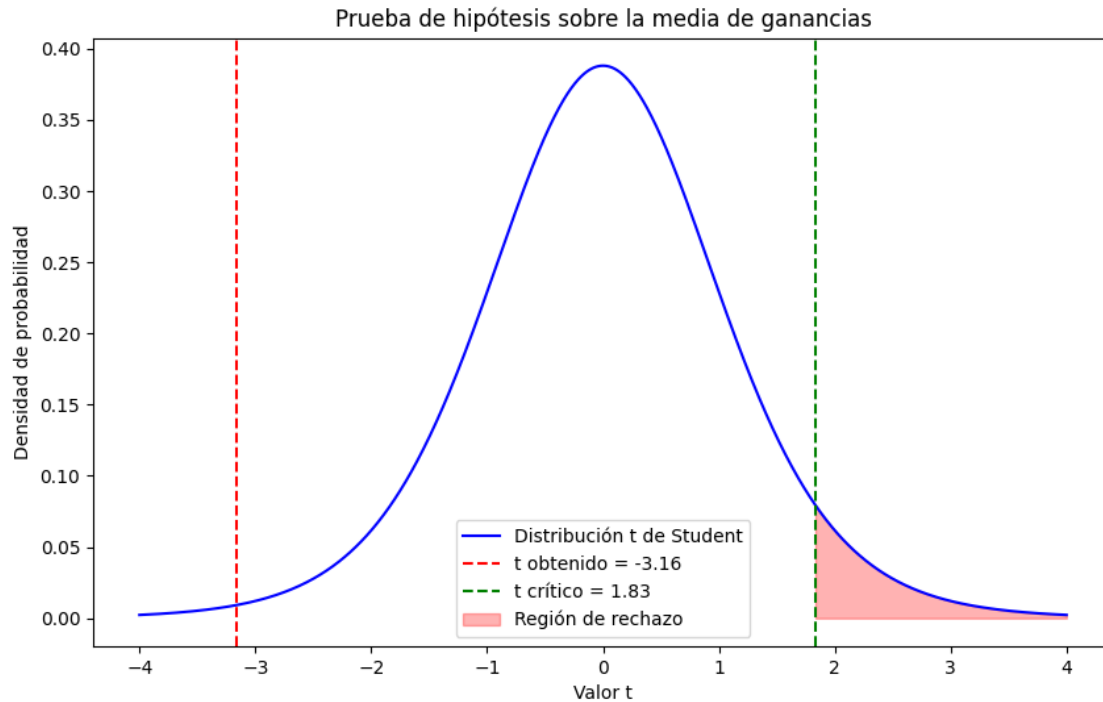
A través de una prueba llamada t de Student, vamos a comparar el promedio de nuestra muestra con un valor teórico (\$3.5). Y también vamos a trabajar con un 95% de confianza.

Una vez que hayamos realizado nuestra prueba, debemos verificar nuestro resultado con una de estas dos situaciones:

- Si el resultado obtenido es mayor que el valor crítico, entonces **hay suficiente evidencia para decir que las ganancias promedio sí son mayores a \$3.5.**
- Si el resultado obtenido no es mayor que el valor crítico, entonces **no podemos asegurarlo.** Puede que la diferencia se deba al azar.

#### 1.3.1 Grafica para poder observar los resultados

- Trazamos la distribución t de Student, que es una campana (Color Azul).
- Marcamos el resultado obtenido de nuestra muestra calculada (Linea roja).
- Marcamos el valor crítico (Linea verde).
- Sombreamos la región de rechazo: **Si el resultado cae dentro de esta región, concluimos que las ganancias son mayores a \$3.5.**



Dado que nuestro resultado es mucho menor que el valor crítico, **NO PODEMOS CONCLUIR QUE LAS GANANCIAS PROMEDIO SUPEREN LOS 3.5 DÓLARES.**

## 1.4 Problema 4

*Armando Romero considera postularse para presidente municipal de Huajuapán de León, Oaxaca.*

*Antes de solicitar la postulación, decide realizar una encuesta entre los electores de Berumen.*

*Una muestra de 400 electores revela que 300 apoyarían en la elecciones de julio de 2024.*

*Romero decide que, si más del 78% de los votantes, se estima lo apoyarán; él se intentaría ser presidente municipal.*

*Además, construya un IC a 99% para su parámetro de estudio. ¿Que le diría usted?*

### 1.4.1 ¿Debería Armando Romero postularse para presidente municipal?

Para resolver este problema vamos a partir de 2 hipótesis o situaciones. Estas nos ayudarán a tomar una decisión con los cálculos que analicemos:

- El apoyo es 78% o menos.
- El apoyo es mayor al 78%.

**SI TENEMOS SUFICIENTE EVIDENCIA DE QUE SUPERA EL 78%, ENTONCES ROMERO DEBERÍA POSTULARSE DE LO CONTRARIO, NO DEBERÍA HACERLO.**

Un dato muy importante que nos ayudará a tomar una decisión es la **Proporción muestral**. pero ¿Qué es?: Es el porcentaje de personas en la muestra que apoyan a Armando Romero.

