

Trường Chính sách công và Quản lý Fulbright

Học kỳ xuân năm 2022

THỐNG KÊ VÀ PHÂN TÍCH DỮ LIỆU

BÀI TẬP 1

Ngày Phát: Thứ bảy 12/02/2022

Ngày Nộp: Thứ hai 14/02/2022

Học viên nộp bản điện tử bao gồm Word và Excel trước 08h20 trên ứng dụng Microsoft Teams

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

Việt Nam thực hiện chuyển đổi nền kinh tế kể từ năm 1986, từ kinh tế kế hoạch hóa tập trung sang kinh tế thị trường định hướng xã hội chủ nghĩa, với nhiều đặc tính khác biệt với các thuyết hệ luận phương Tây (Malesky & London, 2014) và nhà nước kiểm soát Trung Quốc (Pincus, Anh, Nghia, Wilkinson, & Thanh, 2012). Theo đó, sự phát triển đất nước không chỉ đảm bảo các khía cạnh kinh tế mà gắn chặt chẽ với cải thiện các vấn đề xã hội.

Theo hướng nghiên cứu này, Van Le, Tran, and Doan (2022) gần đây chứng tỏ rằng sự phát triển kinh tế tư nhân có đóng góp quan trọng tới giảm nghèo đa chiều tại Việt Nam. Theo đó, với mỗi % gia tăng tỷ lệ doanh nghiệp tư nhân sẽ cải thiện khoảng 0.31% tỷ lệ nghèo đa chiều tại Việt Nam (đọc thêm chi tiết file đính kèm). Trích xuất một phần bộ dữ liệu của nghiên cứu trên, dữ liệu trong bài tập này được sử dụng là dữ liệu bảng thu thập tại 63 tỉnh/thành Việt Nam trong giai đoạn 2016-2019, với các biến được mô tả tại bảng dưới.

Bảng 1. Tên biến, đơn vị và mô tả.

Tên biến	Đơn vị	Mô tả
Province		Tỉnh mã hóa dạng số
year	Năm	Giai đoạn 2016-2019
region	Biển giả	1 = Trung du và miền núi bắc bộ / 2 = Bắc Trung Bộ/ 3= Đồng bằng sông Hồng/ 4= Duyên hải nam trung bộ/ 5 = Tây nguyên/ 6 = Đông Nam Bộ/ 7 = Đồng bằng sông Cửu Long
grdp	Nghìn tỷ đồng	Tổng sản lượng cấp tỉnh
labour	Nghìn người	Tổng số lao động trên 15 tuổi
capital	Triệu USD	Đầu tư trực tiếp nước ngoài
fdi_imp	Triệu đồng	Được biết đến từ investment qua công thức OECD - 2009
mul_pov	%	Tỷ lệ nghèo đa chiều
p_share	Người	Số lao động trong doanh nghiệp tư nhân
s_share	Người	Số lao động trong doanh nghiệp nhà nước
mne_share	Người	Số lao động trong khu vực MNE
idmap		Tỉnh mã hóa dạng chữ (biểu đồ)

Nguồn: ban giảng viên tổng hợp. Ghi chú: dữ liệu về tiền tệ đã được chuyển đổi về năm 2010

Dựa trên bộ dữ liệu được cung cấp, học viên sẽ thực hiện các bước phân tích dữ liệu bước đầu (phân tích mô tả) thông qua:

1. (Nhận diện) Dựa trên dữ liệu được cung cấp tại file **data_BT1**, hãy phân loại loại biến (*định tính hay định lượng, nếu là định lượng thì là biến rời rạc hay liên tục*) và thang đo tương ứng cho các biến tại **bảng 1**? (10 điểm)
2. Dựa trên dữ liệu được cung cấp (**data_BT1**), Học viên hãy:
 - i) Tạo biến logarit tổng vốn nền kinh tế (hàm “log10” trong excel), gọi là log(capital)
 - ii) Tạo biến tỷ lệ phát triển khu vực tư nhân:

$$\text{Trong đó, tỷ lệ phát triển khu vực tư nhân } \pi = \frac{p_share + mne_share}{p_share + mne_share + s_share} \quad (1)$$

