**ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH**

**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**

**KHOA HỆ THỐNG THÔNG TIN**

-----&-----



**BÁO CÁO LAB 1**

**Môn: Phân tích dữ liệu kinh doanh**

Thành viên nhóm:

Đoàn Thục Quyên - 18521320

Nguyễn Thành Trung - 18521554

Nguyễn Đức Anh - 18520454

***Thành viên nhóm:***

|  |  |  |
| --- | --- | --- |
| Số lượng thành viên | MSSV | *Họ và tên* |
| 1 | 18521320 | *Đoàn Thục Quyên* |
| 2 | 18521554 | *Nguyễn Thành Trung* |
| 3 | 18520454 | *Nguyễn Đức Anh* |

***Phân công công việc:***

|  |  |  |  |
| --- | --- | --- | --- |
|  | Đoàn Thục Quyên  18521320 | Nguyễn Thành Trung  18521554 | Nguyễn Đức Anh  18520454 |
| Tìm hiểu ngôn ngữ R | x | x | x |
| Tìm cách import file csv, excel | x | x | x |
| Tập dữ liệu:  *GDP binh quan cua VN* | | | |
| Count, Min, Max, Mean, Median, Mode, Quantile, Range, Variance, Standard Deviation, Coefficient of Deviation, Skewness, Kurtosis. | Count, Min, Max, Mean, Median, Mode, Quantile, Range | Variance, Standard Deviation, Coefficient of Deviation, Skewness, Kurtosis. |  |
| Dùng Histogram để hiện các giá trị |  | x |  |
| Dùng Box Plot để hiện thị các giá trị | x |  |  |
| Kiểm tra |  |  | x |
| Tập dữ liệu:  *Purchase Orders* | | | |
| Count, Min, Max, Mean, Median, Mode, Quantile, Range, Mode, Variance, Standard Deviation, Coefficient of Deviation, Skewness, Kurtosis. |  | Count, Min, Max, Mean, Median, Mode, Quantile, Range | Variance, Standard Deviation, Coefficient of Deviation, Skewness, Kurtosis. |
| Dùng Histogram để hiện các giá trị | x |  |  |
| Dùng Box Plot để hiện thị các giá trị |  | x |  |
| Kiểm tra |  |  | x |
| Tập dữ liệu:  *Computer Repair Times* | | | |
| Count, Min, Max, Mean, Median, Mode, Quantile, Range, Mode, Variance, Standard Deviation, Coefficient of Deviation, Skewness, Kurtosis. | Variance, Standard Deviation, Coefficient of Deviation, Skewness, Kurtosis. |  | Count, Min, Max, Mean, Median, Mode, Quantile, Range |
| Dùng Histogram để hiện các giá trị |  |  | x |
| Dùng Box Plot để hiện thị các giá trị |  | x |  |
| Kiểm tra | x |  |  |
| Câu c | | | |
| Sự khác nhau | x |  |  |
| Cho ví dụ minh họa |  | x | x |
| Kiểm tra | x |  |  |
| Tổng hợp làm file báo cáo | x |  |  |

***CÂU 1:***

# Đề bài:

Với tập tin dữ liệu:

*GDP binh quan cua VN*

|  |  |
| --- | --- |
| Năm | GDP |
| 2000 | 302 |
| 2001 | 315 |
| 2002 | 340 |
| 2003 | 389 |
| 2004 | 453 |
| 2005 | 537 |
| 2006 | 624 |
| 2007 | 730 |
| 2008 | 1065 |
| 2009 | 1132 |
| 2010 | 1234 |
| 2011 | 1443 |
| 2012 | 1655 |
| 2013 | 1811 |
| 2014 | 1916 |
| 2015 | 2109 |
| 2016 | 2215 |

1. Sử dụng ***MS Excel và ngôn ngữ R*** với số liệu về GDP bình quân của VN, hãy tính và giải thích ý nghĩa của các giá trị: Count, Min, Max, Mean, Median, Mode, Quantile, Range, Mode, Variance, Standard Deviation, Coefficient of Deviation, Skewness, Kurtosis.

2. Dùng chức năng Visualization: Histogram, Box Plot hãy hiển thị một số giá trị trên

# Bài làm

# Sử dụng *MS Excel*

# Hàm Count:

* **Mô tả**

Hàm COUNT đếm số lượng ô chứa số và đếm số trong danh sách các đối số. Sử dụng hàm COUNT để lấy số lượng mục trong trường số nằm trong một phạm vi hoặc mảng số.

* **Cú pháp**

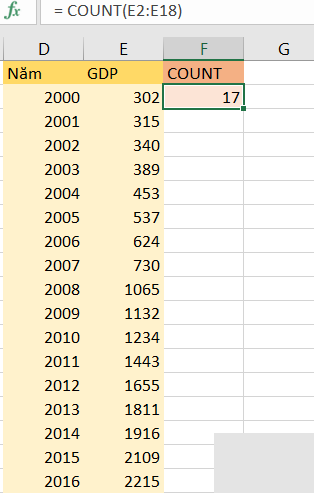
COUNT (value1, [value2], ...)

Cú pháp hàm COUNT có các đối số sau đây:

**value1**  Bắt buộc. Mục đầu tiên, tham chiếu ô hoặc phạm vi trong đó bạn muốn đếm số.

**value2**, ... Tùy chọn. Tối đa 255 mục, tham chiếu ô hoặc phạm vi bổ sung trong đó bạn muốn đếm số.

* **Kết quả**



# Hàm Min:

* **Mô tả**

Hàm MIN trả về số nhỏ nhất trong một tập hợp các giá trị.

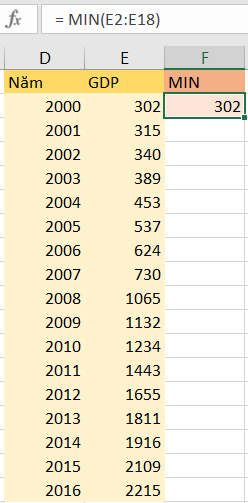
* **Cú pháp trong excel**

MIN (number1, [number2], ...)

MIN(number1, [number2], ...)

Cú pháp hàm MIN có các đối số sau đây:

**Number1, number2**, ... Number1 là tùy chọn, các số sau đó là tùy chọn. 1 tới 255 số mà bạn muốn tìm giá trị nhỏ nhất trong đó.

* **Kết quả**
* 

# Hàm Max:

* **Mô tả**

Hàm Max trả về giá trị lớn nhất trong một tập hợp các giá trị.

