|  |  |  |
| --- | --- | --- |
| **Nhóm 10** | | |
| **STT** | **MSSV** | **HỌ VÀ TÊN** |
| 46 | 21057901 | Nguyễn Đa Nghiêm |
| 47 | 21000615 | Lê Hải Ngọc |
| 48 | 20000841 | Nguyễn Thái Ngọc |
| 49 | 21070711 | Lê Phước Nguyên |
| 50 | 21068021 | Phan Nguyễn Khôi Nguyên |

**Mẫu Ngẫu Nhiên Đơn Giản (Simple Random Sampling):**

Ưu điểm:

Dễ thực hiện và hiểu.

Đảm bảo tính ngẫu nhiên và khả năng biểu diễn của mẫu.

Nhược điểm:

Không đảm bảo sự đại diện của mẫu đối với từng phân khúc trong tổng thể.

**Mẫu Có Hệ Thống (Systematic Sampling):**

Ưu điểm:

Dễ thực hiện và phù hợp khi có sự sắp xếp đều đặn trong tổng thể.

Nhược điểm:

Có thể tạo ra mẫu có đặc điểm đặc biệt nếu có mô hình trong dãy số được sắp xếp.

**Mẫu Phân Tầng (Stratified Sampling):**

Ưu điểm:

Đảm bảo đại diện của mẫu đối với từng phân khúc.

Giảm sai số mẫu so với ngẫu nhiên đơn.

Nhược điểm:

Yêu cầu thông tin về tỷ lệ phân phối trong từng phân khúc.

**Lấy Mẫu Theo Cụm (Cluster Sampling):**

Ưu điểm:

Dễ thực hiện và giảm chi phí so với ngẫu nhiên đơn.

Nhược điểm:

Cụm có thể không đại diện nếu sự đồng nhất không đảm bảo.

**EX**

**Lấy Mẫu Ngẫu Nhiên:**

coffee\_ratings.sample(n=5, random\_state=19000113)

**Lấy Mẫu Có Hệ Thống:**

sample\_size = 5

interval = pop\_size // sample\_size

coffee\_ratings.iloc[::interval]

**Lấy Mẫu Phân Tầng:**

coffee\_ratings\_top = coffee\_ratings[coffee\_ratings['Country of Origin'].isin(top\_counted\_countries)]

coffee\_ratings\_samp = coffee\_ratings\_top.sample(frac=0.1, random\_state=2021)

**Lấy Mẫu Theo Cụm:**

top\_countries\_samp = random.sample(top\_counted\_countries, k=2)

coffee\_ratings\_cluster = coffee\_ratings\_top[coffee\_ratings\_top['Country of Origin'].isin(top\_countries\_samp)]

coffee\_ratings\_cluster.groupby('Country of Origin').sample(n=len(coffee\_ratings\_top) // 6)