A picture containing rectangle, screenshot, text, frame

Description automatically generated

**Giảng viên lý thuyết:** TS. Nguyễn Đình Thuân

**Giảng viên thực hành**: Nguyễn Minh Nhật

**Nhóm sinh viên thực hiện**: Nhóm 20

1. Cao Hoài Sang MSSV: 21522541

2. Thi Thành Công MSSV: 21521897

3. Nguyễn Trần Gia Kiệt MSSV: 21522258

4. Nguyễn Hoàng Đăng Khoa MSSV: 21520999

5. Cù Ngọc Hoàng MSSV: 21552086

**Môn học: PHÂN TÍCH DỮ LIỆU KINH DOANH**

ĐẠI HỌC QUỐC GIA TP.HỒ CHÍ MINH

TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN

KHOA HỆ THỐNG THÔNG TIN

BÁO CÁO LAB01

**BẢNG PHÂN CÔNG CÔNG VIỆC**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Cao Hoài Sang | Thi Thành Công | Nguyễn Trần Gia Kiệt | Cù Ngọc Hoàng | Nguyễn Hoàng Đăng Khoa |
| Câu 1a |  |  |  | x |  |
| Câu 1b | x |  |  |  |  |
| Câu 2 |  | x |  |  |  |
| Câu 3 |  |  | x |  |  |
| Câu 4 |  |  |  |  | x |
| File báo cáo | x | x | x | x | x |
| File thuyết trình | x | x | x | x | x |

**DANH SÁCH HÌNH ẢNH**

[*Figure 1 Kết quả các giá trị tính bằng MS Excel 5*](#_Toc160900198)

[*Figure 2 Code R 6*](#_Toc160900199)

[*Figure 3 Kết quả các giá trị tính bằng R 7*](#_Toc160900200)

[*Figure 4 Code Python 8*](#_Toc160900201)

[*Figure 5 Kết quả các giá trị tính bằng Python 9*](#_Toc160900202)

[*Figure 6 Kết quả các giá tri tính bằng Python 10*](#_Toc160900203)

[*Figure 7 Biểu đồ Histogram GDP Việt Nam 10*](#_Toc160900204)

[*Figure 8 Code R để vẽ biểu đồ histogram 10*](#_Toc160900205)

[*Figure 9 Biểu đồ Histogram GDP Việt Nam vẽ bằng R 11*](#_Toc160900206)

[*Figure 10 Code Python 12*](#_Toc160900207)

[*Figure 11 Biểu đồ Histogram vẽ bằng Python 13*](#_Toc160900208)

[*Figure 12 Vẽ BoxPlot GDP Việt Nam bằng Excel 14*](#_Toc160900209)

[*Figure 13 Code vẽ BoxPlot GDP Việt Nam bằng R 14*](#_Toc160900210)

[*Figure 14 BotPlot GDP Việt Nam 15*](#_Toc160900211)

[*Figure 15 Code Python vẽ biểu đồ BotPlot 16*](#_Toc160900212)

[*Figure 16 Biểu đồ BotPlot vẽ bằng Python 17*](#_Toc160900213)

[*Figure 17 Kết quả các giá trị tính bằng MS Excel 18*](#_Toc160900214)

[*Figure 18 Code R 18*](#_Toc160900215)

[*Figure 19 Kết quả các giá trị tính bằng R 19*](#_Toc160900216)

[*Figure 20 Code Python 20*](#_Toc160900217)

[*Figure 21 Kết quả các giá trị tính bằng Python 21*](#_Toc160900218)

[*Figure 22 Histogram dữ liệu sửa chữa máy tính 22*](#_Toc160900219)

[*Figure 23 Sử dụng ngôn ngữ R 22*](#_Toc160900220)

[*Figure 24 Histogram biểu đồ sửa chữa máy tính 23*](#_Toc160900221)

[*Figure 25 Histogram biểu đồ sửa chữa máy tính bằng Python 24*](#_Toc160900222)

[*Figure 26 BoxPlot dữ liệu sửa chữa máy tính 25*](#_Toc160900223)

[*Figure 27 Đoạn code ngôn ngữ R vẽ BoxPlot 25*](#_Toc160900224)

[*Figure 28 Boxplot biểu đồ sửa chữa máy tính bằng R 26*](#_Toc160900225)

[*Figure 29 Boxplot biểu đồ sửa chữa máy tính bằng Python 27*](#_Toc160900226)

[*Figure 30 Tính giá trị mean và standard deviation của 2 cột 28*](#_Toc160900227)

[*Figure 31 Tính toán các giá trị cần thiết để tính cov 28*](#_Toc160900228)

[*Figure 32 Tính sum 29*](#_Toc160900229)

[*Figure 33 Tính covariance coefficients 29*](#_Toc160900230)

[*Figure 34 Tính correlation 29*](#_Toc160900231)

[*Figure 35 Import dữ liệu từ file .csv 30*](#_Toc160900232)

[*Figure 36 Sử dụng hàm để tính cor 30*](#_Toc160900233)

[*Figure 37 Import các thư viện 30*](#_Toc160900234)

[*Figure 38 Đặt biến 31*](#_Toc160900235)

[*Figure 39 Tính covariance bằng công thức 31*](#_Toc160900236)

[*Figure 40 tính covariance bằng hàm của numpy 31*](#_Toc160900237)

[*Figure 41 Tính correlation 31*](#_Toc160900238)

[*Figure 42 Vào data analysis 32*](#_Toc160900239)

[*Figure 43 Chọn covariance 32*](#_Toc160900240)

[*Figure 44 Chọn các dòng và nhấn OK 32*](#_Toc160900241)

[*Figure 45 Kết quả tính covariance 33*](#_Toc160900242)

[*Figure 46 Thực hiện tính correlation 33*](#_Toc160900243)

[*Figure 47 Kết quả tính correlation 33*](#_Toc160900244)

[*Figure 48 Import dữ liệu 34*](#_Toc160900245)

[*Figure 49 tính cor 34*](#_Toc160900246)

[*Figure 50 tính cor 35*](#_Toc160900247)

[*Figure 51 tính cor 35*](#_Toc160900248)

[*Figure 52 Drop các cột 35*](#_Toc160900249)

[*Figure 53 Dùng hàm tính toán cor 36*](#_Toc160900250)

[*Figure 54: Giá trị trung bình của Square Feet 37*](#_Toc160900251)

[*Figure 55: Giá trị trung bình của Market Value 38*](#_Toc160900252)

[*Figure 56: Độ lệch chuẩn của Square Feet 39*](#_Toc160900253)

[*Figure 57: Độ lệch chuẩn của Market Value 40*](#_Toc160900254)

[*Figure 58: Z-score của Square Feet 40*](#_Toc160900255)

[*Figure 59: Z-score của Market Value 41*](#_Toc160900256)

[*Figure 60: Tìm thấy 2 giá trị z-score > 3 ở Square Feet 42*](#_Toc160900257)

[*Figure 61: Kiến thức được học 43*](#_Toc160900258)

[*Figure 62: Tính toán Q1, Q3 của Square Feet và Market Value 43*](#_Toc160900259)

[*Figure 63: Tính toán IQR dựa vào công thức được cung cấp 44*](#_Toc160900260)

[*Figure 64: Giá trị sàn của Square Feet và Market Value 44*](#_Toc160900261)

[*Figure 65: Giá trị trần của Square Feet và Market Value 45*](#_Toc160900262)

[*Figure 66: Ngoại lai của Square Feet 46*](#_Toc160900263)

[*Figure 67: Ngoại lai của Market Value 47*](#_Toc160900264)

[*Figure 68: Import dữ liệu và lưu vào biến home\_market 48*](#_Toc160900265)

[*Figure 69: Sử dụng attach() 48*](#_Toc160900266)

[*Figure 70: Tính Q3 48*](#_Toc160900267)

[*Figure 71: Tính IQR của Square Feet dựa vào Q3 và Q1 49*](#_Toc160900268)

[*Figure 72: Tính giá trị sàn của Square Feet 49*](#_Toc160900269)

[*Figure 73: Tính giá trị trần của Squart Feet 49*](#_Toc160900270)

[*Figure 74: Ngoại lại của bộ dữ liệu Square Feet 49*](#_Toc160900271)

[*Figure 75: Tính Q1 50*](#_Toc160900272)

[*Figure 76: Tính Q3 50*](#_Toc160900273)

[*Figure 77: Tính IQR của Market Value 50*](#_Toc160900274)

[*Figure 78: Tính giá trị sàn và giá trị trần của Marker Value 50*](#_Toc160900275)

[*Figure 79: Xác định ngoại lai trong Market Value 50*](#_Toc160900276)

[*Figure 80: Giá trị trung bình và độ lệch chuẩn của Square Feet 51*](#_Toc160900277)

[*Figure 81: z-score của các giá trị trong Square Feet 51*](#_Toc160900278)

[*Figure 82: Ngoại lai của Square Feet 52*](#_Toc160900279)

[*Figure 83: Giá trị trung bình, độ lệch chuẩn, z-score của các giá trị trong Market Value 52*](#_Toc160900280)

[*Figure 84: Import thư viện và đọc dữ liệu CSV 53*](#_Toc160900281)

[*Figure 85: Đọc dữ liệu csv 54*](#_Toc160900282)

[*Figure 86: Trích xuất các cột cụ thể 55*](#_Toc160900283)

[*Figure 87: Tính toán và hiển thị các ngoại lai 56*](#_Toc160900284)

[*Figure 88: Giá trị Q1, Q3, IQR, giá trị sàn và giá trị trần của Square Feet và Market Value 57*](#_Toc160900285)

[*Figure 89: Tính toán z-score của các giá trị trong Square Feet 58*](#_Toc160900286)

[*Figure 90: Tính toán z-score của các giá trị trong Market Value 58*](#_Toc160900287)

[*Figure 91: Xác định ngoại lai và in ra 59*](#_Toc160900288)

[*Figure 92: Không xác định được ngoại lai của Market Value với z-score 59*](#_Toc160900289)

[*Figure 93. Xác suất của B, đã biết A 65*](#_Toc160900290)

[*Figure 94. Xác suất của A, đã biết B 66*](#_Toc160900291)

[*Figure 95. Quy tắc tích tổng quát 66*](#_Toc160900292)

[*Figure 96. Xác suất từng phần 68*](#_Toc160900293)

**MỤC LỤC**

[**1.** **Câu 1** 5](#_Toc160900323)

[1.1. Yêu cầu 5](#_Toc160900324)

[1.2. Bài giải 5](#_Toc160900325)

[**2.** **Câu 2** 27](#_Toc160900326)

[2.1. Yêu cầu: 27](#_Toc160900327)

[2.2. Bài Giải: 27](#_Toc160900328)

[**3.** **Câu 3** 36](#_Toc160900329)

[3.1. Yêu cầu 36](#_Toc160900330)

[3.2. Bài giải 36](#_Toc160900331)

[**4.** **Câu 4** 59](#_Toc160900332)

[4.1. Yêu cầu 59](#_Toc160900333)

[4.2. Ví dụ: Bài toán xác suất 64](#_Toc160900334)

[DANH MỤC TÀI LIỆU THAM KHẢO 69](#_Toc160900335)

# **Câu 1**

## Yêu cầu

* Trực quan hóa dữ liệu () và thống kê mô tả () các gói dữ liệu bao gồm:
  + Dữ liệu về GDP của Việt Nam
  + Thời gian sửa chữa máy tính
* Sử dụng các phần mềm: Microsoft Excel, ngôn ngữ lập trình R và Python. Tính toán và phân tích ý nghĩa của các giá trị: Trung bình (mean), trung vị (median), độ phố biển (mode), tứ phân vị (quartile), phương sai (variance), độ lệch chuẩn (standard deviation), giá trị skewness, giá trị kurtosis.
* Trực quan hóa các dữ liệu trên bằng biểu đồ Histogram và Box Plot.

## Bài giải

* + 1. Đối với dữ liệu về GDP của Việt Nam
* Dữ liệu về GDP được thu thập tại website **VietStockFinance (**[*https://finance.vietstock.vn/du-lieu-vi-mo/43/thu-nhap.htm*](https://finance.vietstock.vn/du-lieu-vi-mo/43/thu-nhap.htm)**)**
  + - 1. Tính toán phân tích ý nghĩa các giá trị
* MS ExcelMS Excel



Figure 1 Kết quả các giá trị tính bằng MS Excel

* R



Figure 2 Code R

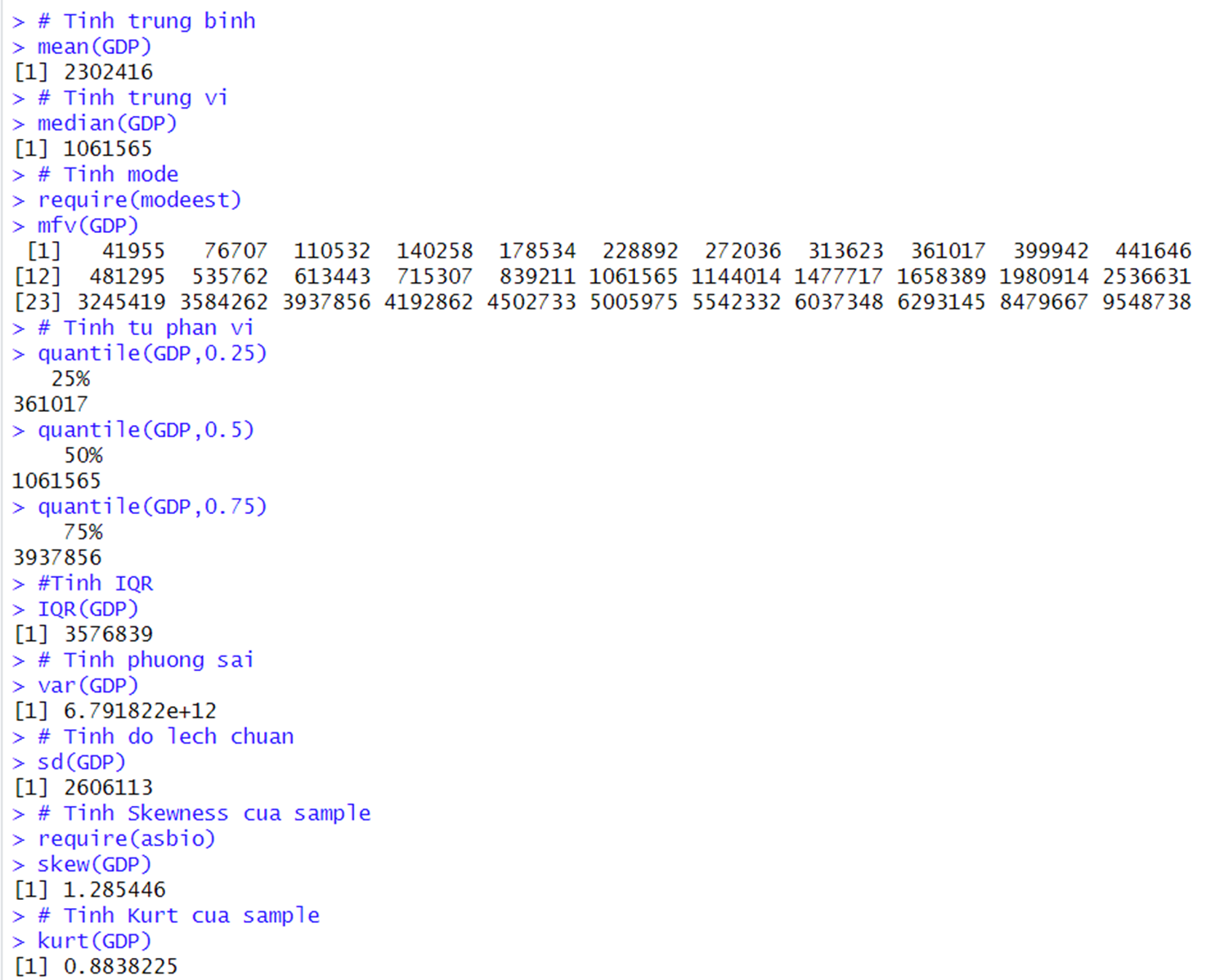


Figure 3 Kết quả các giá trị tính bằng R

* Python

A screenshot of a computer program

Description automatically generated

Figure 4 Code Python

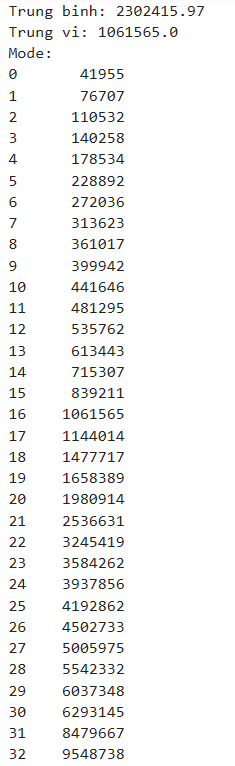


Figure 5 Kết quả các giá trị tính bằng Python

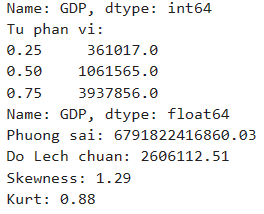


Figure 6 Kết quả các giá tri tính bằng Python

* + - 1. Trực quan hóa dữ liệu
* Biểu đồ Histogram
  + Sử dụng MS Excel: Ta chọn khoảng dữ liệu -> Trong phần Charts của tab Insert. Ta chọn Recommended Charts và chọn loại chart là Histogram