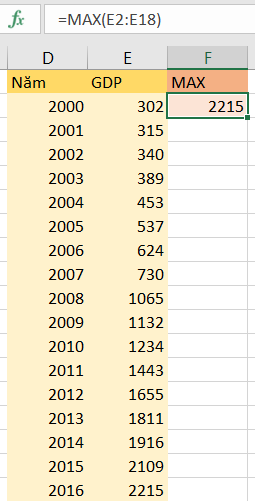
* **Cú pháp trong excel**

MAX (number1, [number2], ...)

Cú pháp của hàm MAX có các đối số sau đây:

**Number1, number2**, ... Number1 là bắt buộc, các số tiếp theo là tùy chọn. 1 tới 255 số mà bạn muốn tìm giá trị lớn nhất trong đó.

* **Kết quả**



# Hàm Mean:

* **Mô tả**

Trong ***MS Excel*** hàm AVERAGEgiống như hàm Mean. Hàm AVERAGE trả về giá trị trung bình (trung bình số học) của các đối số.

* **Cú pháp trong excel**

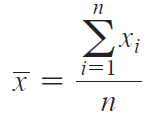
AVERAGE (number1, [number2], ...)

The AVERAGE function syntax has the following arguments:

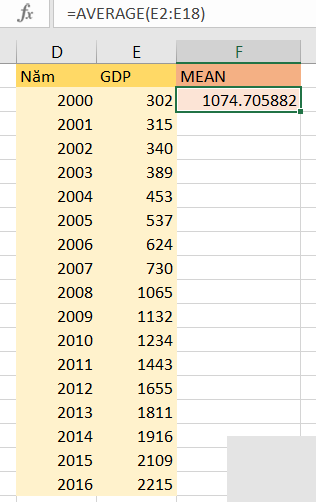
**Number1**    Required. The first number, cell reference, or range for which you want the average.

**Number2, ...**    Optional. Additional numbers, cell references or ranges for which you want the average, up to a maximum of 255.

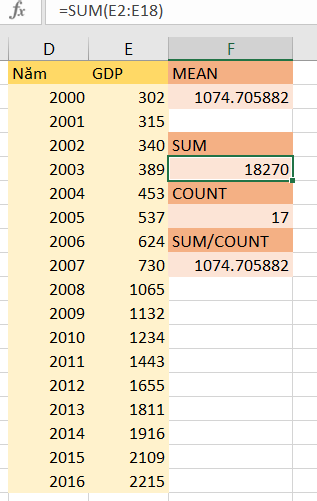
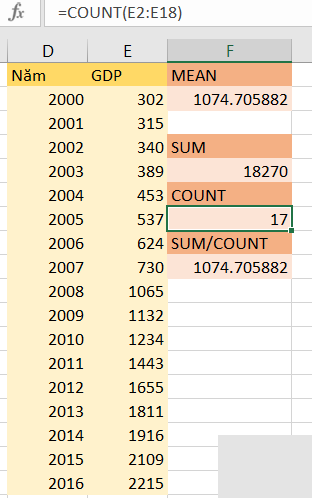
* **Công thức**

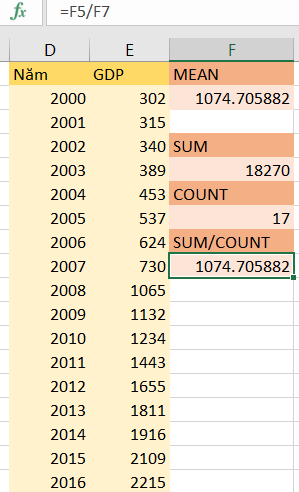


* **Kết quả**



* **Chứng minh**



# Hàm Median

* **Mô tả**

Hàm MEDIAN trả về trung vị của các số đã cho. Trung vị là số ở giữa một tập hợp số. Nếu bộ số có số lượng số chẵn, thì hàm MEDIAN tính toán trung bình của hai số ở giữa. Hãy xem công thức thứ hai trong ví dụ.

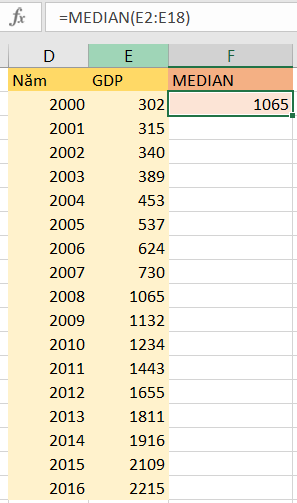
* **Cú pháp**

MEDIAN (number1, [number2], ...)

Cú pháp hàm MEDIAN có các đối số dưới đây:

Number1, number2, ... Number1 là bắt buộc, các số tiếp theo là tùy chọn. 1 tới 255 số mà bạn muốn tìm trung vị.

* **Kết quả**



Cố 17 dòng dữ liệu (số dòng lẻ) nên trung vị là dòng thứ 9.

# Hàm Mode

* **Mô tả**

Hàm MODE trả về giá trị xảy ra thường xuyên nhất hoặc lặp đi lặp lại trong một mảng hoặc phạm vi dữ liệu

* **Cú pháp**

MODE (number1, [number2], ...)

Cú pháp hàm MODE có các đối số dưới đây:

**Number1**     Bắt buộc. Đối số dạng số đầu tiên cho những gì bạn muốn tính số yếu vị.

**Number2, ...** Tùy chọn. Các đối số dạng số từ 2 tới 255 mà bạn muốn tính toán số yếu vị trong đó. Bạn cũng có thể sử dụng một mảng đơn hay tham chiếu tới một mảng thay thế cho các đối số được phân tách bởi dấu phẩy.

* **Kết quả**



Ta có thể thấy các giá trị GDP qua các năm không giống nhau ( Không có giá trị trùng). Vì vậy không có giá trị lặp đi lặp lại. =>giá trị Mode là #N/A

# Hàm Quartile

* **Mô tả**

Hàm QUARTILE trả về tứ phân vị của tập dữ liệu. Tứ phân vị được dùng trong dữ liệu khảo sát và bán hàng để chia tập hợp thành các nhóm. Ví dụ, bạn có thể dùng hàm QUARTILE để tìm ra 25% số người có thu nhập cao nhất trong một tập hợp dân cư.

