



# COLE.VN

Learn Remotely, Work Globally

# TÊN KHÓA HỌC:

Chủ đề:

BÀI TẬP THỰC HÀNH BUỔI 23

TS. Đặng Lê Quang

# BÀI TẬP THỰC HÀNH

## Bộ dữ liệu thực hành: [US Import](#)

Bộ dữ liệu **US Import Cleaned** chứa thông tin chi tiết về **các lô hàng nhập khẩu vào Hoa Kỳ**, bao gồm **xuất xứ, nhà xuất khẩu, nhà nhập khẩu, mã HS, số lượng, trọng lượng, giá trị giao dịch, và phương thức vận chuyển**. Bộ dữ liệu này đặc biệt hữu ích để **phân tích xu hướng nhập khẩu, nhận diện các nhà cung cấp chính, đánh giá chi phí và quản lý chuỗi cung ứng**.

Bộ dữ liệu này thường được sử dụng để thực hành các **kỹ thuật phân tích dữ liệu, máy học và thống kê**.

# BÀI TẬP THỰC HÀNH

Bộ dữ liệu có các cột quan trọng sau:

- **Date (Ngày nhập khẩu)** – Thời điểm giao dịch nhập khẩu được ghi nhận.
- **Import Country (Source)** – Xác nhận rằng tất cả dữ liệu đều liên quan đến nhập khẩu vào Hoa Kỳ.
- **Exporter (Nhà xuất khẩu)** – Doanh nghiệp hoặc tổ chức cung cấp hàng hóa.
- **Importer (Nhà nhập khẩu)** – Công ty hoặc cá nhân nhận hàng tại Hoa Kỳ.
- **HS Code (Mã HS)** – Mã hệ thống hải hòa mô tả loại hàng hóa nhập khẩu.
- **Country of Origin (Quốc gia xuất xứ)** – Quốc gia sản xuất hàng hóa.
- **Country of Destination (Quốc gia đích đến)** – Luôn là Hoa Kỳ.
- **Quantity (Số lượng) & Qty Unit (Đơn vị số lượng)** – Tổng số mặt hàng trong mỗi lô hàng.
- **Weight (Trọng lượng) & Wgt Unit (Đơn vị trọng lượng)** – Trọng lượng của lô hàng.
- **Value (US\$) (Giá trị hàng hóa)** – Giá trị nhập khẩu của lô hàng tính bằng USD.
- **Transport Mode (Phương thức vận chuyển)** – Cách thức vận chuyển hàng hóa
- **Products (Loại sản phẩm)** – Tên loại hàng hóa được nhập khẩu.

# BÀI TẬP THỰC HÀNH

## Bài 1: Lấy mẫu dữ liệu nhập khẩu

- a. Dùng hàm describe để thực hiện thống kê mô tả cho toàn bộ dữ liệu
- b. Lấy mẫu 30 giao dịch nhập khẩu bằng phương pháp **Simple Random Sampling** (*Lấy mẫu ngẫu nhiên đơn giản*)
- c. Lấy mẫu 30 giao dịch nhập khẩu bằng phương pháp **Stratified Sampling** (*Lấy mẫu phân tầng*)
- d. Lấy mẫu 30 giao dịch nhập khẩu bằng phương pháp **Systematic Sampling** (*Lấy mẫu hệ thống*)
- e. Lấy mẫu 30 giao dịch nhập khẩu bằng phương pháp **Cluster Sampling** (*Lấy mẫu cụm*)
- f. Phương pháp lấy mẫu nào cho kết quả trung bình chính xác nhất?
- g. Hãy rút ra kết luận khi nào nên sử dụng các phương pháp lấy mẫu khác nhau

# BÀI TẬP THỰC HÀNH

**Bài 2: Tính khoảng tin cậy:** Tính khoảng tin cậy 95% cho giá trị trung bình của biến Value (US\$) cho từng cách lấy mẫu ở bài 1.

**Bài 3: Ước lượng tỷ lệ tổng thể:** Ước lượng tỷ lệ giao dịch nhập khẩu từ Trung Quốc (China) tại mức độ tin cậy 95% cho từng phương pháp lấy mẫu ở bài 1.

**Bài 4: Kiểm định t-test:** Kiểm định t-test để kiểm tra xem trung bình giá trị nhập khẩu (Value US\$) của phương pháp lấy mẫu Stratified Sampling có lớn hơn 50,000 USD hay không với mức ý nghĩa  $\alpha = 0.05$ . Diễn giải kết quả và rút ra kết luận từ kiểm định.

**Thank you  
for  
listening!**