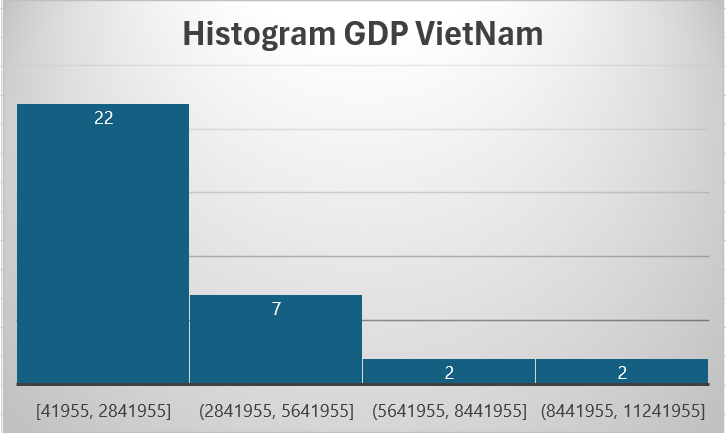


Figure 7 Biểu đồ Histogram GDP Việt Nam

* + Sử dụng ngôn ngữ R

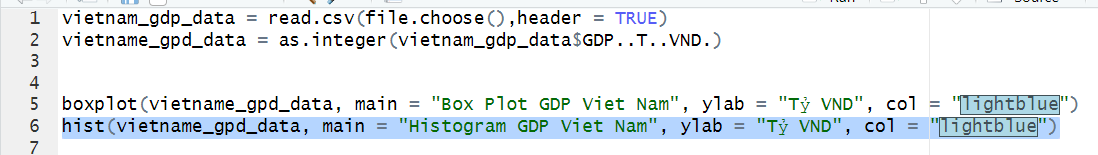


Figure 8 Code R để vẽ biểu đồ histogram

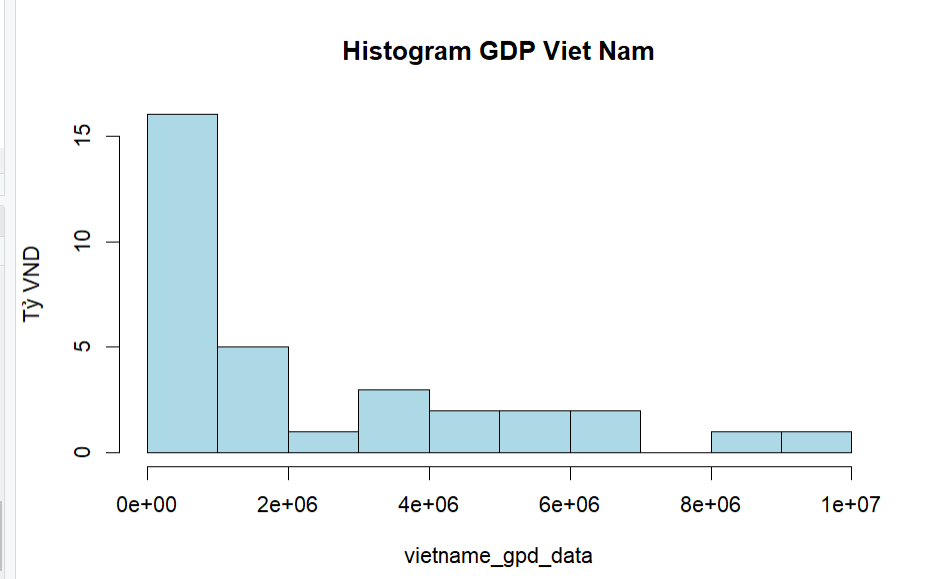


Figure 9 Biểu đồ Histogram GDP Việt Nam vẽ bằng R

* + Sử dụng ngôn ngữ Python

A screenshot of a computer

Description automatically generated

Figure 10 Code Python

A graph with blue squares

Description automatically generated

Figure 11 Biểu đồ Histogram vẽ bằng Python

* Biểu đồ Box Plot
  + Sử dung MS Excel

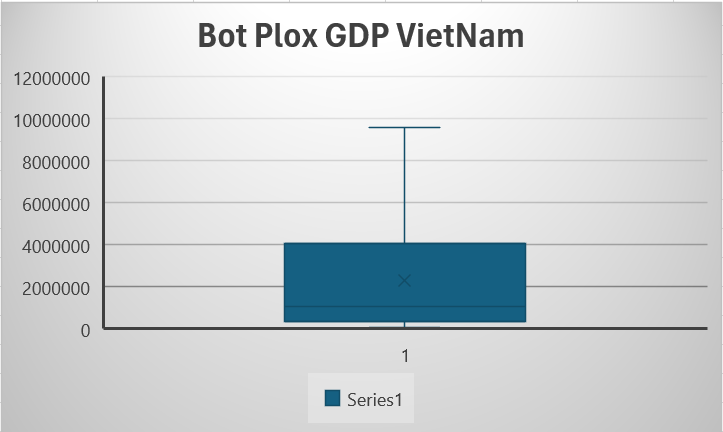


Figure 12 Vẽ BoxPlot GDP Việt Nam bằng Excel

* + Sử dụng ngôn ngữ R

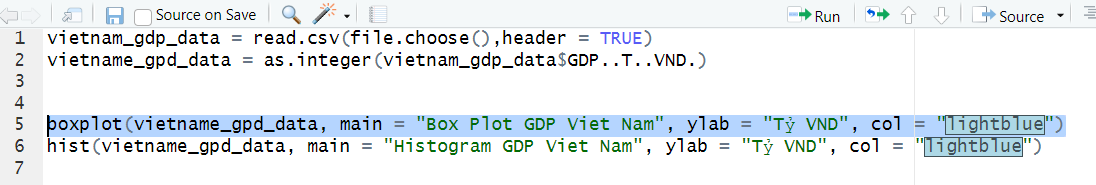


Figure 13 Code vẽ BoxPlot GDP Việt Nam bằng R

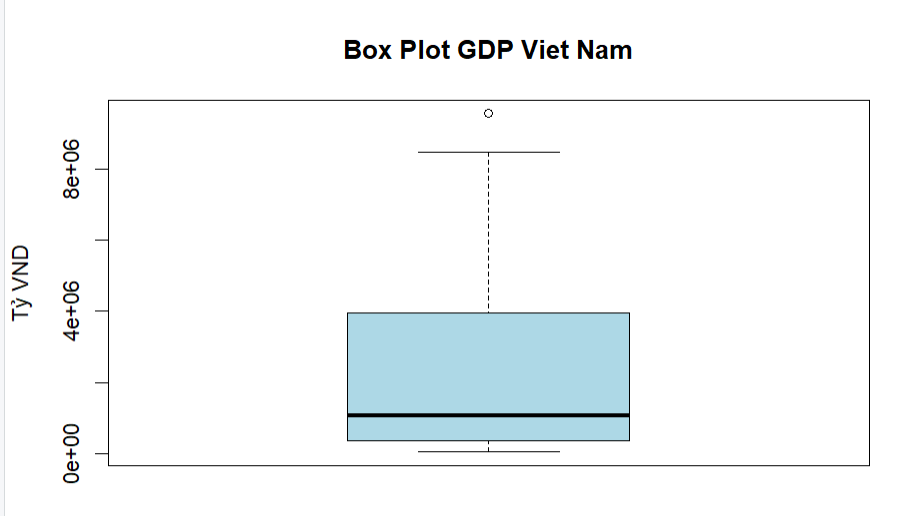


Figure 14 BotPlot GDP Việt Nam

* + Sử dụng ngôn ngữ Python

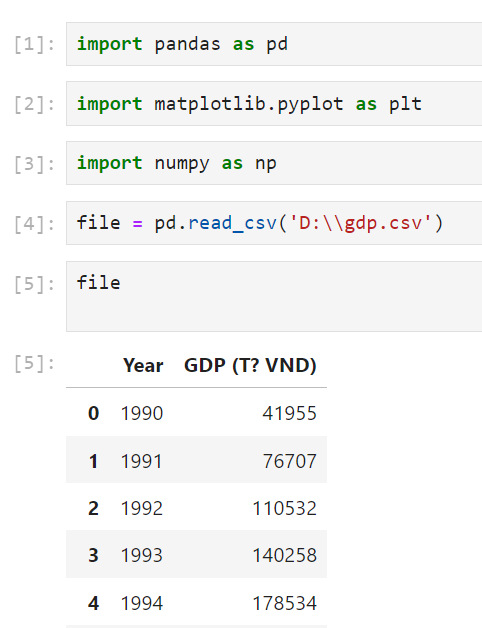


Figure 15 Code Python vẽ biểu đồ BotPlot

A screen shot of a graph

Description automatically generated

Figure 16 Biểu đồ BotPlot vẽ bằng Python

* + 1. Đối với dữ liệu về thời gian sửa chữa máy tính
       1. Tính toán phân tích ý nghĩa các giá trị
* MS Excel