* **Cú pháp**

QUARTILE (array, quart)

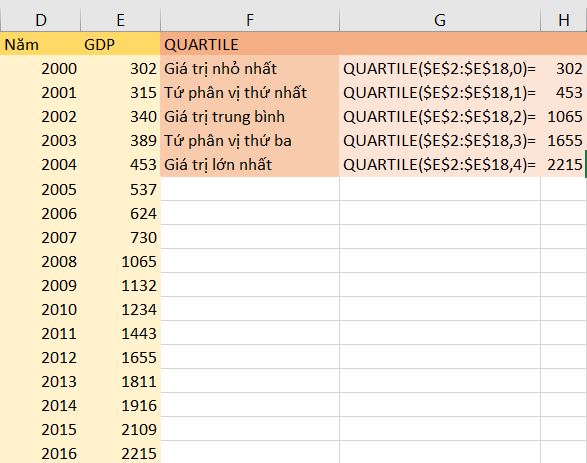
Cú pháp hàm QUARTILE có các đối số sau đây:

**Array**     Bắt buộc. Mảng hoặc phạm vi ô có chứa các giá trị số mà bạn muốn tìm giá trị tứ phân vị.

**Quart**     Bắt buộc. Chỉ rõ giá trị nào cần trả về.

| **Nếu quãng bốn bằng** | **Hàm QUARTILE trả về** |
| --- | --- |
| 0 | Giá trị tối thiểu |
| 1 | Tứ phân vị thứ nhất (phân vị thứ 25) |
| 2 | Giá trị trung vị (phân vị thứ 50) |
| 3 | Tứ phân vị thứ ba (phân vị thứ 75) |
| 4 | Giá trị tối đa |

* **Kết quả:**

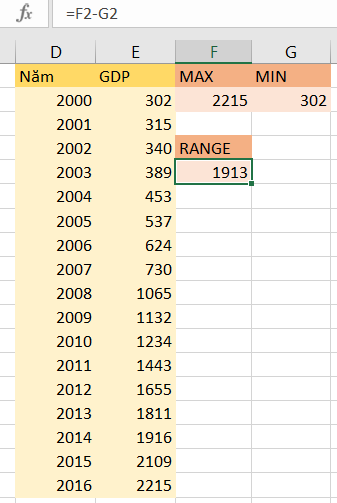


# Hàm Range

* **Mô tả :** Khoảng biến thiên (Range): Là khoảng cách giữa giá trị quan sát nhỏ nhất đến giá trị quan sát lớn nhất
* **Công thức:**



* **Kết quả:**



# Hàm Variance

* **Mô tả**

Phương sai trong thống kê là phép đo mức chênh lệch giữa các số trong một tập dữ liệu. Nghĩa là, nó đo khoảng cách mỗi số trong tập hợp từ giá trị trung bình và do đó từ mọi số khác trong tập hợp.

* **Cú pháp**

VAR (number1, [number2], ...)

Cú pháp của hàm VAR có các đối số như sau:

**Number1**     Bắt buộc. Đối số dạng số đầu tiên tương ứng với mẫu tổng thể.

**Number2, ...**     Tùy chọn. Là các đối số dạng số từ 2 đến 255 tương ứng với một mẫu của một tập hợp.

**Công thức** Công thức

* **Trên toàn bộ tập hợp**

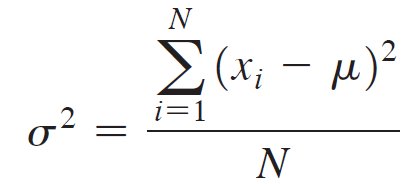
VAR.P(number1, [number2], ...)

Cú pháp hàm VAR.P có các đối số dưới đây:

**Number1**     Bắt buộc. Đối số dạng số đầu tiên tương ứng với tổng thể.

**Number2, ...**     Tùy chọn. Là các đối số dạng số từ 2 đến 254 tương ứng với một tập hợp.

**Công thức:**



* **Trên một mẫu tập hợp**

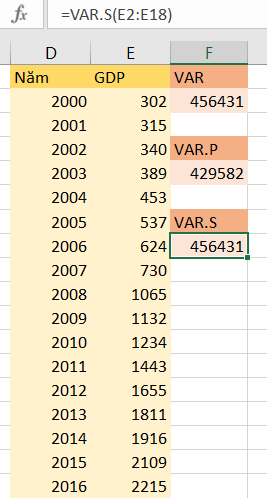
VAR.S(number1, [number2], ...)

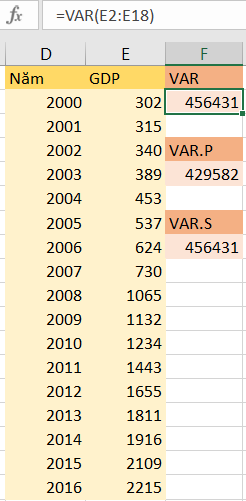
Cú pháp hàm VAR.S có các đối số như sau:

**Number1**     Bắt buộc. Đối số dạng số đầu tiên tương ứng với mẫu tổng thể.

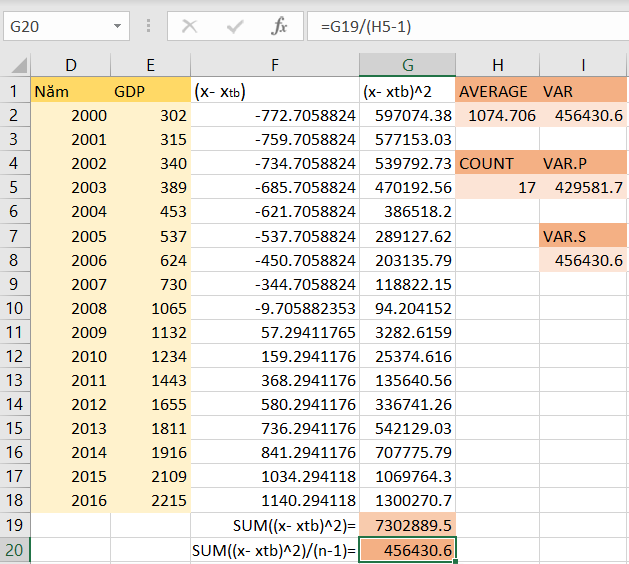
**Number2, ...**     Tùy chọn. Là các đối số dạng số từ 2 đến 254 tương ứng với một mẫu của một tập hợp.

**Công thức** Công thức

* **Kết quả**



* **Chứng minh:**



# Standard Deviation

* **Mô tả**

Khoảng tin cậy trả về khoảng tin cậy của trung bình tổng thể, bằng cách dùng phân bố chuẩn hóa.

* **Cú Pháp**
  + STDEV

Ước tính độ lệch chuẩn dựa trên toàn bộ tổng thể

|  |
| --- |
| STDEV(number1,[number2],...)  Cú pháp hàm STDEV có các đối số sau đây:  **Number1** Bắt buộc. Đối số dạng số đầu tiên tương ứng với mẫu tổng thể.  **Number2**, ... Tùy chọn. Đối số dạng số từ 2 đến 255 tương ứng với mẫu tổng thể. Bạn cũng có thể sử dụng một mảng đơn hay tham chiếu tới một mảng thay thế cho các đối số được phân tách bởi dấu phẩy.  **Công thức:**  Công thức |

* STDEV.S

Ước tính độ lệch chuẩn dựa trên mẫu

STDEV.S(number1,[number2],...)

