

Giả thuyết cần kiểm định:

Theo một cuộc thăm dò do Mashed thực hiện, chỉ khoảng một phần ba số người cho rằng Pepsi ngon hơn Coke. Trong số 36.000 người được hỏi trên toàn thế giới, 64% cho biết họ thích Coca hơn, trong khi chỉ có 36% cho biết họ thích Pepsi hơn.

Nghi ngờ rằng số liệu khảo sát có thể khác so với thực tế rằng số người yêu thích Coca hơn Pepsi ở dưới mức 64%. Tôi đã tiến hành kiểm định và hỏi ngẫu nhiên 50 người khi đi qua máy bán hàng tự động ở căng tin nơi làm việc rằng họ thích Coca hay Pepsi hơn sau đó ghi lại kết quả.

*Sử dụng mức ý nghĩa là 0.05

Nguồn thăm dò:

<https://www.mashed.com/588280/only-about-1-in-3-people-actually-prefer-pepsi-to-coke/>

- Ở đây dữ liệu được lấy từ mẫu gồm 50 người ngẫu nhiên và quần thể là tất cả những người yêu thích đồ uống có ga.

- Phân loại dữ liệu

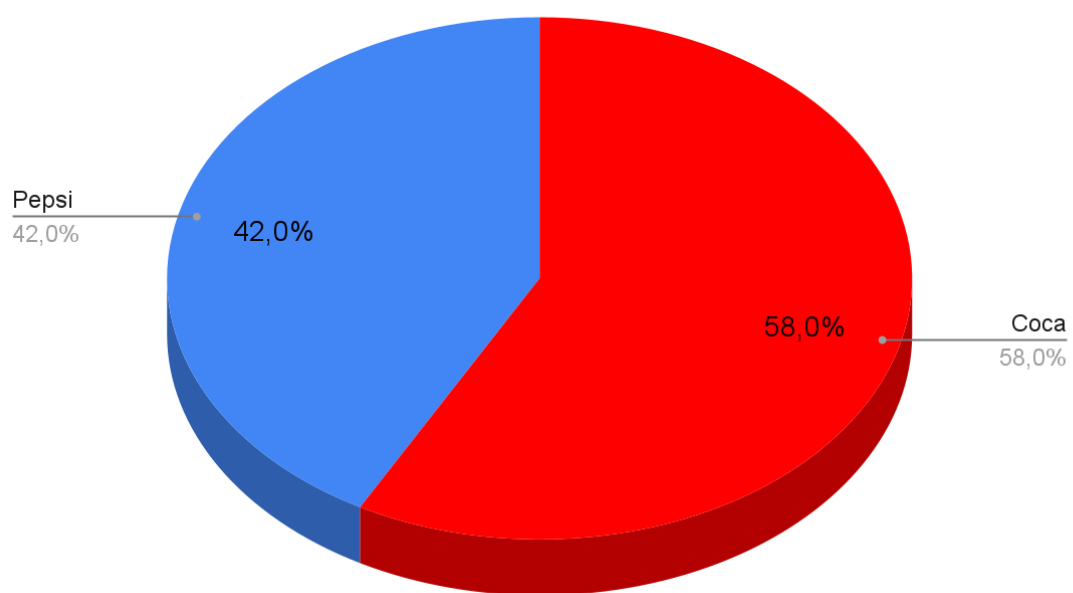
Dữ liệu được phân loại dữ liệu định danh (nominal data) vì chúng không được sắp xếp theo thứ tự

- Cách lấy dữ liệu: Lựa chọn ngẫu nhiên mẫu gồm 50 người bằng cách hỏi ngẫu nhiên 50 người đi ngang qua máy bán hàng tự động ở căng tin nơi làm việc
- Tiến hành thu thập và lấy được dữ liệu: ghi lại kết quả của 50 người ngẫu nhiên sau đó tổng hợp lại kết quả thu được qua bảng hai chiều (mẫu này được phân tầng theo giới tính)

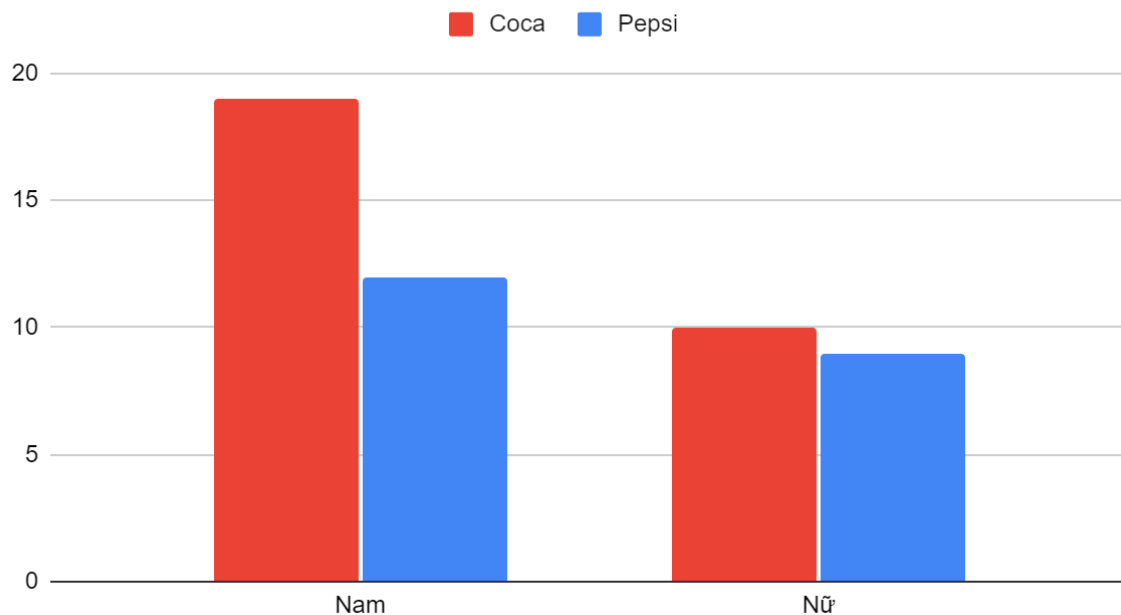
	Coca	Pepsi	Tổng cộng
Nam	19	12	31
Nữ	10	9	19
Tổng cộng	29	21	50