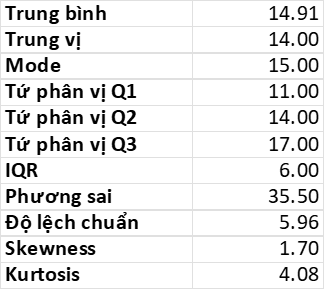


Figure 17 Kết quả các giá trị tính bằng MS Excel

* R

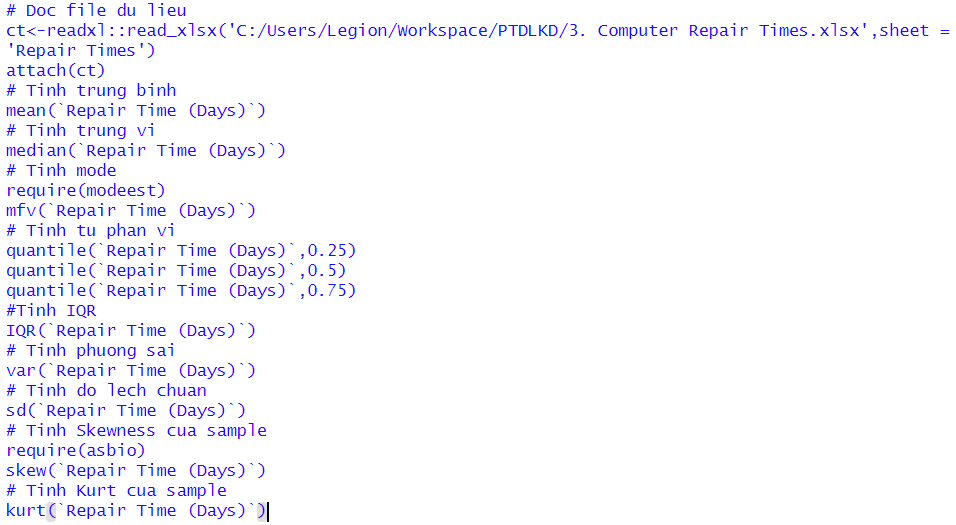


Figure 18 Code R

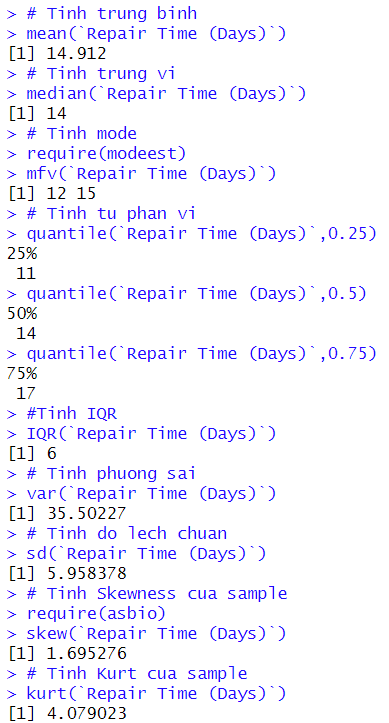


Figure 19 Kết quả các giá trị tính bằng R

* Python

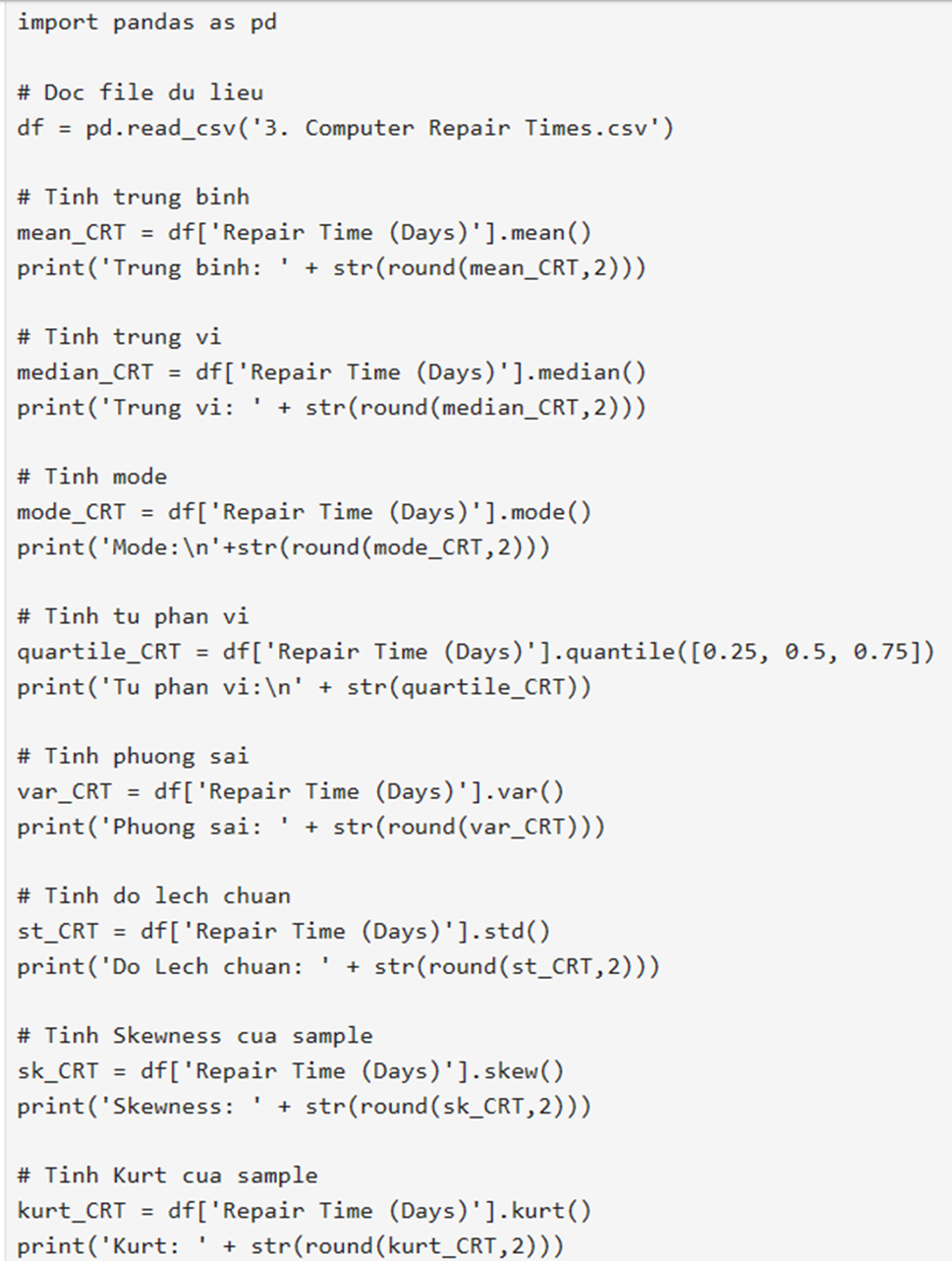


Figure 20 Code Python

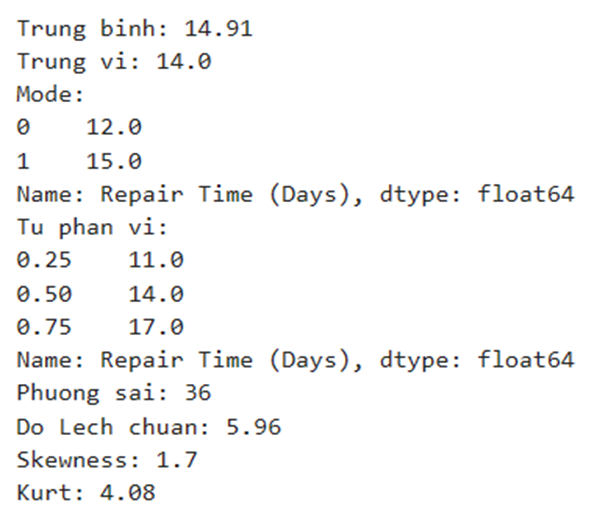


Figure 21 Kết quả các giá trị tính bằng Python

* + - 1. Trực quan hóa dữ liệu
* Biểu đồ Histogram
  + Sử dung MS Excel: Ta chọn khoảng dữ liệu -> Trong phần Charts của tab Insert. Ta chọn Recommended Charts và chọn loại chart là Histogram

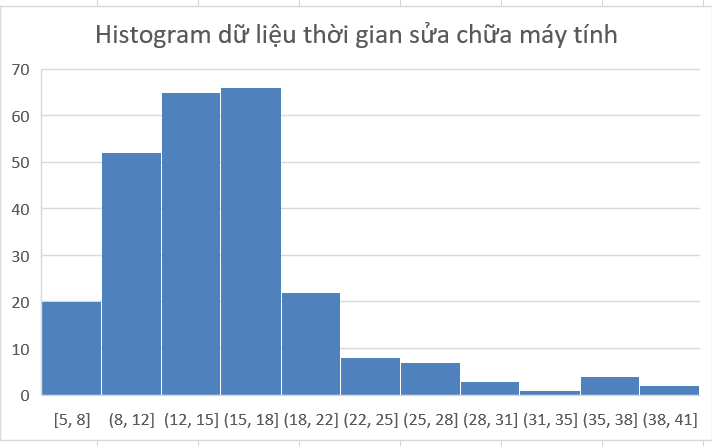


Figure 22 Histogram dữ liệu sửa chữa máy tính

* + Sử dụng ngôn ngữ R

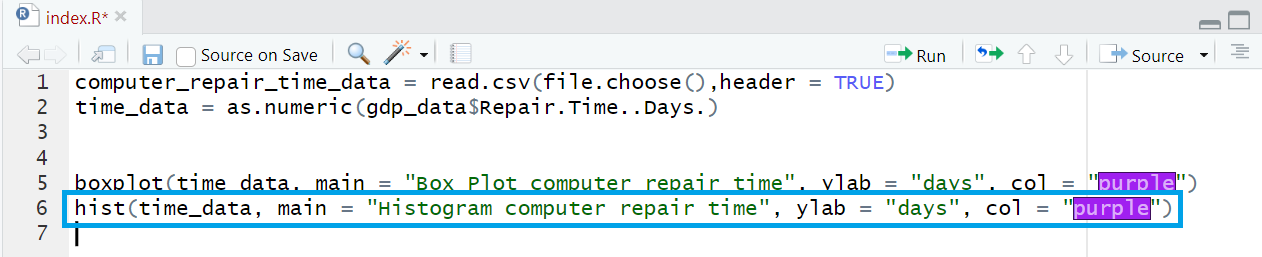


Figure 23 Sử dụng ngôn ngữ R

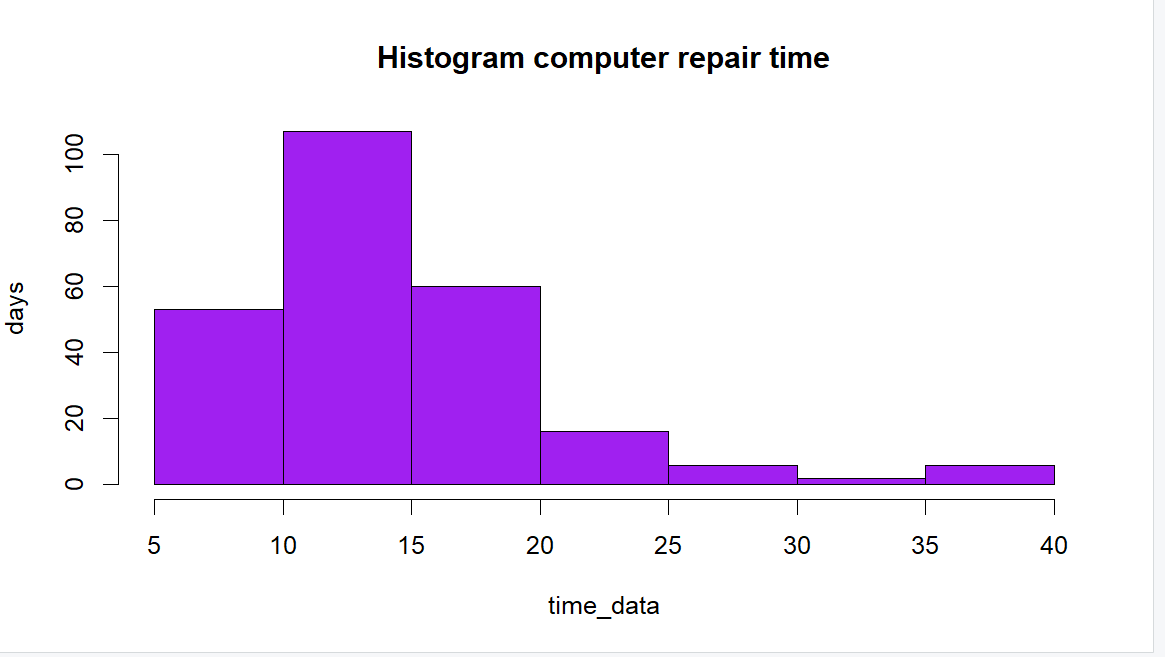


Figure 24 Histogram biểu đồ sửa chữa máy tính

* + Sử dụng ngôn ngữ Python

A screen shot of a graph

Description automatically generated

Figure 25 Histogram biểu đồ sửa chữa máy tính bằng Python

* Biểu đồ Box Plot
  + Sử dung MS Excel: Ta chọn khoảng dữ liệu -> Trong phần Charts của tab Insert. Ta chọn Recommended Charts và chọn loại chart là Box and Whisker