Cú pháp hàm STDEV.S có các đối số sau đây:

**Number1**     Bắt buộc. Đối số dạng số đầu tiên tương ứng với mẫu tổng thể. Bạn cũng có thể sử dụng một mảng đơn hay tham chiếu tới một mảng thay thế cho các đối số được phân tách bởi dấu phẩy.

**Number2**, ...     Tùy chọn. Đối số dạng số từ 2 đến 254 tương ứng với mẫu tổng thể. Bạn cũng có thể sử dụng một mảng đơn hay tham chiếu tới một mảng thay thế cho các đối số được phân tách bởi dấu phẩy.

**Công thức:**

Công thức

* STDEV.P

Ước tính độ lệch chuẩn dựa trên toàn bộ tổng thể

STDEV.P(number1,[number2],...)

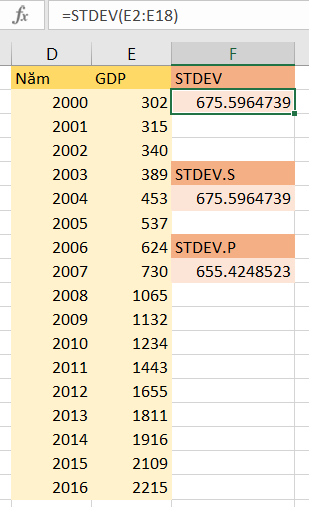
Cú pháp hàm STDEV.P có các đối số sau đây:

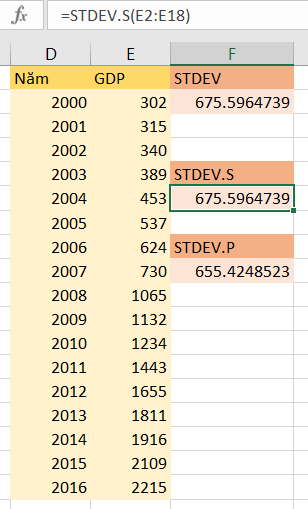
**Number1**    Bắt buộc. Đối số dạng số đầu tiên tương ứng với tổng thể.

**Number2**, ...     Tùy chọn. Đối số dạng số từ 2 đến 254 tương ứng với tổng thể. Bạn cũng có thể sử dụng một mảng đơn hay tham chiếu tới một mảng thay thế cho các đối số được phân tách bởi dấu phẩy.

**Công thức:**

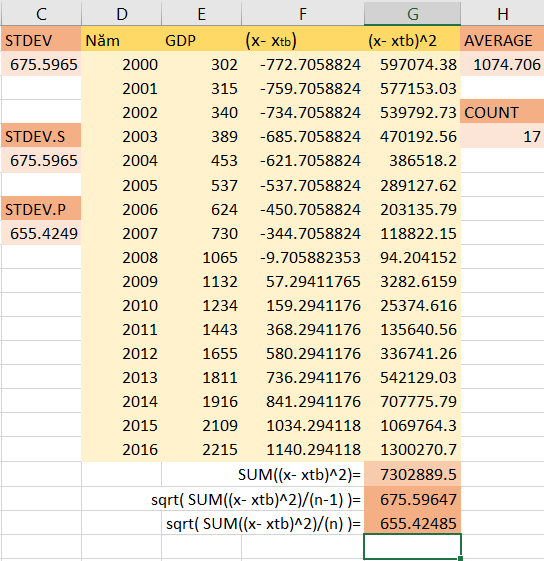
Công thức

* **Kết Quả**





* **Chứng minh**



# Skewness

* **Mô tả**

***Hệ số đối xứng Skewness:*** cho biết dạng phân phối của các giá trị quan sát có thể được sử dụng để kiểm nghiệm tính phân phối chuẩn. Một giá trị Skewness dương cho biết nhánh của phân phối này lệch qua bên phải và ngược lại một trị âm chỉ ra nhánh của phân phối này dài qua bên trái

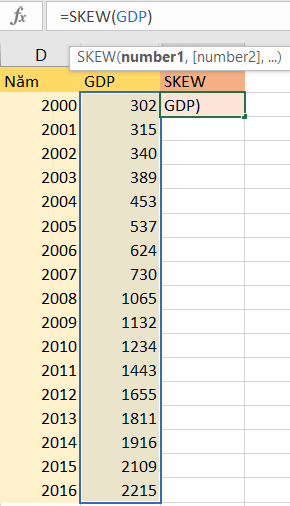
* ***Skewness*** = 0: Các quan sát được phân phối một các đối xứng xung quanh giá trị trung bình
* ***Skewness*** < 0: Các quan sát tập trung chủ yếu vào các giá trị nhỏ nhất
* ***Skewness*** > 0: Các quan sát tập trung chủ yếu vào các giá trị lớn nhất
* **Công thức**



* **Cú pháp**

|  |
| --- |
| SKEW(number1, [number2], ...)  Cú pháp hàm SKEW có các đối số sau đây:  **Number1, number2**, ... Number1 là bắt buộc, các số tiếp theo là tùy chọn. 1 đến 255 đối số mà bạn muốn tính độ lệch cho đối số đó. Bạn cũng có thể sử dụng một mảng đơn hay tham chiếu tới một mảng thay thế cho các đối số được phân tách bởi dấu phẩy. |

* **Kết quả**



# Kurtosis

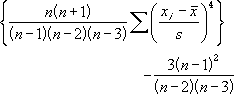
* **Mô tả**

Trả về hệ số nhọn của tập dữ liệu. Hệ số nhọn biểu thị đặc điểm nhọn hoặc phẳng tương ứng của một phân bố so với phân bố thông thường. Hệ số nhọn dương chỉ báo một phân bố tương đối nhọn. Hệ số nhọn âm chỉ báo một phân bố tương đối phẳng.