- iii) Vẽ đồ thị tần suất (histogram) của biến tổng sản phẩm theo của tỉnh (grdp), log(capital), và biến tỷ lệ phát triển khu vực tư nhân (π); sau đó, so sánh rồi đưa ra nhận xét (gợi ý: về vị trí của mean, median, tính đối xứng của đồ thị). (15 điểm)
3. Vẽ đồ thị hình tròn mô tả cơ cấu tổng sản phẩm (GDP = cộng dồn grdp) phân theo 7 vùng kinh tế, qua đó, đưa ra nhận xét về đóng góp kinh tế của các khu vực này (ngắn gọn). (10 điểm)

4. Sử dụng các hàm trong excel để tính toán các thông số về quy mô nền kinh tế Việt Nam theo bảng sau: (15 điểm)

Bảng 2. Yêu cầu tính toán

Tổng sản phẩm tỉnh/thành (GRDP)		Kết quả
A	Mean	
B	Median	MEDIAN
C	Sample Standard Deviation	STDEV.S
D	Sample Variance	
E	Range	
F	Minimum	MIN
G	Maximum	MAX
H	Sum	SUM
I	Count	COUNT
J	Giá trị tại vị trí phân vị 30th	PERCENTILE.EXC
L	Độ lệch (Skewness)	SKEW
M	Tốc độ tăng trưởng bình quân 2016-2019 (Cả nước)	

Lưu ý: Các hàm cần sử dụng để tính toán được trình bày tại bảng. Tốc độ tăng trưởng được tính theo trung bình nhân.

5. (Nhận diện xu hướng) Sử dụng dữ liệu **data_BT1**, học viên hoàn tất bảng sau và đánh giá xu hướng tăng trưởng của Việt Nam năm 2019 so với 2016 theo tỷ lệ nghèo đa chiều, tỷ lệ phát triển khu vực tư nhân, và quy mô nền kinh tế cấp tỉnh/thành (GRDP) bình quân. (15 điểm)

Bảng 3. Yêu cầu tính toán.

	2016		2019	
	MEAN	SD	MEAN	SD
Tỷ lệ nghèo đa chiều				
Tỷ lệ phát triển khu vực tư nhân				
Quy mô nền kinh tế (GRDP bình quân)				

6. Dùng Data Analyst (trong excel) để tính các giá trị thống kê mô tả cho các biến tỷ lệ nghèo đa chiều, tỷ lệ phát triển khu vực tư nhân, và GRDP. Hãy nhận xét về ý nghĩa của độ xiên (skewness) tại các biến. (15 điểm)
7. Xem xét tương quan giữa các biến số sau và đưa ra nhận xét về mức độ tương quan. (10 điểm)
 - ✓ Nghèo đa chiều với GRDP
 - ✓ Nghèo đa chiều với private_sect
 - ✓ Nghèo đa chiều với fdi_imp
 - ✓ GRDP với private_sect
8. Vẽ biểu đồ tương quan (scatter plot) giữa biến nghèo đa chiều và private_sect (gợi ý, có thể vẽ phân theo các vùng kinh tế). (10 điểm)

Chúc các anh/chị học viên làm bài tốt!

Tài liệu tham khảo

- Malesky, E., & London, J. (2014). The political economy of development in China and Vietnam. *Annual Review of Political Science*, 17, 395-419.
- Pincus, J., Anh, V. T. T., Nghia, P. D., Wilkinson, B., & Thanh, N. X. (2012). Structural reform for growth, equity and national sovereignty. *VELP, Harvard Kennedy School, February*, 13-17.
- Van Le, D., Tran, T. Q., & Doan, T. (2022). The private sector and multidimensional poverty reduction in Vietnam: A cross-province panel data analysis. *International Journal of Social Welfare*.