- Đồ thị mô tả dữ liệu:

So sánh sự yêu thích giữa Coca và Pepsi



Sự yêu thích Coca và Pepsi giữa nam và nữ



- Các đặc trưng cần thiết cho bài toán:

- $n = 50, \pi_0 = 0.64$

- $p = 0.58$

- Kiểm tra điều kiện:

- Mẫu là ngẫu nhiên đơn

- Điều kiện phân phối chuẩn: $np \geq 15, n(1 - p) \geq 15$

$n = 50$

$n \cdot p = 50 \cdot 0.58 = 29 > 15$

$n \cdot (1 - p) = 50 \cdot 0.42 = 21 > 15$

→ Phân phối đạt chuẩn

- Xây dựng cặp giả thuyết gốc và giả thuyết đối

Giả thuyết cho rằng số lượng người yêu thích Coca hơn Pepsi sẽ nhỏ hơn 64%, trong khi nghiên cứu chỉ ra rằng còn số này là 64%. Ta có cặp giả thuyết gốc và giả thuyết thay thế sau:

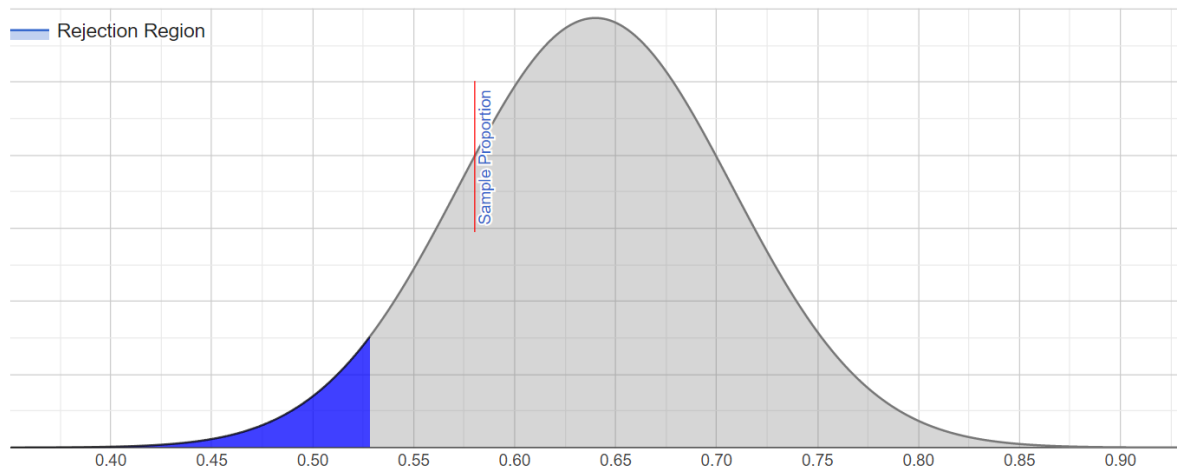
$H_0: \pi = 0.64$

$H_a: \pi < 0.64$

- Tính toán:

$$z = \frac{p - \pi_0}{\sqrt{\frac{\pi_0(1 - \pi_0)}{n}}} = \frac{0.58 - 0.64}{\sqrt{\frac{0.64(1 - 0.64)}{50}}} = -0.88$$

$\alpha = 0.05 \rightarrow$ giá trị tới hạn là $-1.64 \rightarrow$ Miền bác bỏ $z < -1.64$



Vì giá trị z của mẫu nằm ngoài vùng bác bỏ ($z > -1.64$) nên chúng ta không thể bác bỏ giả thuyết gốc H_0

- Kết luận:

Thông qua việc kiểm định xác suất của mẫu cho được kết quả $z = -0.88$, với mức ý nghĩa là 95% ($\alpha = 0.05$) vùng bác bỏ được xác định nhỏ hơn mức tới hạn là -1.64 ($z < -1.64$) vì vậy ta kết luận dữ liệu của mẫu không đủ bằng chứng để có thể phủ định giả thuyết gốc là số người thích Coca hơn Pepsi là 64%, kết quả này có ý nghĩa một cách thống kê.

Do đó, giả thuyết ban đầu được đưa ra là số người thích Coca hơn Pepsi sẽ dưới mức 64% đã không được chấp nhận vì đây là giả thuyết đưa ra đối thuyết với giả thuyết gốc.