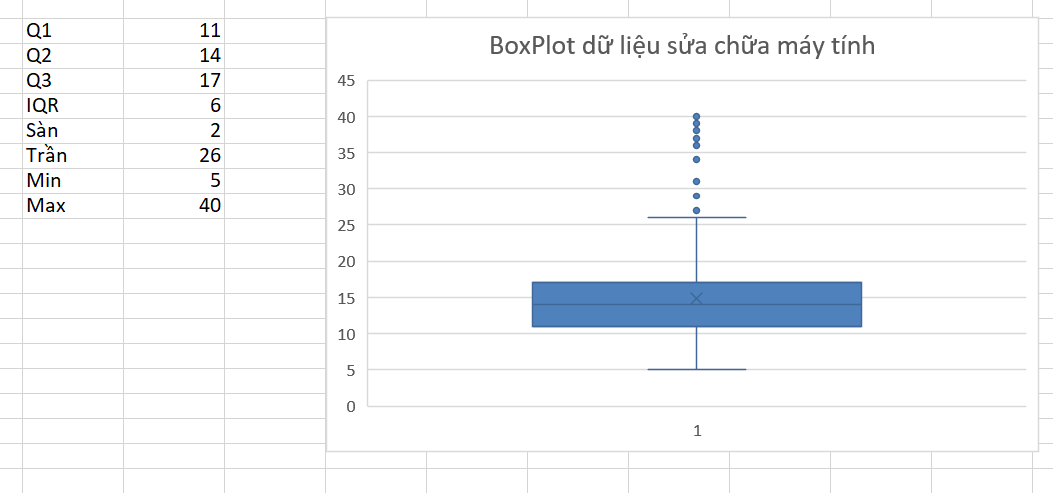


Figure 26 BoxPlot dữ liệu sửa chữa máy tính

* + Sử dụng ngôn ngữ R

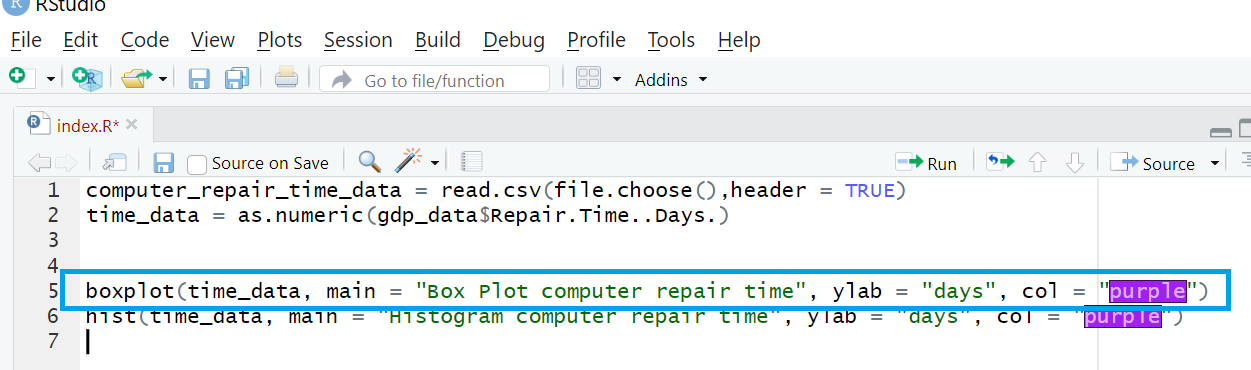


Figure 27 Đoạn code ngôn ngữ R vẽ BoxPlot

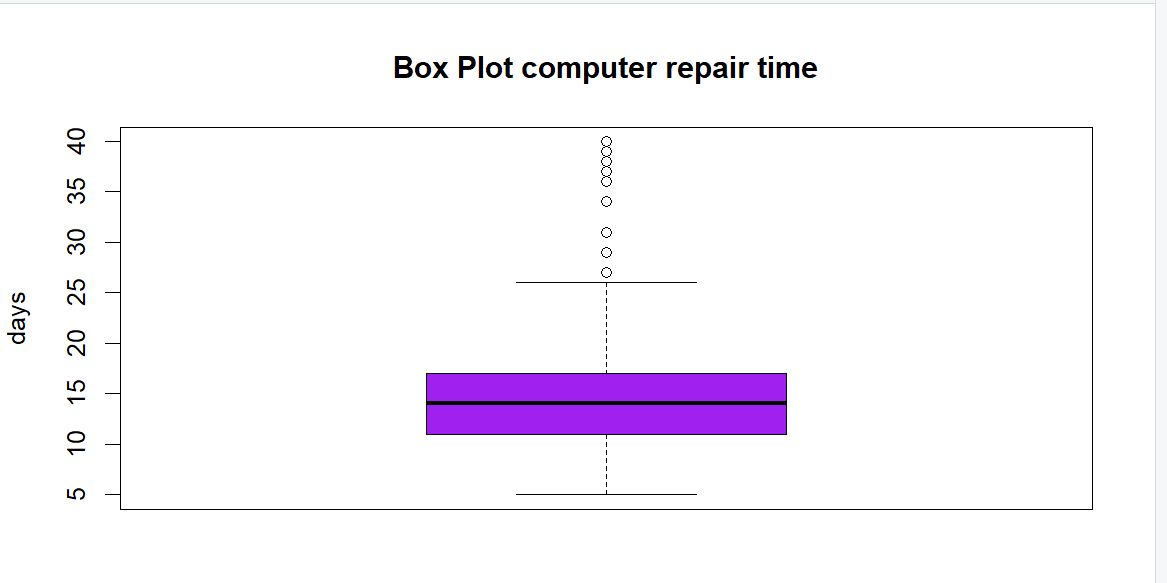


Figure 28 Boxplot biểu đồ sửa chữa máy tính bằng R

* + Sử dụng ngôn ngữ Python

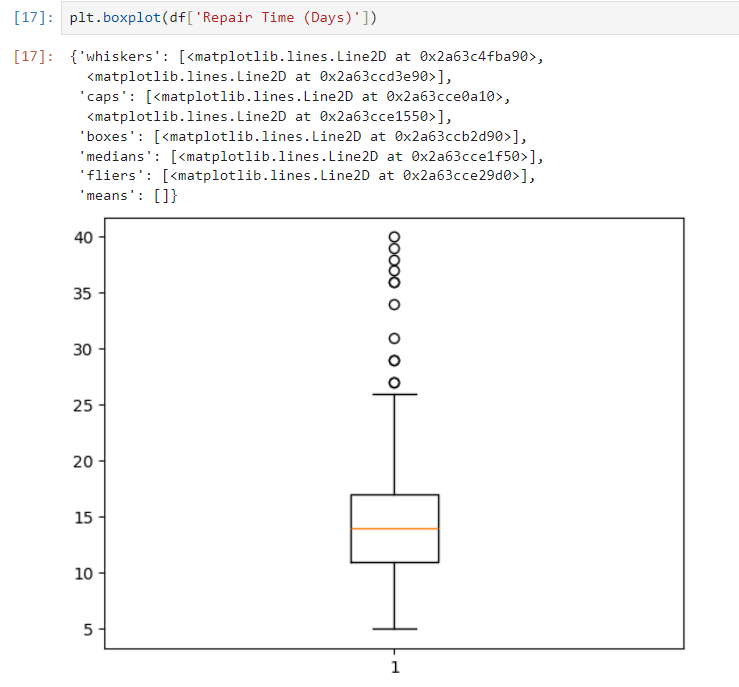


Figure 29 Boxplot biểu đồ sửa chữa máy tính bằng Python

# **Câu 2**

## Yêu cầu:

* Ý nghĩa của các hệ số tương quan và hiệp phương sai.
* Sử dụng MS Excel, R và Python với các ví dụ 4.21 và 4.22
  1. Bài Giải:
* Ý nghĩa của các hệ số tương quan và hiệp phương sai.
  + Hệ số tương quan: Hệ số tương quan là chỉ số thống kê đo lường mức độ tương quan của mối quan hệ giữa hai biến số. Hệ số tương quan thường được ký hiệu là "r" và có giá trị nằm trong khoảng -1 đến 1:
    - r<0 => 2 biến có mối quan hệ nghịch biến/tương quan âm (nghịch tuyệt đối khi r=-1)
    - r>0 => 2 biến có mối quan hệ đồng biến/tương quan dương(đồng biến tuyệt đối khi r=1)
    - r=0 => 2 biến độc lập với nhau
  + Hiệp phương sai(cov): là độ đo sự biến thiên cùng nhau của 2 biến ngẫu nhiên:
    - Cov > 0: nếu 2 biến có xu hướng thay đổi cùng nhau
    - Cov < 0: nếu 1 biến nằm trên giá trị kì vọng và biến kia có xu hướng nằm dưới giá trị kì vọng
    - Cov = 0: 2 biến độc lập tuyến tính
* Sử dụng MS Excel, R và Python với các ví dụ 4.21 và 4.22:
  + 4.21:
    - Sử dụng excel:
      * Tính giá trị mean và standard deviation của từng dòng:

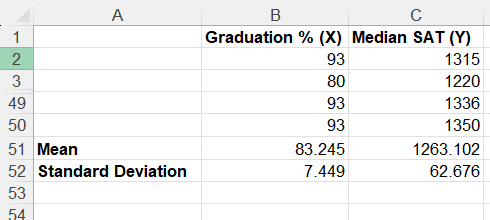


Figure 30 Tính giá trị mean và standard deviation của 2 cột

* + - * Tính giá trị X-Mean(X), X-Mean(Y) và (X-Mean(X))(X-Mean(Y))

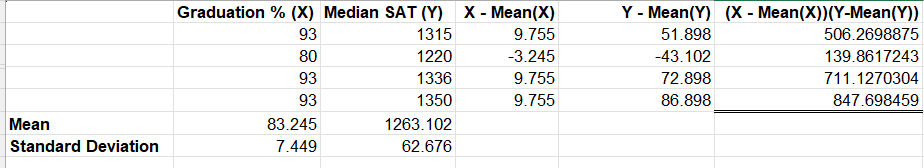


Figure 31 Tính toán các giá trị cần thiết để tính cov

* + - * Tính sum của (X-Mean(X))(X-Mean(Y)) và count:

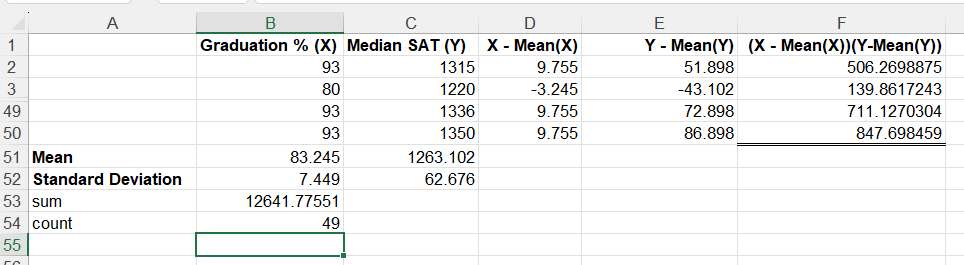


Figure 32 Tính sum

* + - * Tính hệ số tương quan:

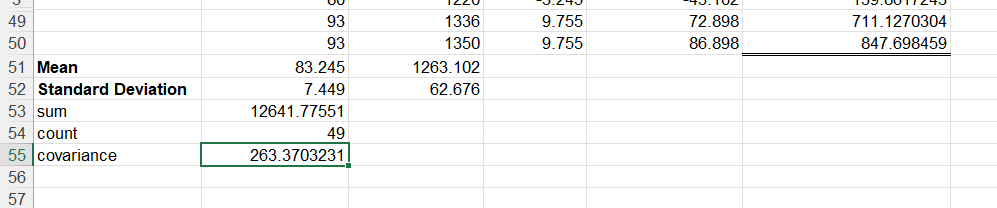


Figure 33 Tính hiệp phương sai

* + - * Sau đó có thể tính hệ số tương quan 1 cách dễ dàng:

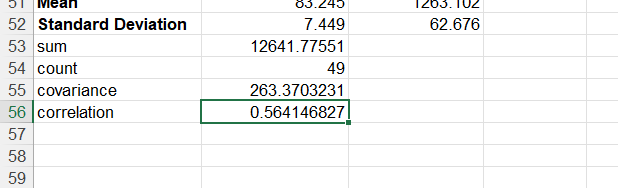


Figure 34 Tính correlation

* Sử dụng R:
  + Thêm dữ liệu vào từ file .csv:

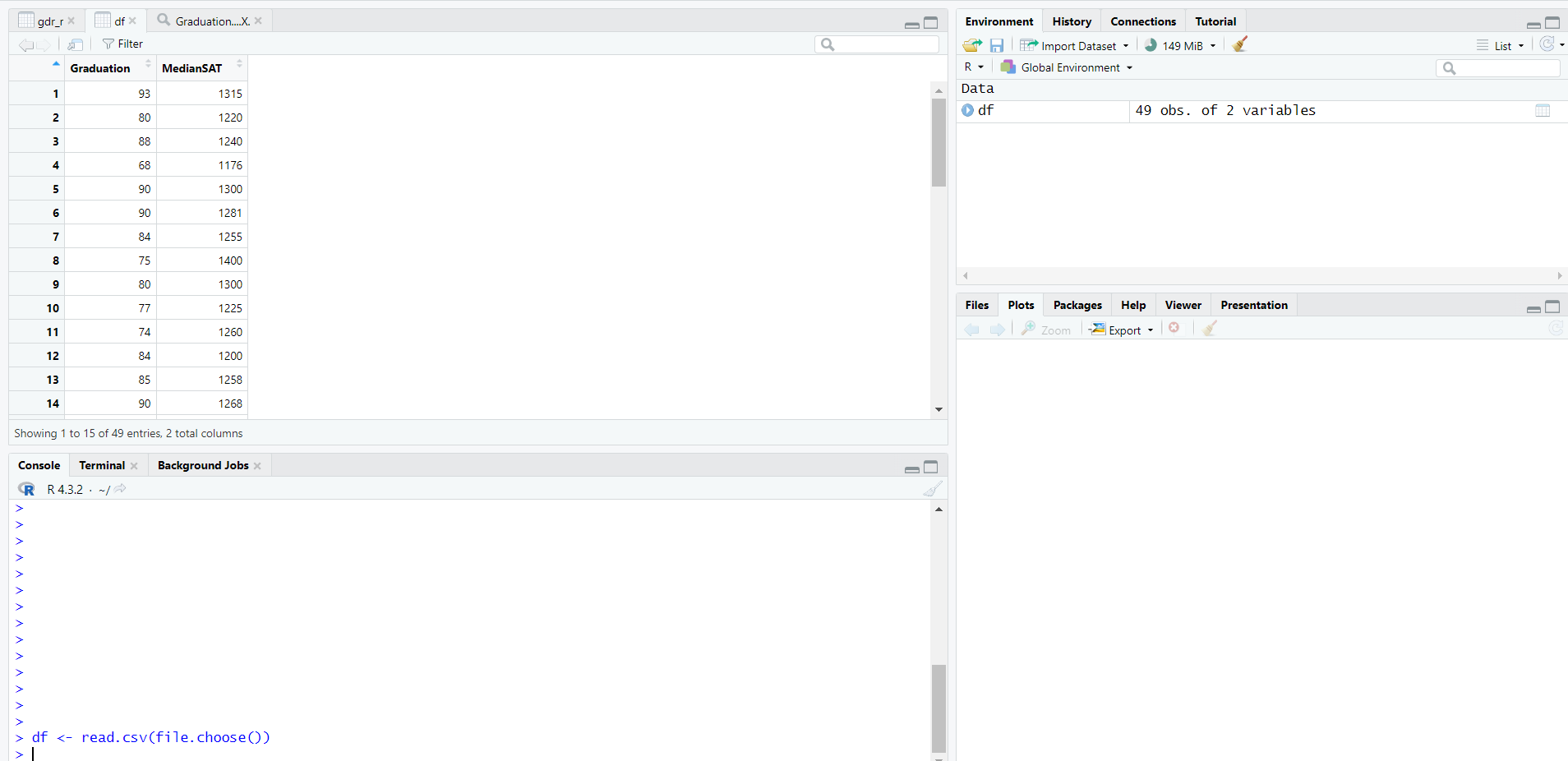


Figure 35 thêm dữ liệu từ file .csv

* + Tính hệ số tương quan giữa MedianSAT và graduation:

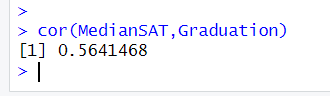


Figure 36 Sử dụng hàm để tính cor

* Sử dụng python
  + Thêm các thư viện cần thiết:

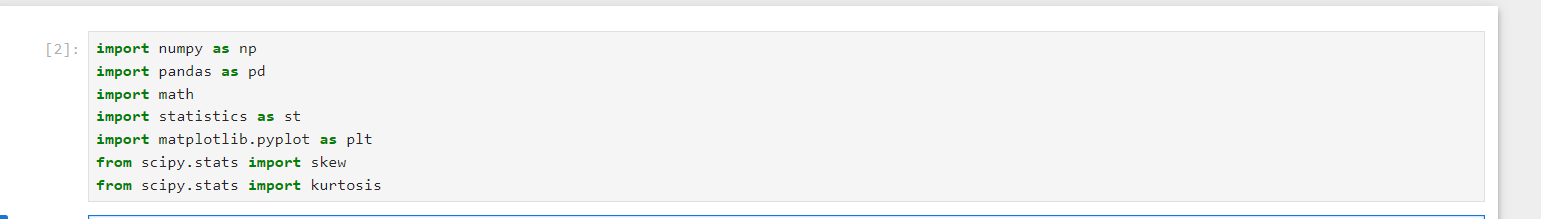


Figure 37 thêm các thư viện

* + thêm dữ liệu vào



* + Đặt biến để dễ truy xuất các cột vừa thêm

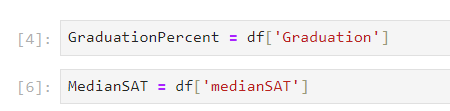
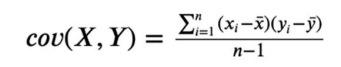


Figure 38 Đặt biến

* + Tính hiệp phương sai với công thức 

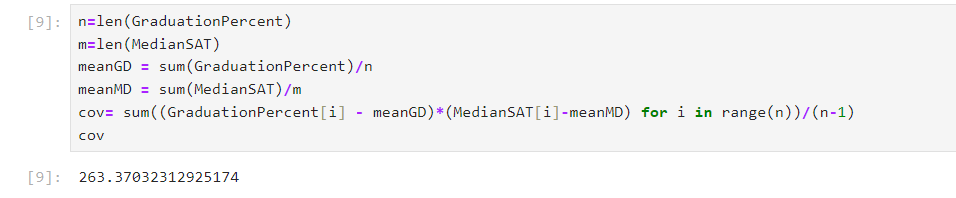


Figure 39 Tính hiệp phương sai bằng công thức

* + Hoặc tính hiệp phương sai với hàm của thư viện numpy:

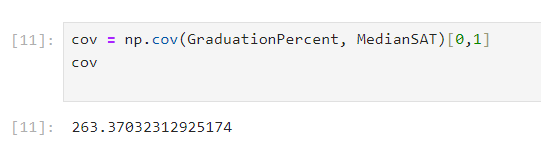


Figure 40 tính hiệp phương sai bằng hàm của numpy

* + Sau đó tính he số tương quan bằng thư viện của numpy

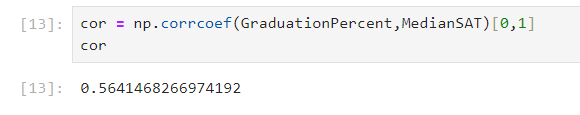


Figure 41 Tính correlation

* 4.22:
  + Sử dụng Excel:
    - Tính covariance bằng tool của excel:
      * + Vào Data Analysis ở cửa sổ data:

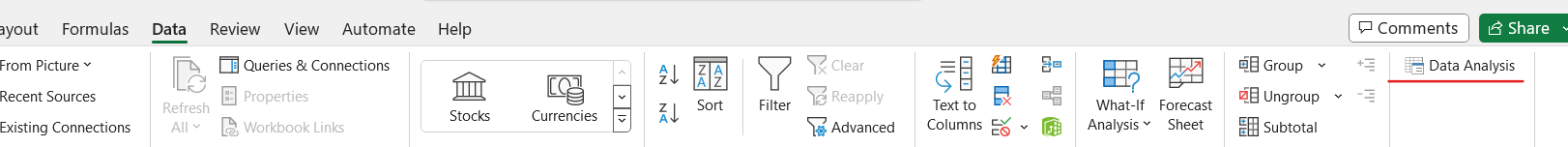


Figure 42 Vào data analysis

* + - * + Chọn covariance:

A screenshot of a computer

Description automatically generated

Figure 43 Chọn covariance

* + - * + Nhập input range, chọn label in first row và nhấn ok:

A screenshot of a computer

Description automatically generated

Figure 44 Chọn các dòng và nhấn OK

* + - * + Sau khi nhấn ok sẽ được kết quả như sau:

A screenshot of a calculator

Description automatically generated

Figure 45 Kết quả tính hiệp phương sai

* + - Tính hệ số tương quan bằng công cụ của excel:
      * + Tương tự vào data analyst và chọn correlation:

A screenshot of a computer

Description automatically generated

Figure 46 Thực hiện tính hệ số tương quan

* + - * + Kết quả sau khi nhấn OK:

A screenshot of a spreadsheet

Description automatically generated

Figure 47 Kết quả tính các hệ số tương quan

* + Sử dụng R:
    - Thêm dữ liệu vào:

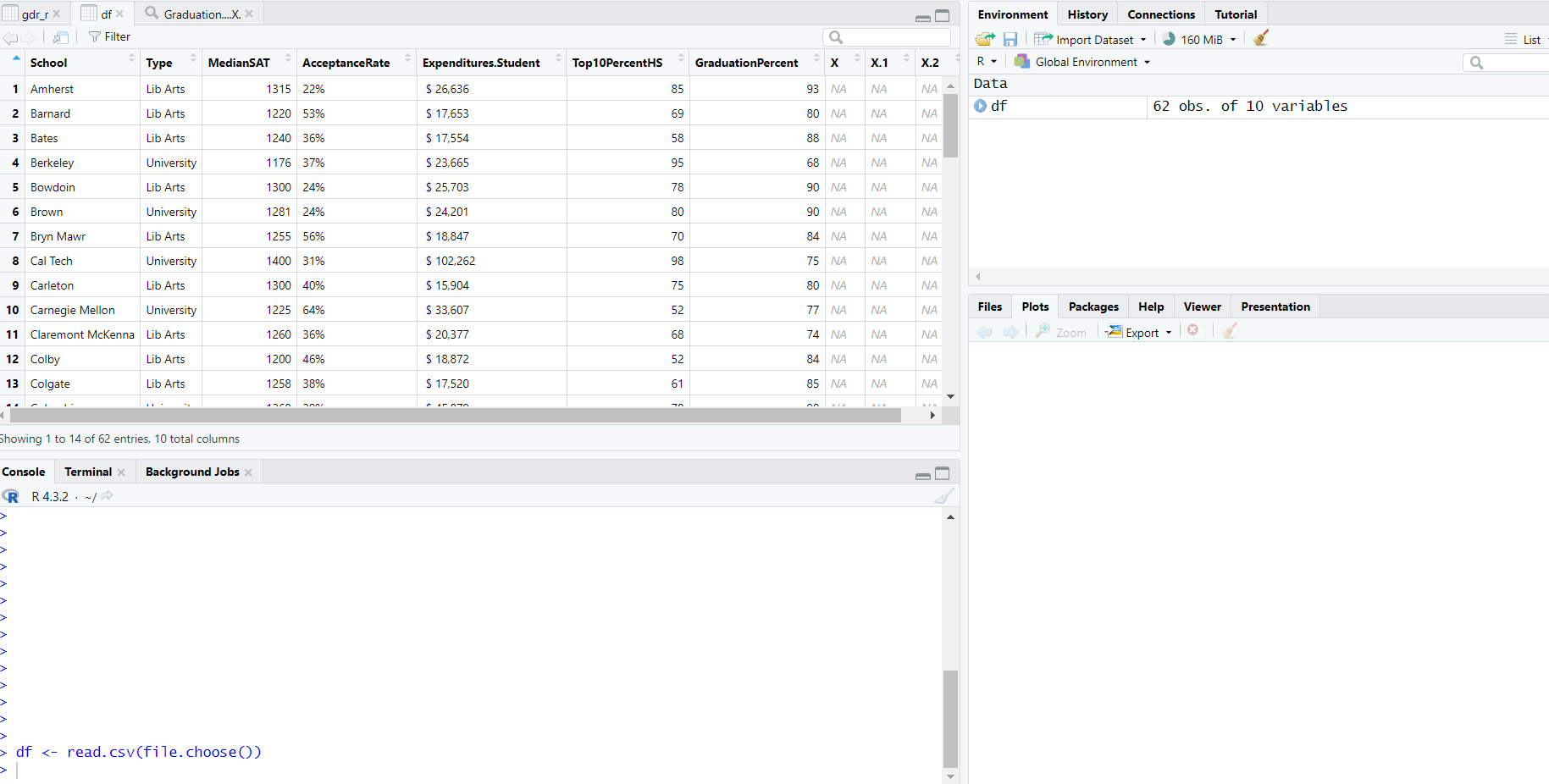


Figure 48 Thêm dữ liệu

* + - Chuyển đổi 1 số dữ liệu thành kiễu dữ liệu số:



* + - Tính toán các hệ số tương quan:

A screenshot of a computer

Description automatically generated

Figure 49 tính các hệ số tương quan

A screenshot of a computer program

Description automatically generated

Figure 50 tính các hệ số tương quan

A screenshot of a computer

Description automatically generated

Figure 51 tính các hệ số tương quan

* + - Tương tự với hiệp phương sai
  + Sử dụng Python:
    - Xoá các cột khác với kiễu dữ liệu số đi:

A screenshot of a computer

Description automatically generated

Figure 52 Xoá các cột

* + - Dùng hàm để tính cor của các thuộc tính:

A screenshot of a computer

Description automatically generated

Figure 53 Kết quả dùng hàm tính toán hệ số tương quan

# **Câu 3**

## Yêu cầu

* Ý nghĩa của Ngoại lai và sử dụng thống kê để xác định giá trị Ngoại lai.
* Sử dụng MS Excel, R và Python với các ví dụ 4.23.

## Bài giải

* Ý nghĩa của ngoại lai:
  + Ngoại lai là dữ liệu nằm tách biệt so với phần còn lại trong một tập hợp mẫu được chọn ngẫu nhiên từ một nhóm dữ liệu lớn. Nói cách khác, việc xác định điểm bất thường phụ thuộc vào người phân tích (hoặc theo một quy trình thống nhất) để quyết định xem điều gì sẽ được coi là bất thường. Trước khi phân loại các quan sát bất thường, cần phải hiểu rõ đặc điểm của các quan sát thông thường.
* Sử dụng MS Excel với ví dụ 4.23
  + Sử dụng Z-score
    - Đầu tiên, ta tính giá trị trung bình của SquareFeet và Market Value

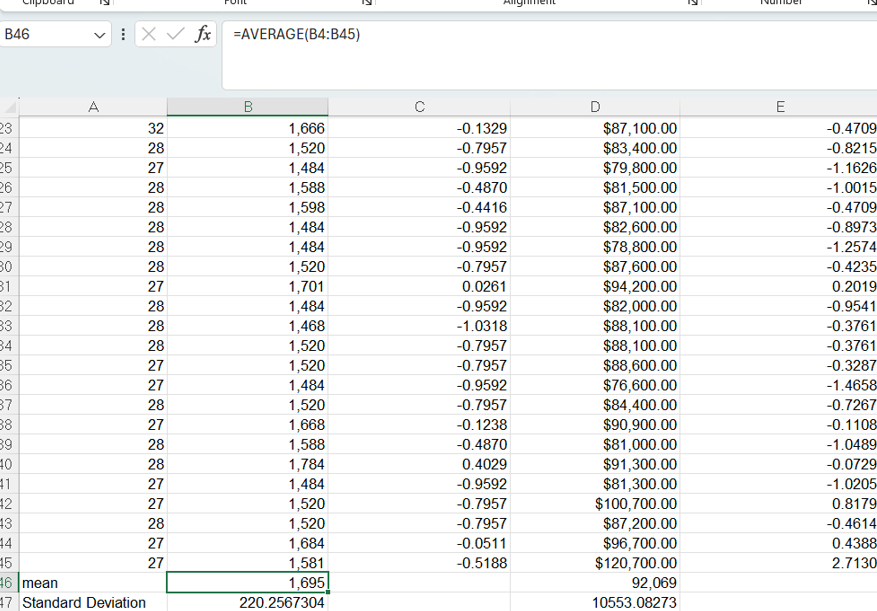


Figure 54: Giá trị trung bình của Square Feet

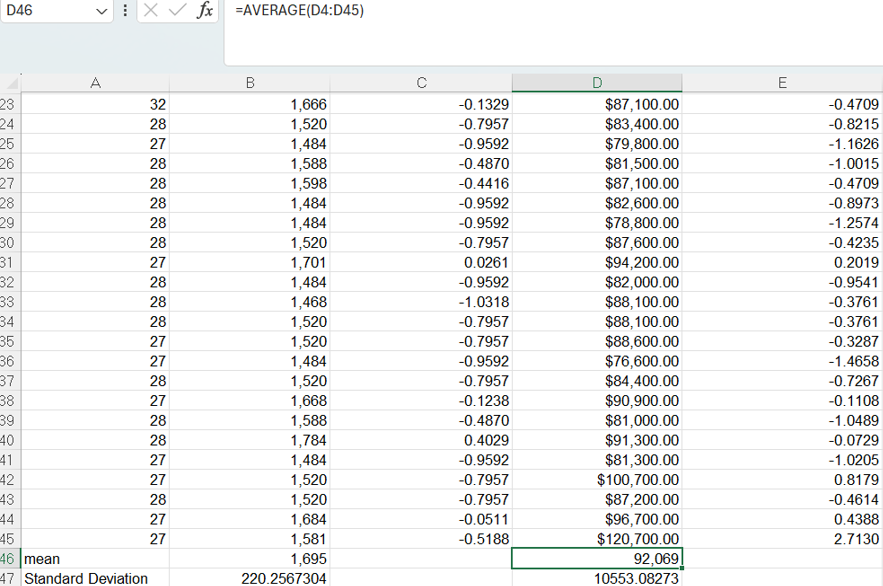


Figure 55: Giá trị trung bình của Market Value

* + - Tiếp theo, ta tính độ lệch chuẩn của SquareFeet và Market Value

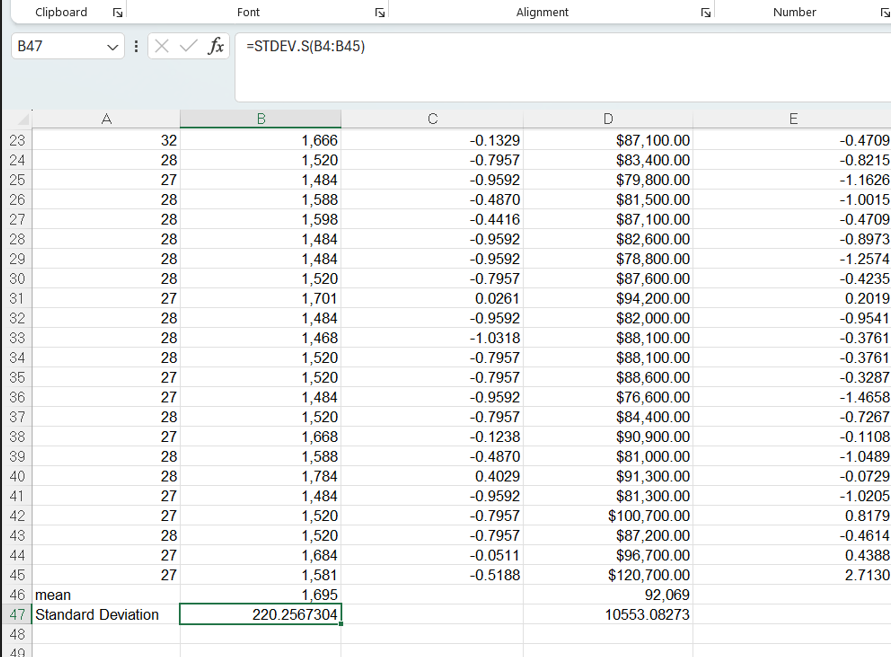


Figure 56: Độ lệch chuẩn của Square Feet

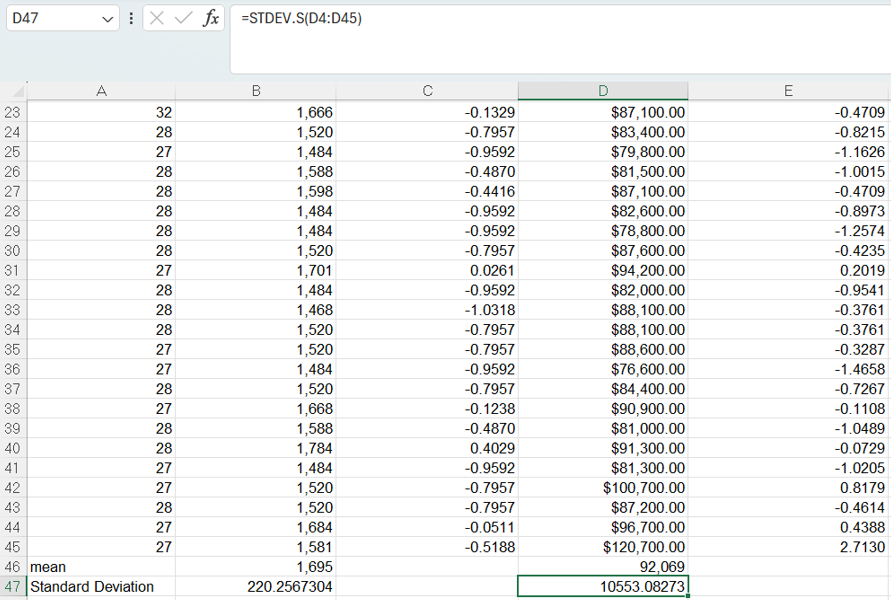


Figure 57: Độ lệch chuẩn của Market Value

* + - Sau khi có đủ các giá trị như giá trị trung bình và độ lệch chuẩn của Square Feet và Market Value với công thức

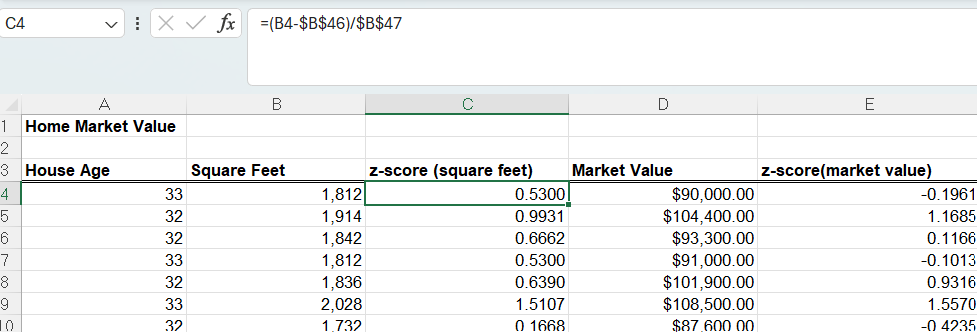


Figure 58: Z-score của Square Feet

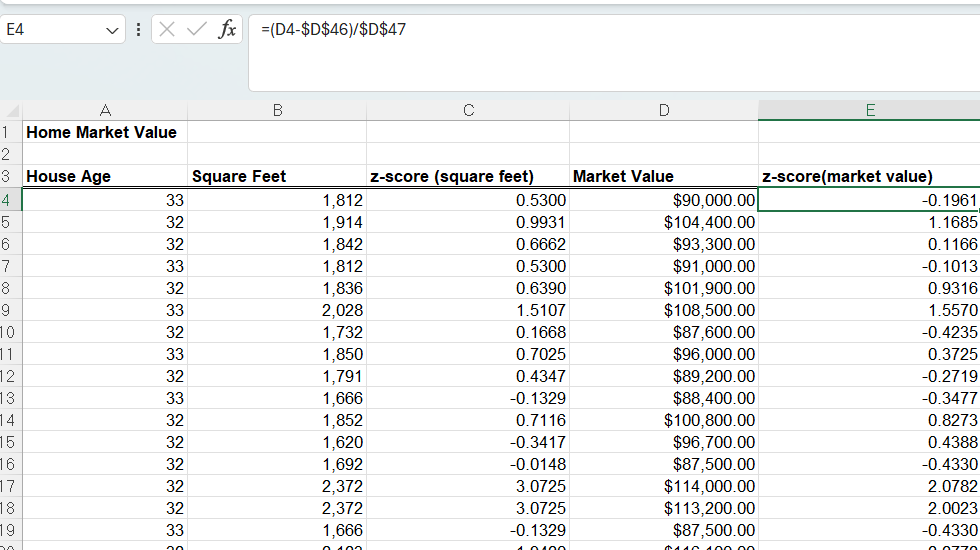


Figure 59: Z-score của Market Value

* + - Sau khi ta đã có đủ z-score cho từng giá trị trong Square Feet, ta sẽ bắt đầu so sánh, nếu z-score > 3 hoặc z-score < -3 thì giá trị đó là ngoại lệ (outlier)

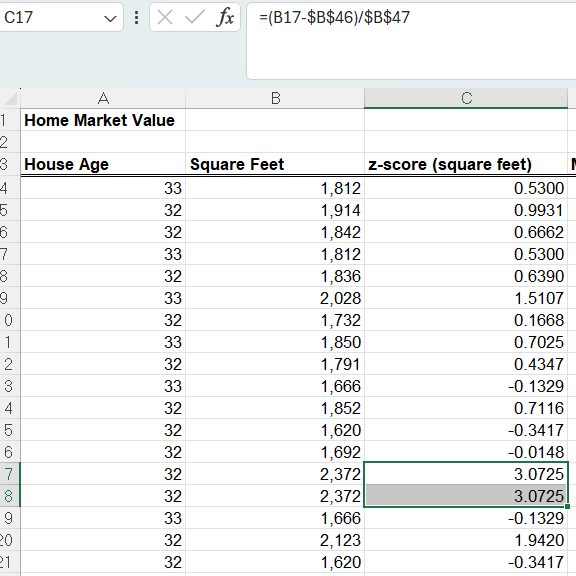


Figure 60: Tìm thấy 2 giá trị z-score > 3 ở Square Feet

* + - Như hình ở trên, ta thấy ở Square Feet có 2 giá trị có z-score > 3 nên đó là right outliers. Còn ở Market Value ta không tìm thấy giá trị nào có z-score > 3 hoặc z-score < -3
  + Sử dụng IQR:
    - Dựa vào công thức được học, bước đầu tiên ta phải tính toán được Q1, Q3 của Square Feet và Market Value.

