

# MAS291 Computer Project

Tình hình sản xuất lúa gạo ở các tỉnh thành theo năm.

Giảng viên:

-Nguyễn Việt Anh

Thành viên:

-Chu Văn Luân

-Trần Hải Nam

-Hà Trung Hiếu

## I. Số liệu

Về phần số liệu, nhóm đã thống kê ở Website <https://www.gso.gov.vn/>.

Đây là số liệu thực về tình hình sản lượng lúa ở các khu vực ( Đồng bằng sông Hồng, Trung du và miền núi phía bắc, Đồng bằng sông Cửu Long,..) và các tỉnh thành ( Hà Nội, Hải Phòng, Hậu Giang, Tiền Giang,...) vào các năm từ năm 1995 đến 2020. Ta gọi số liệu này là Population.

(Tham khảo số liệu ở File DuLieu.xlsx ở Sheet “Population”).

Tại đây, ta lấy ra sản lượng lúa của 6 tỉnh Bắc Trung Bộ (gồm có Thanh Hóa, Nghệ An, Hà Tĩnh, Quảng Bình, Quảng Trị, Huế) tại 15 năm (từ 1995 đến 2010) làm Sample để quan sát và đưa ra đánh giá.

(Tham khảo số liệu ở File DuLieu.xlsx ở Sheet “Sample”).

## II. Đánh giá

## 1. Khoảng tin cậy (CI)

Ta sẽ tính khoảng tin cậy (CI) của số liệu tại độ tin cậy 95%.

Để làm ví dụ cụ thể, ta giả sử sản lượng lúa của Quảng Trị theo các năm từ 1995 đến 2010 là sản lượng trung bình cả nước với độ tin cậy là 95%.

Ta xây dựng khoảng tin cậy (CI) theo công thức

$$CI = (\bar{X} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} ; \bar{X} + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}})$$

Trong đó  $\sigma$  là POPULATION DEVIATION,  $\bar{X}$  là SAMPLE MEAN.  $n$  là số năm trong SAMPLE.

Bằng các công thức

$$\bar{X} = \frac{\sum_{i=1}^n x_i}{n}$$

$$\mu = \frac{\sum_{i=1}^N x_i}{N}$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - \mu)^2}{N}}$$

Hoặc tra cứu ở File DuLieu.xlsx tại các Sheet "Population" và "Sample".

Ta có được  $\bar{X} = 324,9069, \sigma = 64,69715148, n = 16$

$$\alpha = 0,025 \Rightarrow Z_{\alpha/2} = 1,96 \text{ (bảng Z - Table)}$$

Thay số vào, tính được  $CI = (293,2 ; 356,6)$  hay  $324,9069 \pm 31,7$

Từ con số này, chúng ta có thể tin rằng các giá trị của sản lượng lúa của tỉnh Quảng Trị nằm trong khoảng (293,2 ; 356,6) với độ tin cậy tới 95%. Nghĩa là, sản lượng lúa bình quân của Quảng Trị thay đổi không quá rõ rệt, và cũng không quá cao nếu so sánh với 2 vựa lúa lớn nhất cả nước (thường vượt quá 1000).

## 2. REGRESSION

### Giải thích:

Ta thấy, năng suất lúa tại Quảng Trị hằng năm luôn tăng. Ta sẽ thử kiểm chứng là có phải là tuyến tính hay không.

Nhóm đã tạo bảng Regression bên trong Sheet "Sample". Ta sẽ sử dụng bảng này làm số liệu để khảo sát.

### Phân tích:

Trước tiên, ta quan tâm vào chỉ số Multiple R. Nếu  $R > 0,9$  thì chỉ số này rất chặt chẽ, nếu  $0,9 > R > 0,5$  thì chỉ số này sẽ nói là khá chặt chẽ. Nếu  $R < 0,5$  thì chỉ số này sẽ không chặt chẽ tí nào. Ở đây, Multiple  $R = 0,88$  nên có thể coi như số liệu có tuyến tính khá chặt chẽ.

Nghĩa là, ta có thể coi như là sản lượng lúa ở Quảng Trị tuyến tính theo năm. Bây giờ, ta kiểm tra là tuyến tính này tăng hay giảm.

Bảng công thức:  $Y = \beta_0 * x + \beta_1$

Quan sát 2 chỉ số ở cột Coefficients, ta có thể viết được phương trình tuyến tính chính xác là gì, cụ thể là

$$Y = 9,645632353X - 18990,47191$$

Trong đó, Y là sản lượng, X là năm.

Bằng cách này, ta thấy sản lượng lúa ở Quảng Trị tăng trung bình 9 tấn/năm. Đây là 1 sự tiến triển khá tốt với nông nghiệp nước nhà, đặc biệt là một nơi không chuyên về lúa nước như Quảng Trị.

### 3. TEST HYPOTHESIS

Kiểm tra sản lượng lúa của Quảng Trị nó có trung bình là 300 tấn/ năm. Với độ tin cậy 95%.

Hypothesis :  $H_0 : \mu_0 = 300$

$H_1 : \mu_1 \neq 300$

Test Statistic:

$$T = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} = 1,54$$

$$\alpha = 0,05 \Rightarrow Z_{\alpha/2} = 1,96$$

Ta thấy rằng  $T < Z_{\alpha/2}$  nên ta không bác bỏ  $H_0$ , và bác bỏ  $H_1$ .

Nghĩa là giả thuyết sản lượng lúa trung bình năm của tỉnh Quảng Trị là 300 tấn được chấp nhận.

### III. Các công thức về excel

- AVERAGE: tính trung bình.

- VAR.S tính Sample Variance.

- SQRT(VAR.S) là Sample Deviation.

- VAR.P tính Population Variance.

- Các bảng dữ liệu phân tích hypothesis testing và phần hồi quy tuyến tính được generate bằng phần data analysis của Excel.