* **Cú pháp**

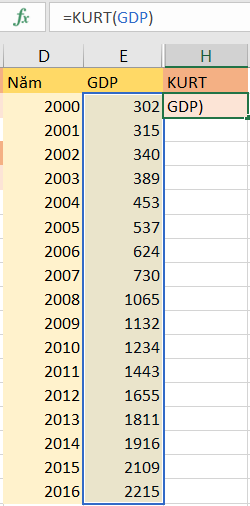
|  |
| --- |
| KURT(number1, [number2], ...)  Cú pháp hàm KURT có các đối số sau đây:  **Number1, number2**, ... Number1 là bắt buộc, các số tiếp theo là tùy chọn. 1 đến 255 đối số mà bạn muốn tính toán hệ số nhọn. Bạn cũng có thể sử dụng một mảng đơn hay tham chiếu tới một mảng thay thế cho các đối số được phân tách bởi dấu phẩy. |

* **Công thức**



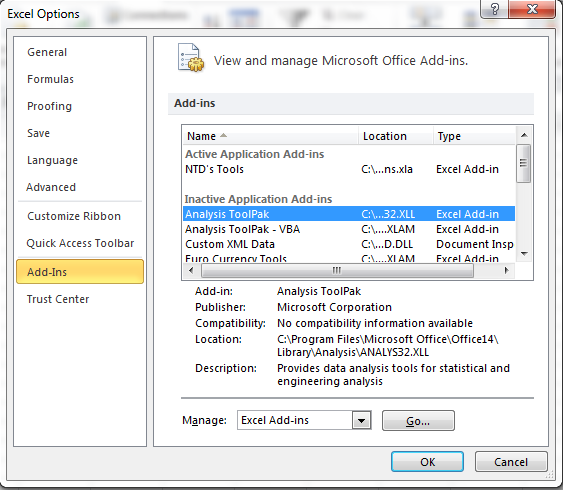
trong đó s là độ lệch chuẩn mẫu.

* **Kết quả**

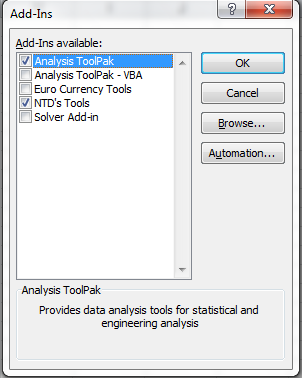


* **Kích hoạt Data Analysis trong Excel**

Bước 1: Chọn **“File”**=> chọn **“Options”**

Bước 2: Vào mục**“Add-Ins**” => chọn **“Analysis ToolPak”** sau đó bấm vào**“Go”**[](http://tinhocmos.edu.vn/wp-content/uploads/2016/10/nhung-dieu-can-biet-ve-cong-cu-data-analysis.png)

Bước 3: Sau khi bấm **“Go”** thì sẽ hiện ra 1 giao diện cửa sổ **“Add-Ins”.** Ban chọn **“Analysis ToolPak”**  rồi bấm **“Ok”**

[](http://tinhocmos.edu.vn/wp-content/uploads/2016/10/nhung-dieu-can-biet-ve-cong-cu-data-analysis1.png)

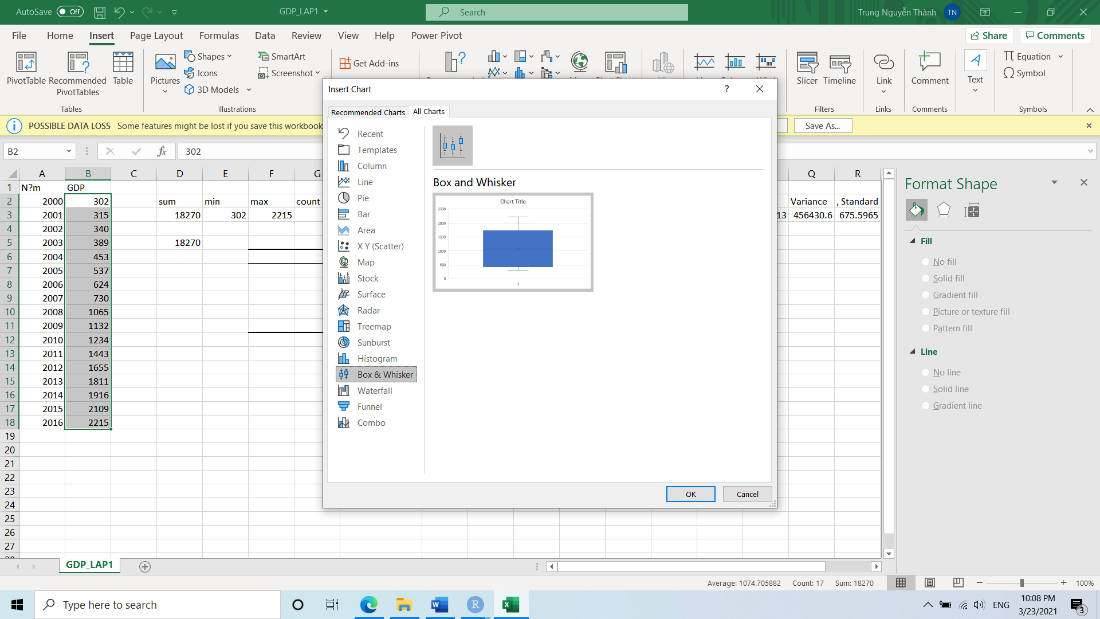
* **Box Plot**

**Box Plot trong MS Excel**

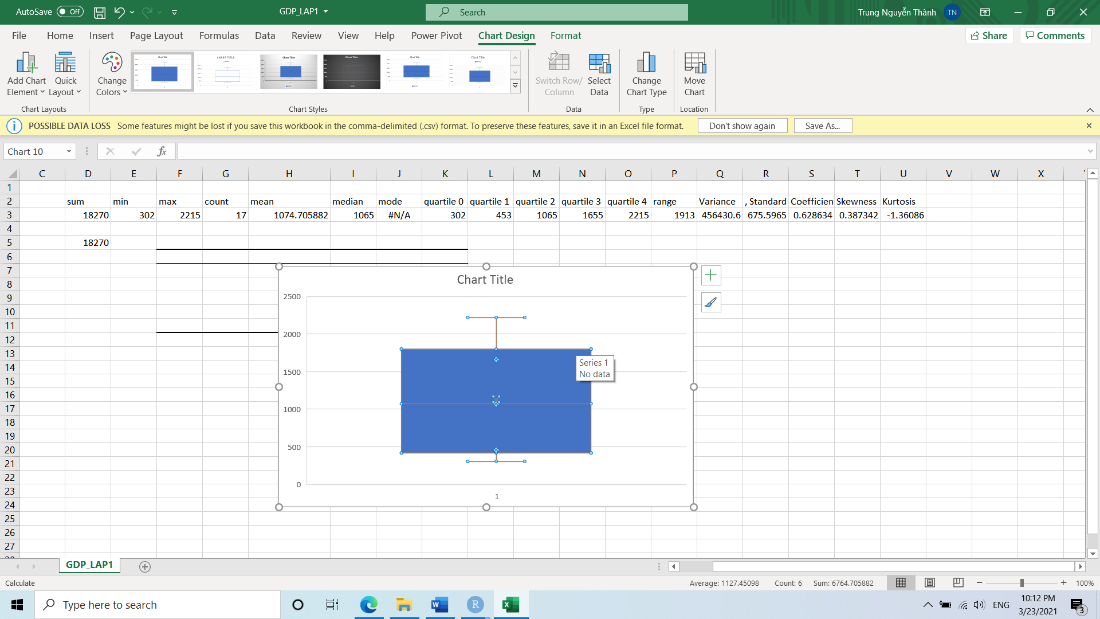
* **Các bước thực hiện**
* Tính các giá trị trước khi vẽ biểu đồ

