

CÁC PHƯƠNG PHÁP ĐỊNH LƯỢNG 1**BÀI TẬP 1****THỐNG KÊ MÔ TẢ VÀ XÁC SUẤT**

Ngày phát: 25/10/2022

Hạn nộp: 8h20, 08/11/2022

Bài làm được yêu cầu chỉ nộp bản điện tử trên Microsoft Teams

Tập tin *PCI2021.csv* chứa dữ liệu Chỉ số Năng lực Cạnh tranh cấp Tỉnh (PCI) của 63 tỉnh thành được công bố vào năm 2021.

Các cột dữ liệu được mô tả ngắn gọn trong Bảng 1. Thông tin chi tiết về các chỉ số thành phần cũng như phương pháp tính toán tổng điểm PCI được trình bày tại trang web: pcivietnam.vn

Tên cột	Mô tả	Trọng số
Province	Tên tỉnh/thành	N/A
Entry_cost	Chỉ số thành phần - Chi phí gia nhập thị trường thấp	0.05
Access_to_land	Chỉ số thành phần - Tiếp cận đất đai dễ dàng	0.05
Transparency	Chỉ số thành phần - Môi trường kinh doanh minh bạch và thông tin kinh doanh công khai	0.20
Time_costs	Chỉ số thành phần - Thời gian thanh tra, kiểm tra và thực hiện các quy định thủ tục hành nhanh chóng	0.05
Informal_charges	Chỉ số thành phần - Chi phí không chính thức thấp	0.10
Business_environment	Chỉ số thành phần - Môi trường cạnh tranh bình đẳng	0.05
Proactivity	Chỉ số thành phần - Chính quyền địa phương năng động sáng tạo trong giải quyết vấn đề cho doanh nghiệp	0.05
Business_support	Chỉ số thành phần - Các chính sách hỗ trợ doanh nghiệp hiệu quả	0.20
Labour_policy	Chỉ số thành phần - Chính sách đào tạo lao động tốt	0.20
Law_and_order	Chỉ số thành phần - Thủ tục giải quyết tranh chấp công bằng, hiệu quả và an ninh trật tự được duy trì	0.05
PCI_score	Điểm PCI - Tổng có trọng số các chỉ số thành phần	N/A
Ranking	Xếp hạng PCI (theo điểm PCI)	N/A
Tier	Xếp loại PCI (dựa trên độ lệch chuẩn của điểm PCI)	N/A

Anh/Chị sử dụng bộ dữ liệu này và các thông tin liên quan để đánh giá năng lực cạnh tranh của các tỉnh/thành theo các yêu cầu cụ thể dưới đây:

Câu 1. Anh/Chị hãy phân loại các biến trong bộ dữ liệu theo loại, bao gồm biến định danh (nominal), biến thứ tự (ordinal), biến khoảng (interval), và biến tỷ lệ (ratio).

Câu 2. Anh/Chị hãy tính các trị thống kê mô tả cơ bản của từng chỉ số thành phần của 63 tỉnh/thành và bình luận ngắn gọn về các trị thống kê này.

Câu 3. Anh/Chị vẽ đồ thị tần suất theo khoảng giá trị (histogram) và biểu đồ hộp (boxplot) của từng chỉ số thành phần của 63 tỉnh/thành. Anh/Chị có kết luận gì về phân phối xác suất của các chỉ số thành phần? Các kết luận này có phù hợp với các kết luận thu được từ các trị thống kê được tính ở Câu 2 không?

Câu 4. Lập bảng giá trị phân vị theo các mức 5%, 10%, 25%, 50%, 75%, 90%, 95% của từng chỉ tiêu thành phần của 63 tỉnh/thành. Dựa trên bảng giá trị phân vị, anh/chị hãy đưa ra một số nhận định về thành tích của các tỉnh/thành theo từng chỉ tiêu thành phần.

Câu 5. Giả sử các tỉnh/thành được chia thành 3 nhóm. Nhóm 1 bao gồm các tỉnh/thành được xếp hạng ‘Tốt’ và ‘Rất tốt’. Nhóm 2 bao gồm các tỉnh/thành được xếp hạng ‘Khá’. Nhóm 3 bao gồm các tỉnh/thành được xếp hạng ‘Trung bình’, ‘Tương đối thấp’, và ‘Thấp’.

Đối với mỗi nhóm, anh/chị hãy vẽ đồ thị tần suất theo khoảng giá trị (histogram) và biểu đồ hộp (boxplot) của từng chỉ số thành phần và của tổng điểm có trọng số (cột ‘PCI_score’).

Dựa trên các đồ thị cho mỗi nhóm như trên, anh/chị so sánh và đưa ra nhận định ngắn gọn về phân phối điểm của ‘PCI_score’ và của từng chỉ số thành phần giữa 3 nhóm.

Câu 6. Anh/Chị tính hệ số tương quan giữa các cặp chỉ số thành phần của 63 tỉnh/thành, kiểm chứng bằng các đồ thị scatter plot tương ứng, và nhận định ngắn gọn về các kết quả.

Câu 7. Anh/Chị lập bảng tần suất 2 chiều phù hợp để trả lời các câu hỏi sau:

- Xác suất để chọn 1 tỉnh thuộc vùng Đồng Bằng Sông Cửu Long được xếp hạng PCI loại ‘Khá’ hoặc cao hơn.
- Xác suất để chọn 1 tỉnh thuộc vùng Bắc Trung Bộ và Duyên Hải Miền Trung, vùng Tây Nguyên và vùng Đông Nam Bộ được xếp hạng PCI loại ‘Trung bình’.
- Xác suất để chọn 1 tỉnh thuộc vùng Đồng Bằng Sông Hồng và vùng Trung Du và Miền Núi Phía Bắc được xếp hạng PCI loại ‘Tốt’.

Danh sách các tỉnh/thành nằm trong các vùng kinh tế xã hội trong Câu 7 được định nghĩa bởi Tổng cục Thống kê tại đường link <https://tinyurl.com/2x9rpehu>.

---HẾT---