

**Trường Chính sách công và Quản lý Fulbright**

Học kỳ Xuân năm 2020

**THỐNG KÊ VÀ PHÂN TÍCH DỮ LIỆU**

**BÀI TẬP 1**

**Ngày Phát: Thứ bảy 07/03/2020**

**Ngày Nộp: Thứ hai 09/03/2020**

Học viên nộp bản điện tử bao gồm Word và Excel trước 13h20 trên ứng dụng Microsoft Teams

Học viên cần tính toán trên Excel và trình bày rõ ràng trên Word

**Câu 1:**

Bạn hãy sử dụng file dữ liệu NGTK.xlsx về các dữ liệu thống kê kinh tế xã hội của Việt Nam được lấy từ GSO (<https://www.gso.gov.vn/>) để trả lời các câu hỏi sau:

	Đơn vị	Diễn giải
<b>province</b>		Tỉnh
<b>region</b>		Vùng
<b>acreage</b>	km <sup>2</sup>	Diện tích
<b>population</b>	ngàn người	Dân số trung bình
<b>density</b>	người/km <sup>2</sup>	Mật độ dân số/km <sup>2</sup>
<b>labor</b>	ngàn người	Lực lượng lao động từ 15 tuổi trở lên
<b>scale</b>		Phân loại mật độ dân số (Thấp: mật độ < 400 người/km <sup>2</sup> Trung Bình: 400 < mật độ < 800 Cao: mật độ ≥ 800 người/km <sup>2</sup> )
<b>migrate</b>	%	Tỷ suất di cư thuần
<b>ind-index</b>	%	Chỉ số sản xuất công nghiệp
<b>s-market</b>	siêu thị	Số lượng siêu thị
<b>project</b>	dự án	Số dự án FDI
<b>capital</b>	triệu USD	Tổng vốn FDI đăng ký

- Phân loại loại biến và thang đo cho 5 biến province, region, population, scale, ind-index?
- Vẽ đồ thị tần số về dân số phân theo vùng?

- c. Dùng hàm trong excel (không được dùng công cụ Data Analysis) để tính toán điền vào ô trống ở các dòng A, C, F, J, K, M (các ô bôi màu cam)

project		results
A	Mean	?
C	Median	?
E	Standard Deviation	?
F	Sample Variance	?
I	Range	?
J	Minimum	?
K	Maximum	?
L	Sum	?
M	Count	?
N	Coefficient of Variation	?
P	Giá trị tại vị trí phân vị 45th	?

- d. Từ kết quả ở các ô bôi màu cam, diễn giải cách tính cho các ô còn lại E, I, L, N (không được dùng hàm và công cụ Data Analysis)
- e. Tính mode của project nếu cho bin bằng 700?
- f. Trình bày cách làm để có được số vốn FDI đăng ký ứng với mỗi mức mật độ dân số (scale) ở mỗi vùng?

**Câu 2:** Vẫn sử dụng bộ dữ liệu NGTK.xlsx để trả lời những câu hỏi sau:

- a. Dùng Data Analyst để tính các giá trị thống kê mô tả cho toàn bộ các biến trong bộ dữ liệu (ngoại trừ province, region và scale)
- b. Kiểm tra độ lệch xiên của những biến vừa thống kê mô tả ở câu 2a bằng dấu của skewness?
- c. Vẽ đồ thị phân tán biểu diễn mối quan hệ giữa vốn FDI đăng ký với từng biến trong 7 biến acreage, population, density, labor, migrate, ind-index, s-market. Từ đồ thị anh/chị hãy nhận xét về mức độ tương quan giữa vốn FDI đăng ký với 7 biến này.
- d. Kiểm tra sự tương quan giữa vốn FDI đăng ký với từng biến trong 7 biến trên bằng hệ số tương quan. So sánh với nhận xét ở câu c.