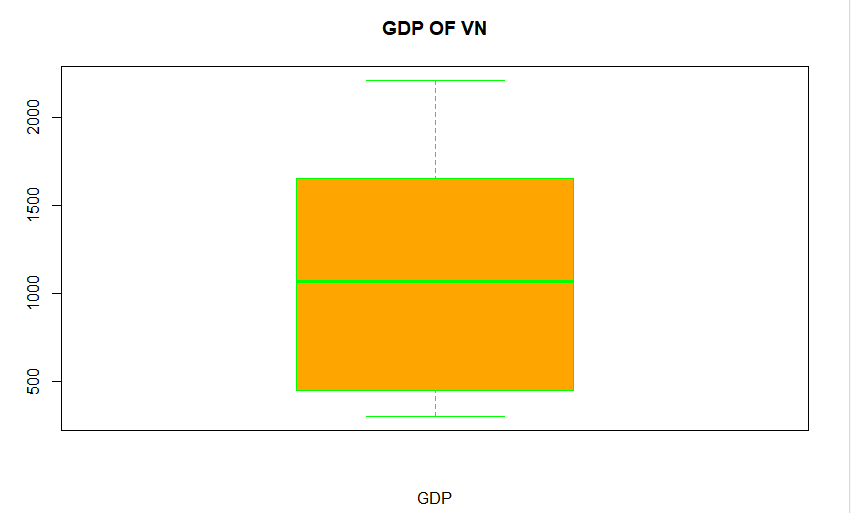
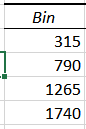
|  |  |  |
| --- | --- | --- |
| **Tên hàm** | **Thao tác** | **Giá trị** |
| Min | =Min (B2:B18) | 302 |
| Max | =Max (B2:B18) | 2215 |
| Mean | =Average (B2:B18) | 1074.706 |
| Median | =Median (B2:B18) | 1065 |
| Quartile 1 | = Quartile (B2:B18, 1) | 453 |
| Quartile 3 | = Quartile (B2:B18, 3) | 1655 |

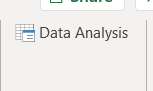
* Chọn tab Chèn trong MS Excel -> Chọn Box Plot (Hộp và dải dữ liệu trung bình)



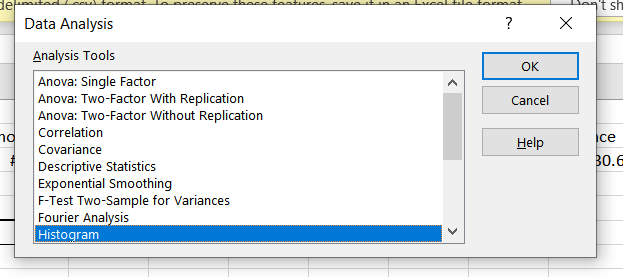
* **Bảng kết quả:**



* **Box Plot trong ngôn ngữ R**
* 
* **Histogram**
* **Histogram trong MS Excel**
* Trước tiên xác định các giá trị Bin
* 
* Trong MS Excel, vào tab Data chọn Data Analysis



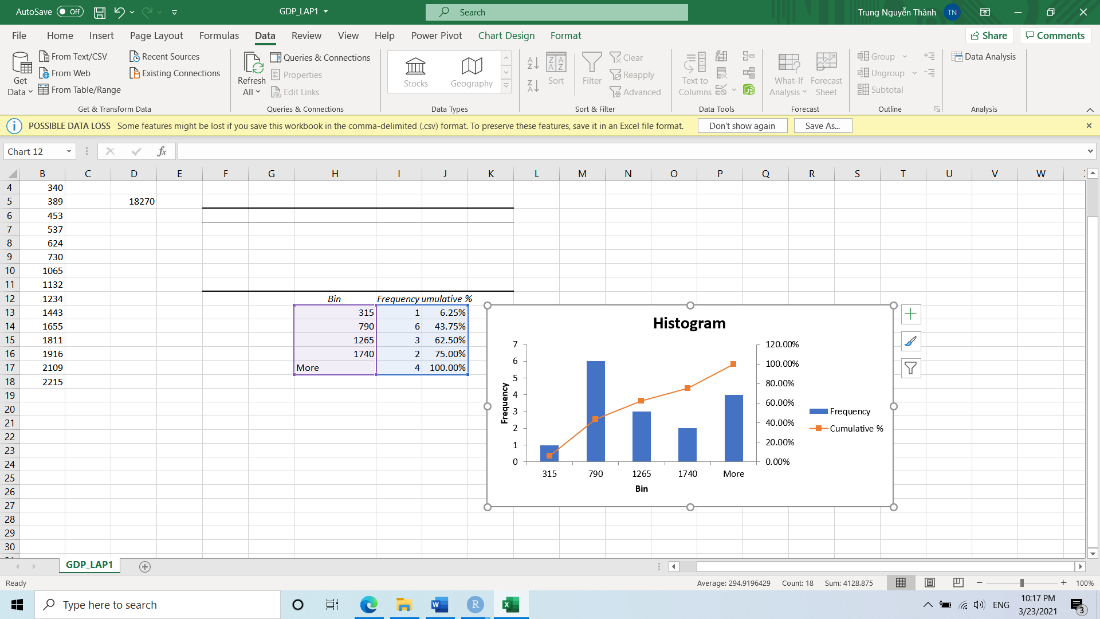
* Trong tab Data Analysis -> chọn Histogram -> chọn OK