A white background with black text

Description automatically generated

Figure 61: Kiến thức được học

A screenshot of a computer

Description automatically generated

Figure 62: Tính toán Q1, Q3 của Square Feet và Market Value

* + - Tính toán IQR theo công thức đã được học.

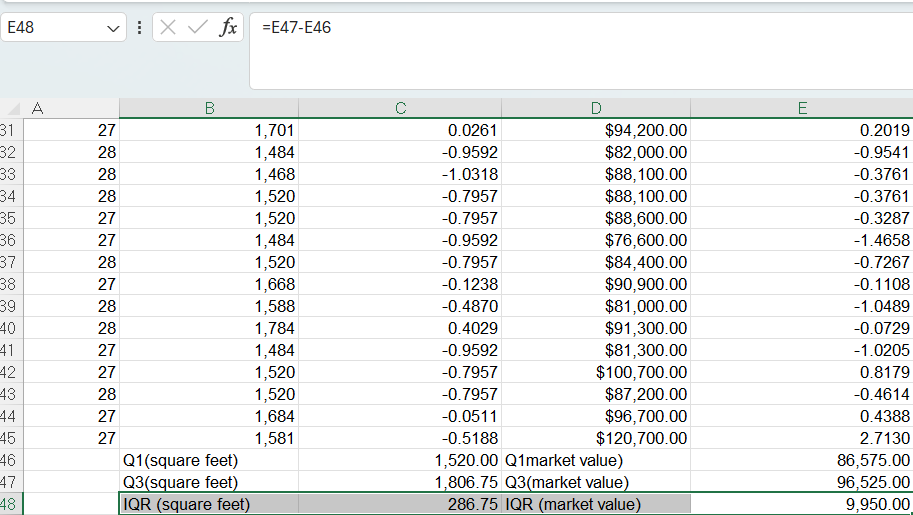


Figure 63: Tính toán IQR dựa vào công thức được cung cấp

* + - Dựa vào kết quả IQR, tính toán giá trị sàn = Q1 – 1.5\*IQR và giá trị trần = Q3 + 1.5\*IQR.

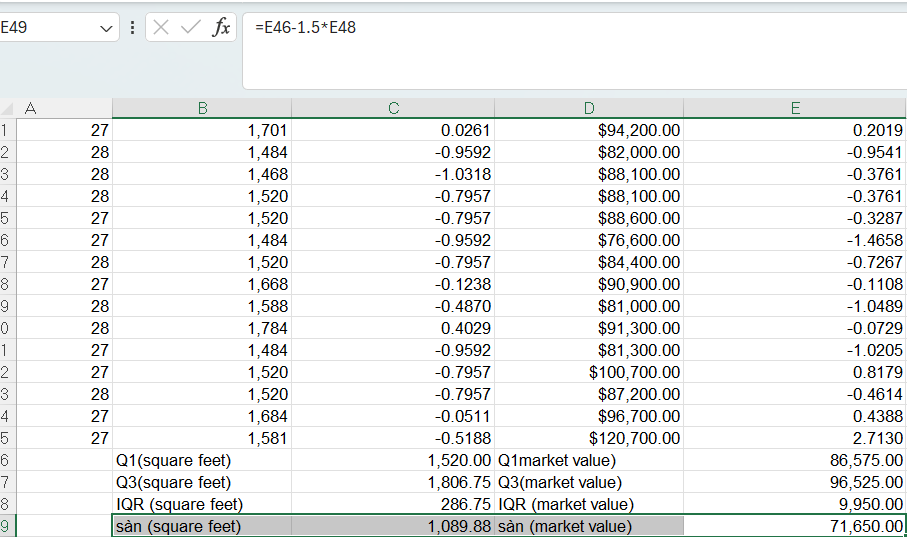


Figure 64: Giá trị sàn của Square Feet và Market Value

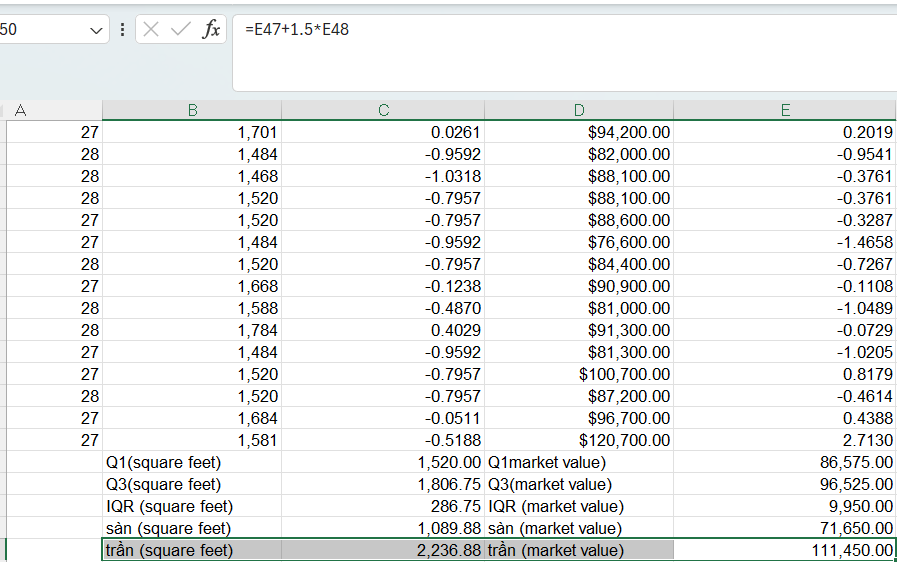


Figure 65: Giá trị trần của Square Feet và Market Value

* + - Dựa vào giá trị sàn và giá trị trần ta xác định được ngoại lai của Square Feet là 2372, 2372. Còn đối với Market Value là: 120700, 116100, 113200, 114000.

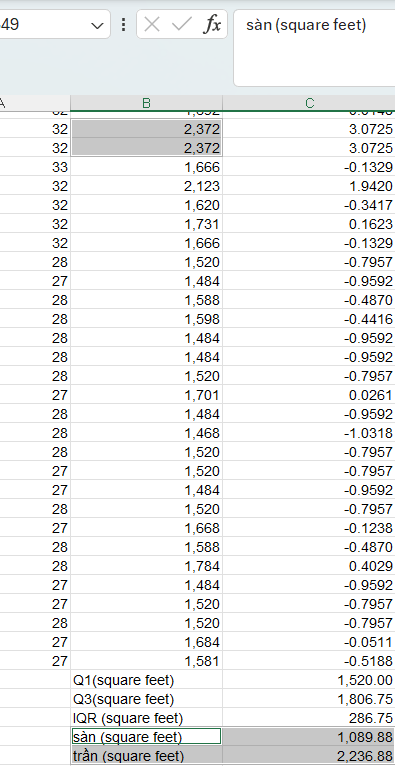


Figure 66: Ngoại lai của Square Feet

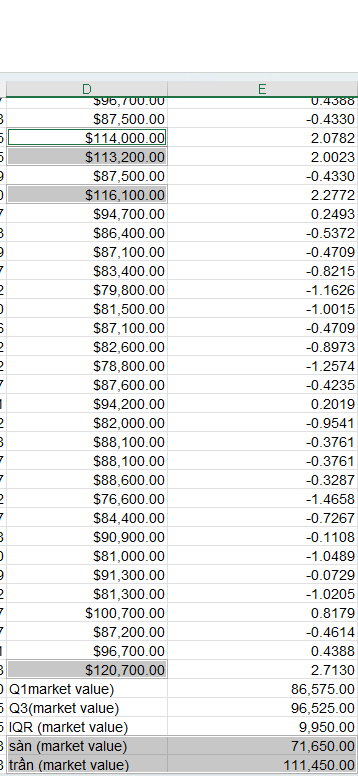


Figure 67: Ngoại lai của Market Value

* Sử dụng R
  + Sử dụng IQR
    - Đầu tiên, import dữ liệu bằng cách sử dụng hàm read.csv() để đọc dữ liệu từ một tập CSV, sau đó lưu vào một biến có tên home\_market.

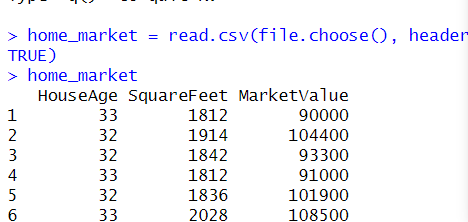


Figure 68: Import dữ liệu và lưu vào biến home\_market

* + - Để có thể làm việc được với biến home\_market, ta cần phải thêm bộ dữ liệu vào không gian làm việc bằng cách sử dụng attach(), lúc này bộ dữ liệu sẽ được gắn vào đường dẫn tìm kiếm của R.

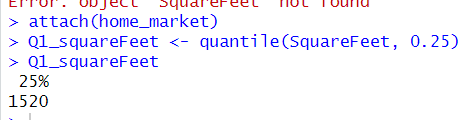


Figure 69: Sử dụng attach()

* + - Tiếp sau đó ta lần lượt tính Q1, Q3, IQR để có thể tìm được sàn và trần của tứ phân vị nhằm xác định được ngoại lai (outliers) của Square Feet.

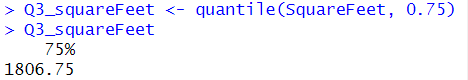


Figure 70: Tính Q3

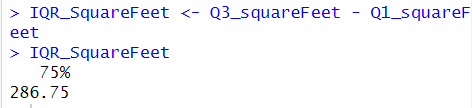


Figure 71: Tính IQR của Square Feet dựa vào Q3 và Q1

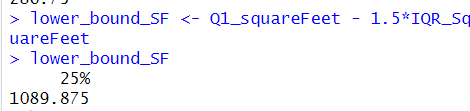
z

Figure 72: Tính giá trị sàn của Square Feet

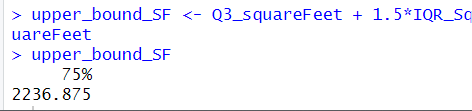


Figure 73: Tính giá trị trần của Squart Feet

* + - Xác định ngoại lai của bộ dữ liệu Square Feet

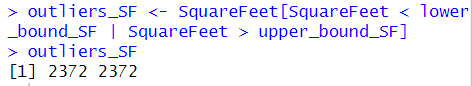


Figure 74: Ngoại lại của bộ dữ liệu Square Feet

* + - Như vậy ngoại lai trong Square Feet là 2372, 2372.
    - Thực hiện tương tự trên Market Value.

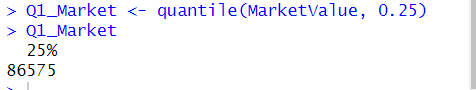


Figure 75: Tính Q1

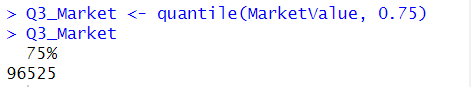


Figure 76: Tính Q3

A close-up of a text

Description automatically generated

Figure 77: Tính IQR của Market Value

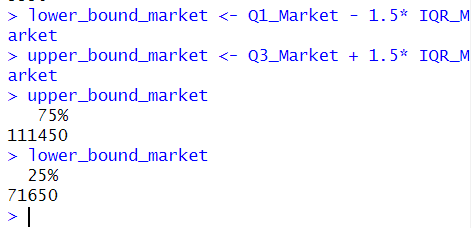


Figure 78: Tính giá trị sàn và giá trị trần của Marker Value

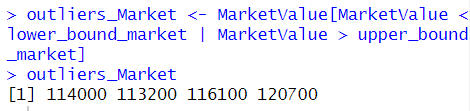


Figure 79: Xác định ngoại lai trong Market Value

* + - Như vậy với việc sử dụng IQR ta tìm được ngoại lai trong Market Value là 114000, 113200, 116100, 120700
  + Sử dụng z-score
    - Ta cũng tính toán lần lượt giá trị trung bình, độ lệch chuẩn và giá trị z-score của các giá trị trong Square Feet và Market Value và từ đó kiểm tra để xác định được ngoại lai như khi sử dụng Excel

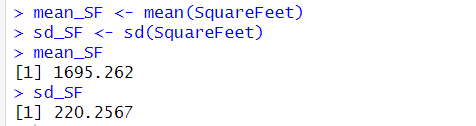


Figure 80: Giá trị trung bình và độ lệch chuẩn của Square Feet

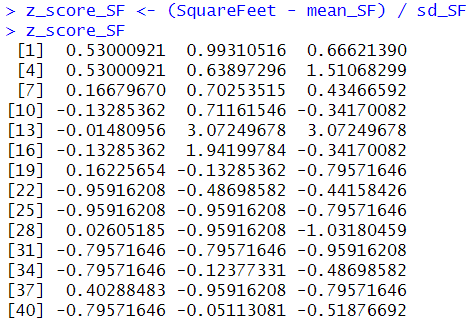


Figure 81: z-score của các giá trị trong Square Feet

* + - Sau khi có được z-score, ta so sánh xem giá trị nào có z-score > 3 hoặc z-score < -3 và sử dụng hàm which để có thể biết được vị trí của giá trị đó trong Square Feet

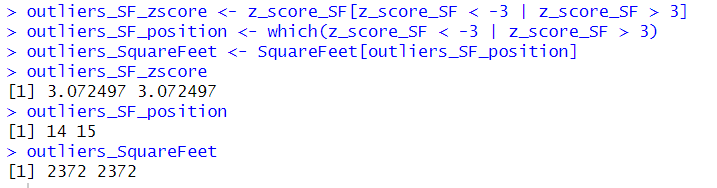


Figure 82: Ngoại lai của Square Feet

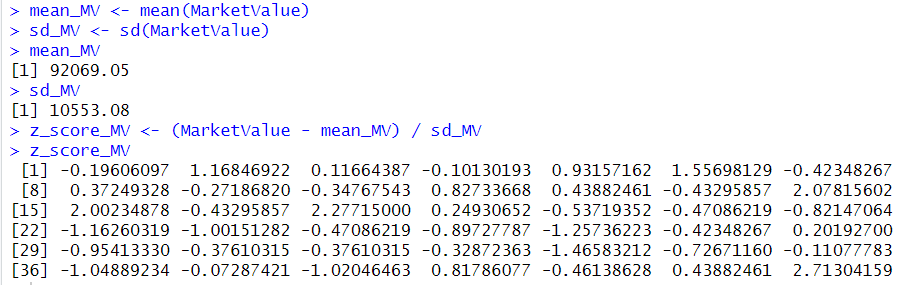
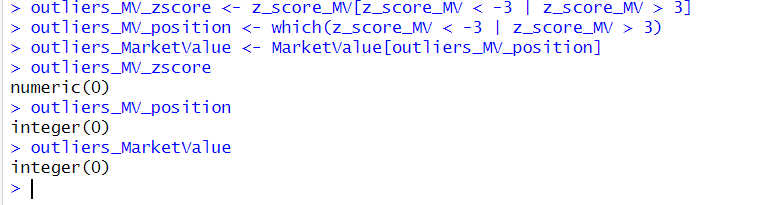


Figure 83: Giá trị trung bình, độ lệch chuẩn, z-score của các giá trị trong Market Value

* + - Và giống khi sử dụng Excel để xác định ngoại lai với z-score ta không tìm thấy giá trị ngoại lai nào trong Market Value



* Sử dụng Python
  + Sử dụng IQR:
    - Bước 1: Import các thư viện.