En este caso: De 400 electores encuestados, 300 dijeron que lo apoyarían. Entonces, la proporción muestral es:

$$300 / 400 = 0.75$$

En la muestra, el 75% de los electores apoyan a Romero.

Nosotros necesitamos saber si la diferencia que existe entre 78% y 75% es importante, para esto usamos una prueba de hipótesis con una distribución normal. Calculamos un valor llamado **estadístico z**, que nos dice qué tan lejos está el 75% del 78%.

Si este valor z es muy pequeño (es decir, si el 75% y el 78% no son muy distintos), no podremos decir con seguridad que Romero tiene más apoyo.

Usamos un nivel de confianza del 99% Esto significa que solo aceptaremos que más del 78% lo apoyan si tenemos una evidencia muy fuerte (99% de confianza).

El valor crítico es un número que nos ayuda a decidir si el resultado es significativo. Para un nivel de confianza del 99%, el **valor crítico es 2.3263**. Esto significa que, si el valor de Z es mayor que 2.3263, podríamos concluir que más del 78% de los votantes lo apoyarían.

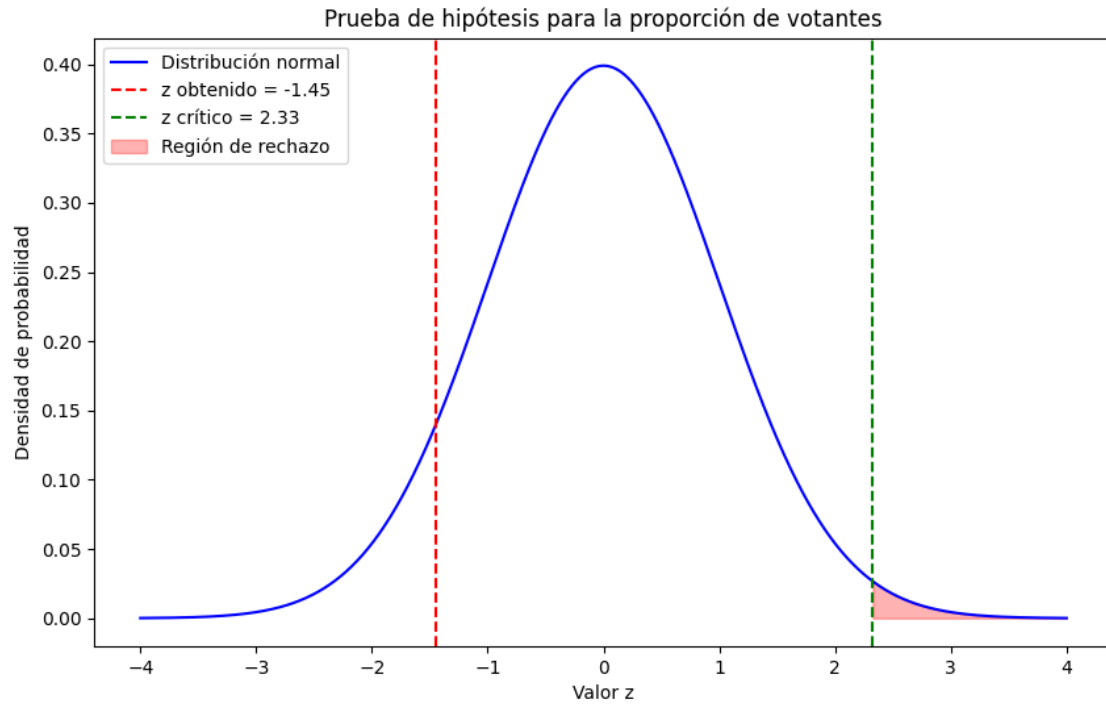
Con esta información vamos a definir un rango de valores para saber donde esté la verdadera proporción de votantes que apoyarían a Romero. Si contamos con un 99% de confianza, la proporción de votantes que apoyarían a Romero está entre 69.42% y 80.58%.

#### **1.4.2 Gráfica para interpretar los resultados**

Una campana (distribución normal) que representa las posibles variaciones en la proporción de votantes.

Un área roja (región crítica) que indica los valores donde la proporción sería significativamente mayor.

Una línea azul punteada que representa nuestro resultado. Como está fuera del área roja, no podemos concluir que más del 78% de los votantes lo apoyarían.



Por lo tanto, **NO PODEMOS CONCLUIR QUE MÁS DEL 78% DE LOS VOTANTES APOYARÍAN A ROMERO.**

Si bien no hay evidencia suficiente para afirmar que más del 78% lo apoyarían, el intervalo sugiere que aún tiene un apoyo significativo. Sería recomendable que Romero realice más encuestas o análisis antes de tomar una decisión final.