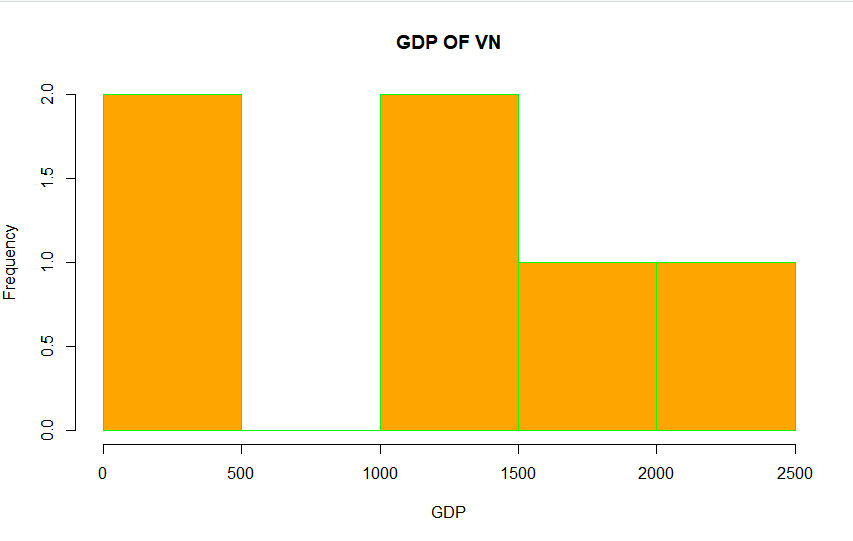
* Nhập vào các trường giá trị:

+ Input range: $B$2:$B$18

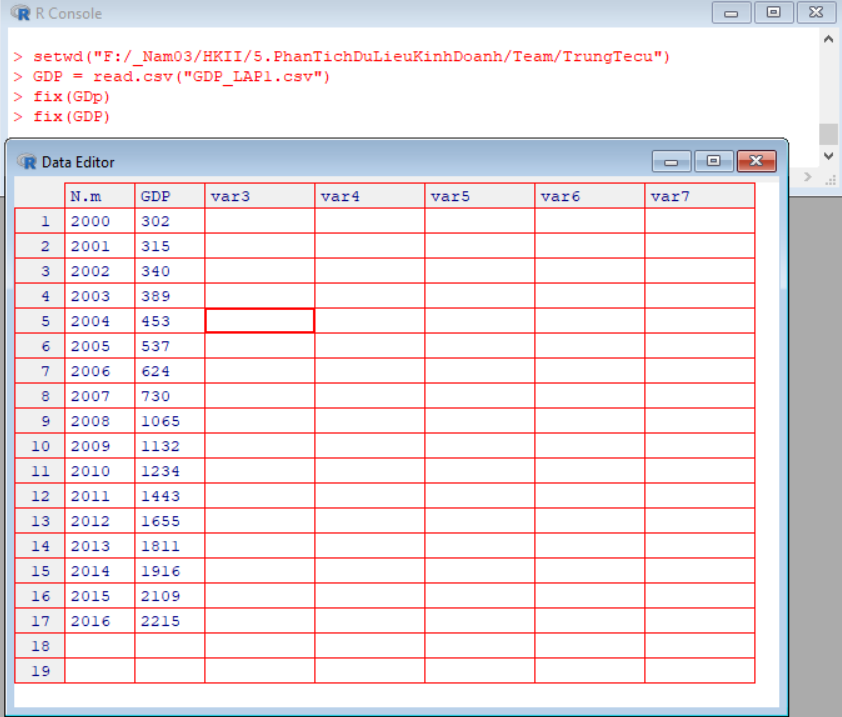
+ Bin range: $F$7:$F$10



* **Histogram trong ngôn ngữ R**



# II.Sử dụng *R*

* **Import dữ liệu từ Excel sang R** 
* **Hàm Max**



* **Hàm Min**



* **Hàm Median**



* **Hàm Mean**



* **Hàm Variance**



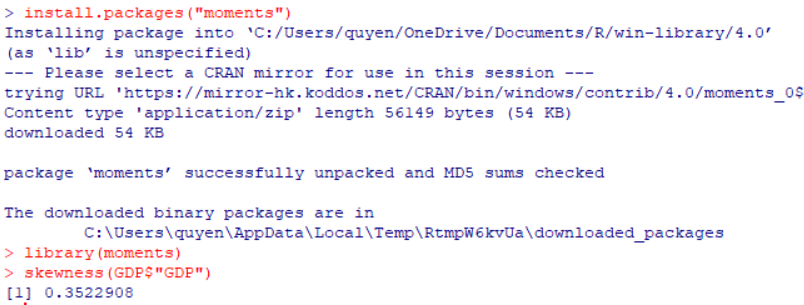
* **Hàm Range**



* **Hàm Standard Deviation**



* **Hàm Skewness**



* **Hàm Kurtosis**



* **Hàm Quantile**



# Bài 1b:

* **Purchase Orders**

**Phát biểu bài toán:**

Tập tin dữ liệu cho ta biết được thông tin về người cung cấp, mã số order, giá cả của mặt hàng mà ta cần để ta thống kê thông tin

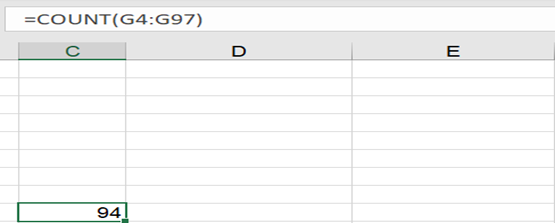
Với tập tin dữ liệu: **Purchase Orders**

Dùng công cụ trực quan hóa dữ liệu (Data Visualization) và Thống kê mô tả (Descriptive Statistics) thực hiện yêu cầu sau:

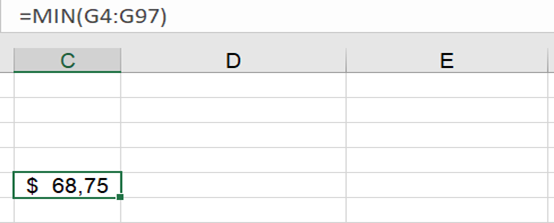
**Câu 1:**

**Sử dụng MS Excel, SPSS và ngôn ngữ R với số liệu về GDP bình quân của VN từ 2010 đến 2016 hãy tính và giải thích ý nghĩa của các giá trị: Count, Min, Max, Mean, Median, Mode, Quantile, Range, Mode, Variance, Standard Deviation, Coefficient of Deviation, Skewness, Kurtosis.**