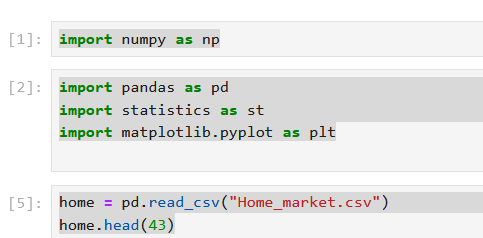


Figure 84: Import thư viện và đọc dữ liệu CSV

* + - Bước 2: Đọc dữ liệu CSV vào DataFrame home.

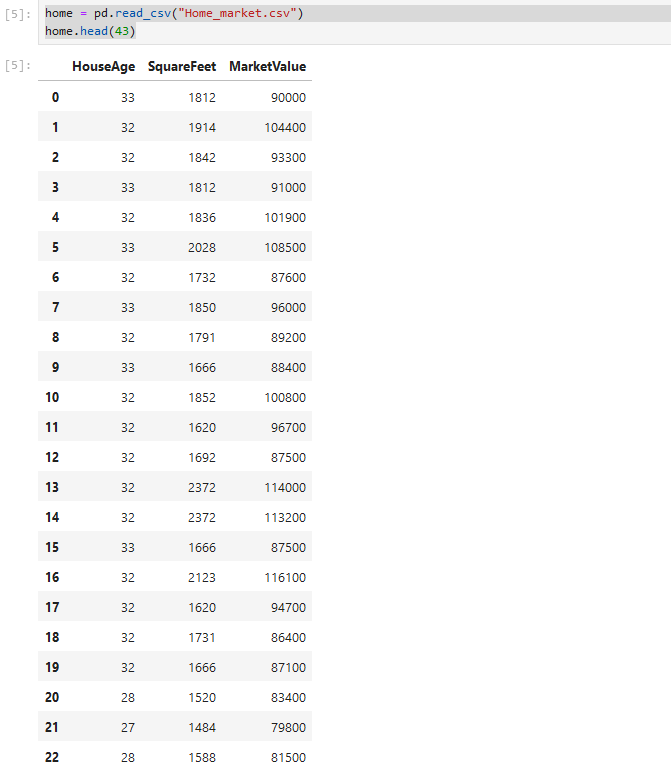


Figure 85: Đọc dữ liệu csv

* + - Bước 3: Trích xuất các cột cụ thể vào 2 biến sf và mv.

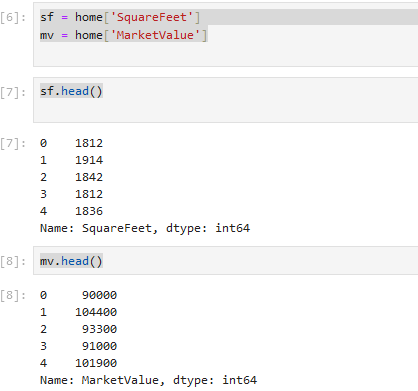


Figure 86: Trích xuất các cột cụ thể

* + - Bước 4 thực hiện tích toán Q1, Q3, IQR, giá trị sàn và giá trị trần từ đó so sánh và đưa ra ngoại lai như khi sử dụng Excel, R.

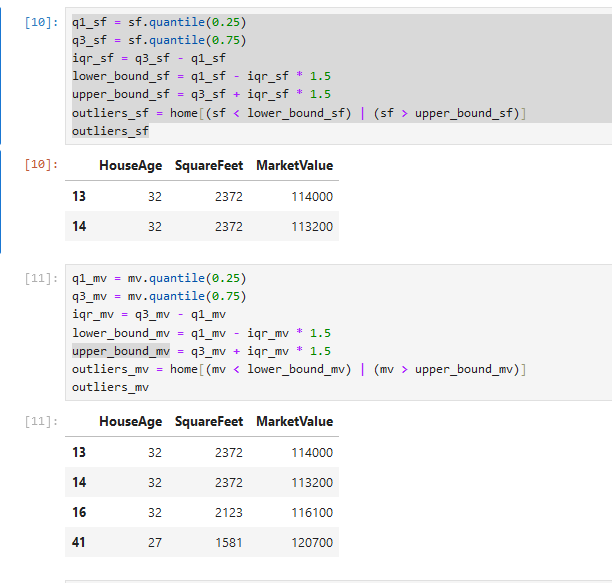


Figure 87: Tính toán và hiển thị các ngoại lai

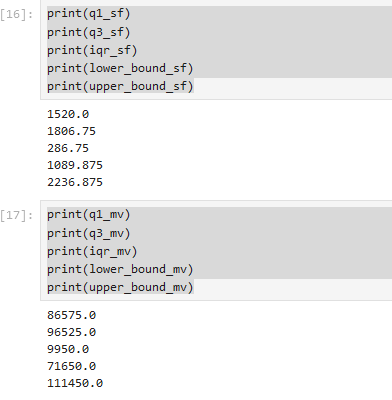


Figure 88: Giá trị Q1, Q3, IQR, giá trị sàn và giá trị trần của Square Feet và Market Value

* + Sử dụng z-score:
    - Ta tính toán lần lượt giá trị trung bình, độ lệch chuẩn và giá trị z-score của các giá trị trong Square Feet và Market Value và từ đó kiểm tra để xác định được ngoại lai như khi sử dụng Excel

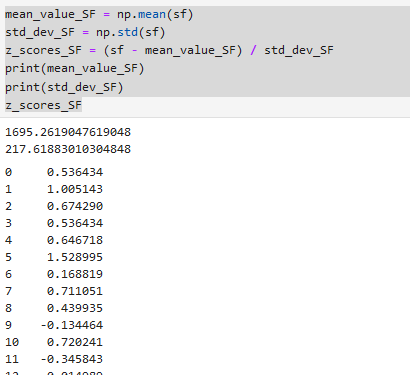


Figure 89: Tính toán z-score của các giá trị trong Square Feet

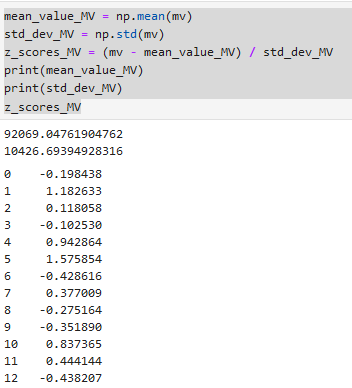


Figure 90: Tính toán z-score của các giá trị trong Market Value

* + - Sau đó ta kiểm tra xem ngoại lai thuộc về giá trị nào thông qua vị trị của giá trị đó
    - Đối với Square Feet thì đó là vị trị 13, 14. Hai giá trị thuộc vị trí đó là 2372, 2372

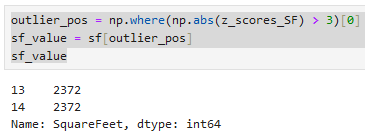


Figure 91: Xác định ngoại lai và in ra

* + - Đối với Market Value, thì không tìm được ngoại lai bằng z-score nên không thể tìm được vị trị của giá trị đó.

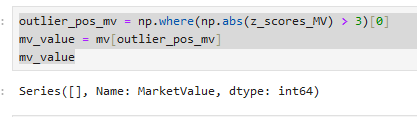


Figure 92: Không xác định được ngoại lai của Market Value với z-score

# **Câu 4**

## Yêu cầu

* Tư duy thống kê trong quyết định kinh doanh là gì? Đưa ra ví dụ minh họa

1. Định nghĩa Tư duy thống kê:

* Tư duy thống kê được định nghĩa như một triết lý học hỏi và hành động, một cách tư duy nhận diện hệ thống sản xuất là một chuỗi các quá trình liên kết với nhau. Trong đó, cần xác định, định lượng và giảm thiểu sự biến động của các quá trình này để tìm ra cơ hội cải tiến. Nó tạo ra tác động và tính hợp lý lớn hơn cho các hành động bằng cách nỗ lực cải thiện các quá trình thông qua việc giảm thiểu tính biến đổi [1].
* Và từ định hướng suy nghĩ đó, chúng ta có Phân tích nghiệp vụ doanh nghiệp (BA).
* Phân tích nghiệp vụ doanh nghiệp là một lĩnh vực rộng bao gồm việc sử dụng dữ liệu, thuật toán thống kê và công nghệ để trích xuất thông tin chi tiết và hỗ trợ ra quyết định trong các tổ chức. Mục tiêu của Phân tích doanh nghiệp là biến đổi dữ liệu thành những hiểu biết có thể hành động, giúp định hướng chiến lược và thúc đẩy cải tiến thông qua sự kết hợp của các phân tích mô tả, phân tích chẩn đoán, phân tích dự báo, cung cấp các cấp độ hiểu biết khác nhau và hỗ trợ các kiểu ra quyết định khác nhau [2].
* Phạm vi của Phân tích Doanh nghiệp bao gồm nhiều hoạt động và lĩnh vực rộng lớn trong một tổ chức, bao gồm:
  + **Thu thập và quản lý dữ liệu** (Data Collection and Management): Quy trình thu thập, lưu trữ và tổ chức dữ liệu từ các nguồn khác nhau theo một cách có cấu trúc.
  + **Phân tích dữ liệu** (Data Analysis): Quy trình sử dụng các kỹ thuật thống kê và toán học để xác định các mẫu hình và mối quan hệ trong dữ liệu, nhằm trích xuất thông tin chi tiết về các vấn đề kinh doanh.
  + **Lập mô hình dự báo** (Predictive Modeling): Sử dụng các thuật toán thống kê và kỹ thuật học máy để đưa ra dự đoán về các sự kiện hoặc xu hướng trong tương lai dựa trên dữ liệu lịch sử.
  + **Trực quan hóa dữ liệu** (Data Visualization): Quy trình tạo các biểu diễn trực quan của dữ liệu để giúp hiểu và truyền đạt các phân tích và thông tin hiệu quả hơn.
  + **Hỗ trợ ra quyết định** (Decision-Making Support): Sử dụng phân tích để cung cấp thông tin chi tiết và đề xuất cho người ra quyết định, giúp họ đưa ra những lựa chọn sáng suốt hơn.
  + **Phân tích hành vi khách hàng** (Customer Behavior Analysis): Quy trình phân tích dữ liệu khách hàng để có được thông tin chi tiết về hành vi và sở thích của họ, đồng thời đưa ra thông tin cho chiến lược kinh doanh.
  + **Nghiên cứu thị trường** (Market Research): Quy trình thu thập và phân tích dữ liệu về thị trường, khách hàng và đối thủ cạnh tranh để đưa ra thông tin cho chiến lược kinh doanh.
  + **Tối ưu hóa vận hành** (Operations Optimization): Sử dụng phân tích để tối ưu hóa các quy trình và hoạt động kinh doanh, đồng thời cải thiện hiệu quả, năng suất và sự hài lòng của khách hàng.
  + **Phân tích bán hàng và marketing** (Sales and Marketing Analysis): Đánh giá hiệu quả của các chiến lược bán hàng và marketing, xác định các cơ hội cải thiện.
  + **Tối ưu hóa chuỗi cung ứng** (Supply Chain Optimization): Tối ưu hóa hoạt động của chuỗi cung ứng, chẳng hạn như quản lý hàng tồn kho, logistics và vận tải.
  + **Phân tích, dự báo và báo cáo tài chính** (Financial Analysis and Reporting): Phân tích dữ liệu tài chính để hỗ trợ lập ngân sách, dự báo và ra quyết định.
  + **Quản lý và phân tích nhân sự** (Human Resource Management and Analysis): Kiểm tra dữ liệu nhân sự để cải thiện kế hoạch nguồn nhân lực, quản lý tài năng và sự hài lòng của nhân viên.

1. Định nghĩa quyết định kinh doanh

* Quyết định kinh doanh là một trong những yếu tố quan trọng nhất để điều hành doanh nghiệp. Nó bao gồm việc lựa chọn sản phẩm hoặc dịch vụ cung cấp, cách định giá, bán ở đâu và quảng bá chúng như thế nào. Ra quyết định đúng đắn có thể ảnh hưởng đến sự sống còn của một doanh nghiệp.
* Quyết định kinh doanh có thể ảnh hưởng đến mọi thứ, từ sản phẩm và dịch vụ của người kinh doanh đến cách quản lý tài chính. Có ba loại quyết định chính trong kinh doanh: chiến lược, tác chiến và vận hành:
  + **Quyết định chiến lược** (Strategic decisions): Là những quyết định dài hạn và thường liên quan đến những thay đổi lớn về định hướng của doanh nghiệp. Chúng thường được đưa ra bởi ban quản lý cấp cao và yêu cầu lập kế hoạch cẩn thận.
  + **Quyết định tác chiến** (Tactical decisions): Là những quyết định ngắn hạn và thường liên quan đến các nhiệm vụ hoặc dự án cụ thể. Chúng thường được đưa ra bởi các nhân viên tuyến đầu và thường liên quan đến việc tối ưu hóa cục bộ.
  + **Quyết định vận hành** (Operational decisions): Là những quyết định trung hạn và thường liên quan đến hoạt động hàng ngày của doanh nghiệp. Chúng thường được đưa ra bởi các nhà quản lý trung cấp.
* Loại quyết định người kinh doanh cần đưa ra sẽ phụ thuộc vào tình huống mà đang phải đối mặt. Ví dụ, nếu một thương lái đang tung ra một sản phẩm mới, họ sẽ cần đưa ra các quyết định chiến lược về thị trường muốn hướng đến và mức giá muốn đặt.
* Ra quyết định là cả một quy trình mà các doanh nghiệp sử dụng để xác định và lựa chọn phương hướng hành động tốt nhất để đạt được mục tiêu mong muốn. Các bước trong quy trình này thường là:

1. **Xác định vấn đề hoặc cơ hội** (Define the problem or opportunity): Xác định vấn đề là nền tảng vững chắc để phát triển các chiến lược phù hợp để đạt được kết quả mong muốn.
2. **Thu thập thông tin và lựa chọn** (Gather information and options): Bao gồm nghiên cứu các phương án hành động tiềm năng, đánh giá rủi ro và lợi ích, xem xét các khả năng thay thế và tham khảo ý kiến của các chuyên gia.
3. **Phân tích thông tin và các lựa chọn** (Analyse the information and options): Bước này đảm bảo bạn đã cân nhắc tất cả các khía cạnh của một quyết định tiềm năng, bao gồm cả rủi ro và lợi ích tiềm ẩn, cũng như bất kỳ vấn đề đạo đức nào có thể liên quan.
4. **Lựa chọn lựa chọn tốt nhất** (Select the best option): Để chọn được lựa chọn tốt nhất, điều quan trọng là phải cân nhắc dữ liệu và bằng chứng hỗ trợ cho từng lựa chọn và tham khảo ý kiến của các chuyên gia, những người có thể có kiến thức.
5. **Lập kế hoạch thực thi** (Plan for implementation): Một kế hoạch thành công sẽ giúp tối đa hóa cơ hội thành công bằng cách giảm thiểu thời gian, công sức hoặc tiền bạc lãng phí đồng thời hướng dẫn cách tiến hành khi đối mặt với những trở ngại hoặc thay đổi bất ngờ.
6. **Thực thi kế hoạch** (Implement the plan): Thực hiện kế hoạch là bước cuối cùng trong quy trình ra quyết định kinh doanh. Nó liên quan đến việc đưa vào hành động tất cả các quyết định được đưa ra trong các bước lập kế hoạch, phân tích và đánh giá trước đó.’
7. **Đánh giá kết quả** (Evaluate the results): Đánh giá kết quả của một quyết định là một bước quan trọng trong quy trình ra quyết định. Nó cung cấp phản hồi về việc liệu phương hướng hành động đã chọn có đạt được kết quả mong muốn hay không.
8. Định nghĩa của tư duy thống kê trong quyết định kinh doanh:

* Tư duy thống kê trong kinh doanh đề cập đến quá trình sử dụng các phương pháp toán học và thống kê để thu thập, tổ chức và diễn giải dữ liệu của doanh nghiệp. Dữ liệu này có thể được sử dụng để tìm hiểu thêm về những gì người kinh doanh đang làm hiện tại và giúp dự đoán trong tương lai.
* Doanh nghiệp cần biết khách hàng đang mua gì, có bao nhiêu khách hàng và những quyết định nào mà khách hàng đưa hàng ngày. Tất cả các quyết định được đưa ra trong doanh nghiệp cần dựa trên thông tin do phân tích thống kê cung cấp để đảm bảo kết quả tích cực.
* Những lợi ích của Tư duy thống kê trong kinh doanh:
  + Xác định cơ hội (Identifying Opportunities).
  + Hiểu biết hành vi khách hàng (Understanding Customer Behavior).
  + Xác định thị trường mục tiêu chính xác (Determining the Correct Target Market).
  + Đánh giá sản phẩm hoặc dịch vụ (Evaluating Products or Services).
* Tư duy Thống kê trong Quyết định Kinh doanh (Statistical thinking in business decision-making) bao gồm một số yếu tố then chốt, mỗi yếu tố đóng vai trò quan trọng trong việc tận dụng hiệu quả dữ liệu để ra các quyết định sáng suốt:
  + **Thu thập dữ liệu và đảm bảo chất lượng** (Data Collection and Quality Assurance): Xác định các nguồn dữ liệu liên quan và đảm bảo chất lượng của dữ liệu được thu thập. Điều này bao gồm việc xác định các số liệu, thiết lập quy trình thu thập dữ liệu và triển khai các biện pháp kiểm soát chất lượng để đảm bảo tính chính xác và tin cậy của dữ liệu.
  + **Hiểu và quản lý rủi ro** (Understand and manage risk factor): Các quyết định kinh doanh hiếm khi được đưa ra với thông tin hoàn hảo. Tư duy thống kê giúp các doanh nghiệp định lượng và cân nhắc bài toán rủi ro, cho phép họ đưa ra quyết định tốt hơn ngay cả khi thiếu thông tin đầy đủ.
  + **Xác định và phân tích các mẫu hình** (Identify and analyze patterns): Doanh nghiệp thu thập rất nhiều dữ liệu và tư duy thống kê cung cấp các công cụ và kỹ thuật để phân tích dữ liệu này và xác định các mẫu hình và xu hướng. Các mẫu hình này có thể được sử dụng để đưa ra quyết định về mọi thứ, từ các chiến dịch marketing đến phát triển sản phẩm.
  + **Cải tiến liên tục và giám sát quyết định** (Continuous Improvement and Decision Monitoring): Doanh nghiệp cần thường xuyên đánh giá kết quả của các quyết định và sử dụng vòng lặp phản hồi để lặp lại và tinh chỉnh các chiến lược của mình dựa trên bằng chứng thực nghiệm.

## Ví dụ: Bài toán xác suất

* Xác suất (Probability) là một khái niệm thống kê quan trọng khác đối với nhà quản lý. Các quyết định của nhà quản lý mang tính xác suất. Do đó, luôn luôn có yếu tố không chắc chắn trong hầu hết các quyết định mà nhà quản lý đưa ra. Nếu nhà quản lý đưa ra quyết định chắc chắn dẫn đến kết quả cụ thể, thì bài toán đó được gọi là xác định (deterministic). Đối với nhà quản lý, việc ra quyết định giống như đang lựa chọn ngẫu nhiên. Họ không thể chắc chắn về kết quả của dự án của mình do một số yếu tố may rủi ro ảnh hưởng đến kỳ vọng của họ. T ất cả các quyết định của nhà quản lý, nằm giữa hoàn toàn không chắc chắn và hoàn toàn chắc chắn, đều có xác suất nằm trong khoảng từ 1 đến 0. Xác suất không thể lớn hơn 1 và không có xác suất âm. Nói cách khác, xác suất của bất kỳ sự kiện quản lý nào cũng nằm trong khoảng từ 0 đến 1 [3].
* Trong kinh doanh, các nhà quản lý đưa ra nhiều quyết định không hoàn toàn chắc chắn, và xác suất giúp họ hiểu mức độ có thể xảy ra của các kết quả khác nhau.
* **Định nghĩa**: Xác suất của một sự kiện được tính bằng cách chia số lượng kết quả thuận lợi (số cách sự kiện có thể xảy ra) cho tổng số các kết quả có thể (tất cả các kết quả có thể xảy ra trong một thí nghiệm). Tỷ lệ này được gọi là xác suất.
* Ví dụ:
  + Tung một đồng xu: Xác suất xuất hiện mặt sấp hoặc mặt ngửa đều là ½ (0.5).
  + Tung một xúc xắc: Xác suất đổ được mặt có 4 chấm là ⅙ (khoảng 0.17).
* Xác suất trong Quyết định Kinh doanh (Probability in Business Decisions):
* Khi nhà quản lý dự đoán doanh số bán hàng đạt 20.000 khối trong một tuần, họ đang đưa ra một tuyên bố mang tính xác suất. Dự đoán này dựa trên các sự kiện trong quá khứ. Kết quả thực tế có thể thấp hơn 20.000 hoặc cao hơn một chút. Về cơ bản, xác suất cung cấp cho nhà quản lý một thước đo thay thế cho sự chắc chắn tuyệt đối. Do ảnh hưởng của nhiều nhân tố lẫn trong và ngoài, chính vì vậy ta có xác suất điều kiện (Conditional Probability).
* Trong một số trường hợp, xác suất của một sự kiện có thể thay đổi tùy thuộc vào việc một sự kiện khác đã xảy ra hay chưa. Khái niệm này được gọi là xác suất điều kiện. Xác suất điều kiện ký hiệu là P(B|A), đọc là "xác suất của B cho biết A đã xảy ra" và được thể hiện như sau:

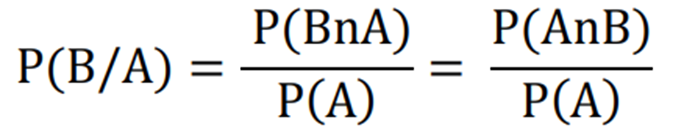


Figure 93. Xác suất của B, đã biết A

* Ở đây, chúng ta giả định rằng P(A) > 0. Trong khi đó, xác suất có điều kiện rằng A sẽ xảy ra khi đã xảy ra B, được viết là P(A/B) và được định nghĩa là:

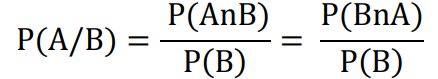


Figure 94. Xác suất của A, đã biết B

* Kế tiếp, chúng ta giả định rằng P(A) > 0:

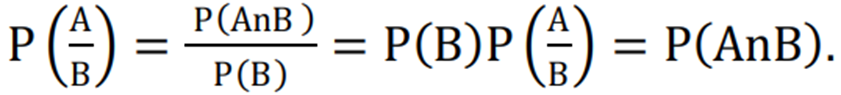


Figure 95. Quy tắc tích tổng quát

* Công thức nhân xác suất tổng quát là một công cụ quan trọng để tính xác suất của hai sự kiện A và B xảy ra cùng nhau, được ký hiệu là P(AB). Có hai cách để tính toán P(AB):
  + Cách 1: P(AB) = P(A) \* P(B|A)
  + Trong đó:
    - P(AB): Xác suất của cả hai sự kiện A và B xảy ra cùng nhau.
    - P(A): Xác suất của sự kiện A xảy ra.
    - P(B|A): Xác suất của sự kiện B xảy ra cho biết sự kiện A đã xảy ra.
  + Cách 2: P(AB) = P(B) \* P(A|B)
  + Trong đó:
    - P(AB): Xác suất của cả hai sự kiện A và B xảy ra cùng nhau.
    - P(B): Xác suất của sự kiện B xảy ra.
    - P(A|B): Xác suất của sự kiện A xảy ra cho biết sự kiện B đã xảy ra.
* Hai sự kiện A và B được gọi là độc lập khi và chỉ khi:
  + P(A|B) = P(A)
  + P(B|A) = P(B)
* Điều kiện (1) và (2) tương đương nhau. Nghĩa là nếu điều kiện (1) thỏa mãn thì điều kiện (2) cũng thỏa mãn, và ngược lại.
* **Minh họa**: Một công ty nghiên cứu thị trường đã tiến hành khảo sát về sở thích vị của nước ngọt. 1000 người tiêu dùng nước ngọt được yêu cầu chọn giữa hai loại nước ngọt khác nhau mà họ thích. Mỗi người cũng được hỏi họ thích nước ngọt ngọt hay rất ngọt. Thật không may, công ty nghiên cứu đã mất một số thông tin khảo sát do sự cố máy chủ xảy ra vào năm ngoái. Tuy nhiên, một số thông tin đã được phục hồi.
* Thông tin khảo sát có được:
  + 68% người tiêu dùng thích nước loại 1 hơn loại 2.
  + 60% người tiêu dùng thích nước ngọt (hơn rất ngọt).
  + 80% người tiêu dùng thích nước "ngọt” thì thích nước loại 1 hơn loại 2.
* **Thách thức**: Công ty hiện đang gặp thách thức khôi phục các thông tin bị thiếu hụt khác. Để phục hồi tất cả thông tin khảo sát bị mất, hãy cân nhắc chọn ngẫu nhiên một trong 1000 người tham gia khảo sát và định nghĩa các sự kiện sau:
  + T1: Người được chọn ngẫu nhiên thích nước loại 1.
  + T2: Người được chọn ngẫu nhiên thích nước loại 2.
  + S: Người được chọn ngẫu nhiên thích nước ngọt.
  + V: Người được chọn ngẫu nhiên thích nước rất ngọt.

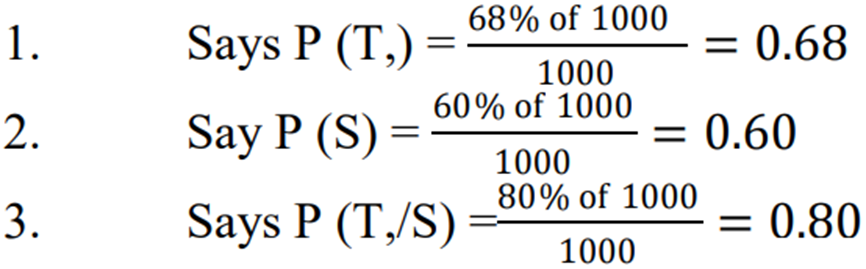


Figure 96. Xác suất từng phần

* Mục tiêu: Khôi phục tất cả thông tin khảo sát bị mất.
* Phương pháp: Tìm P(T1 \* S), xác suất người được chọn ngẫu nhiên thích nước loại 1 (T1) và thích nước ngọt (S).
* Áp dụng công thức nhân xác suất tổng quát:

P(T1 \* S) = P(T1) \* P(S/T1) = P(S) \* P(T1/S)

* Như thông tin đã biết, P(S) = 0.6 (60% thích ngọt) và P(T1/S) = 0.8 (80% thích ngọt chọn loại 1). Do đó:

P(T1 \* S) = (0.6) \* (0.8) = 0.48.

* Diễn giải: 0.48 trong tổng số 1000 người tiêu dùng thích nước loại 1 và thích nước ngọt ngọt. Nghĩa là có 480 người chọn loại 1 ngọt.
* Tổng hợp thông tin: Biết rằng 68% (680 người) thích loại 1 (T1) và 480 người thích loại 1 ngọt (T1 \* S). Từ đó, ta suy ra:
  + Số người thích loại 2 (T2) = 1000 - 680 = 320 người (vì chỉ có 2 loại nước).
  + Số người chỉ thích loại 1 ngọt (T1 \* S) = 480 người (đã tính toán ở trên).
  + Số người thích ngọt nhưng không nhất thiết là loại 1 (S) = 600 người (đã biết từ đầu).
  + Số người thích rất ngọt (V) = 600 người (tổng số thích ngọt) - 480 người (thích loại 1 ngọt) = 120 người.
* Bằng cách tính toán xác suất P(T1 \* S) và áp dụng logic, chúng ta đã phục hồi được thông tin cho tất cả các sự kiện T1, T2, S, V và T.

**CÁC TÀI NGUYÊN KHÁC**

|  |  |
| --- | --- |
| Tên tài nguyên | Đường dẫn |
| Mã nguồn | [Ấn vào đây](https://uithcm-my.sharepoint.com/:f:/g/personal/21522541_ms_uit_edu_vn/Ei-0B7hlf_BMowV5Z9ICg-0BKZmme9vpYboFjJ8oWBDClg?e=pnHhHh) |

**DANH MỤC TÀI LIỆU THAM KHẢO**

|  |  |
| --- | --- |
| [1] | J. C. d. Toledo, F. L. Lizarelli, A. B. d. Santos and A. Ishizaka, "Statistical Thinking in quality improvement: use, difficulties and benefits of its implantation in industries of the Brazilian State of São Paulo," pp. 14: 2-3, 2018. |
| [2] | M. Imran, Business Analytics, D. S. Kashyap, Ed., Jalandhar Delhi GT road, Phagwara - 144411: Lovely Professional University. |
| [3] | M. Odoh and C. Ihedigbo, "The Application of Statistics to the Different Areas of Business," *IOSR Journal of Business and Management (IOSR-JBM),* vol. XVI, no. 11, pp. 48-49, 2014. |
| [4] | K. H. t. T. t. tin. |

**­­­­**