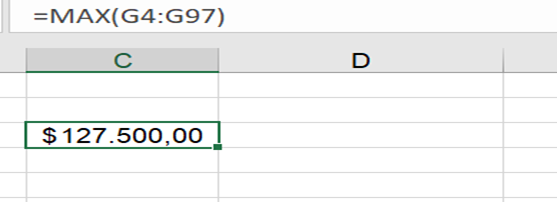
* **MS Excel**
* **Hàm Count**



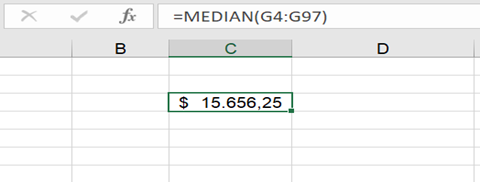
* **Hàm Min**



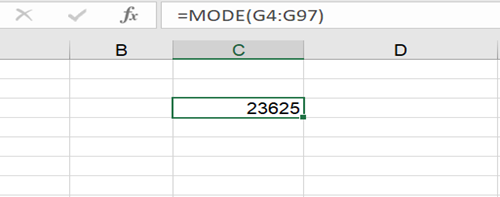
* **Hàm Max**



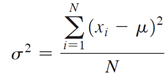
* **Hàm Median**



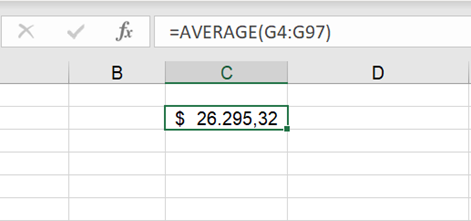
* **Hàm Mode**

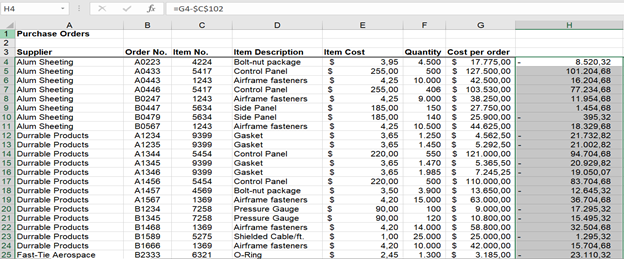


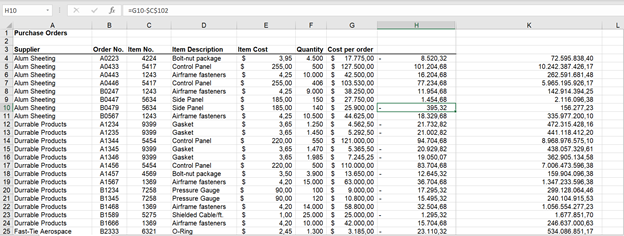
* **Hàm Variance**
* **Công thức**

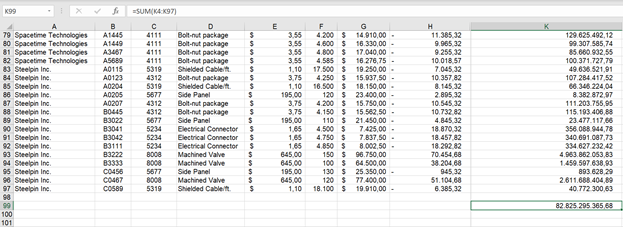


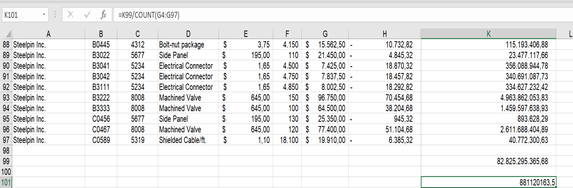
* **Chứng minh công thức**



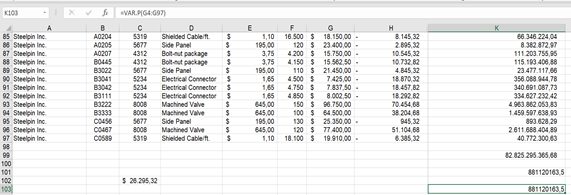




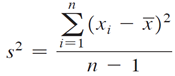




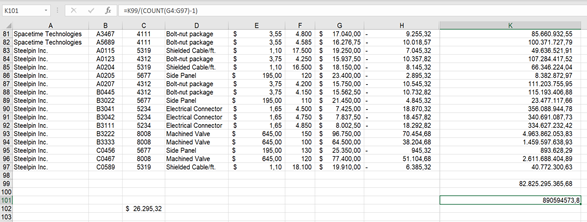
* **So sánh kết quả với hàm trong excel**

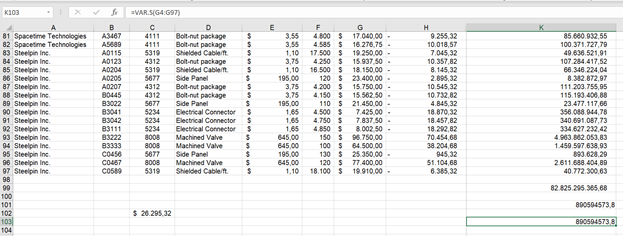


* **Công thức (For a sample)**

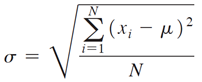


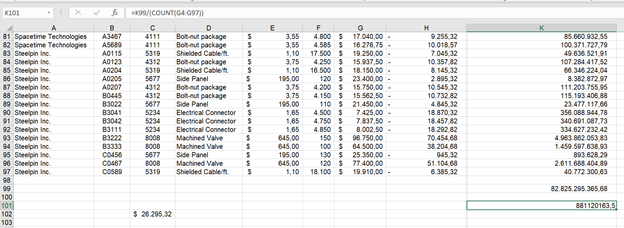
***(Các bước khác tiến hành như công thức (For a population) như trên.)***

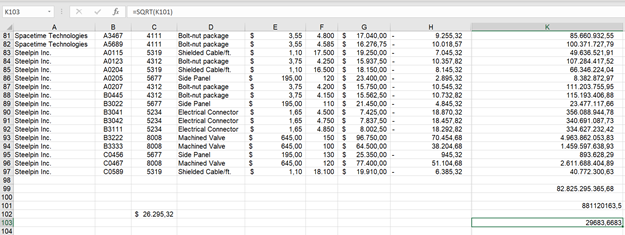




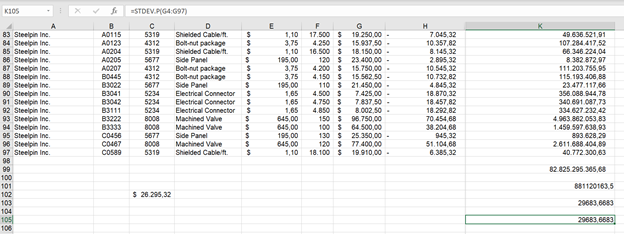
* **Hàm Standard Deviation**
* **Công thức (For a population)**



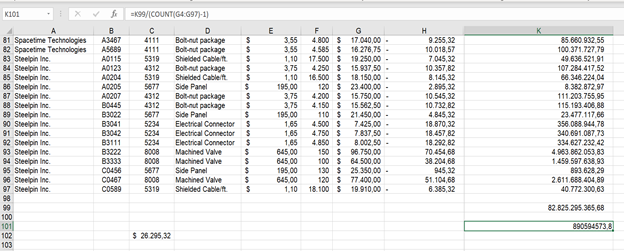


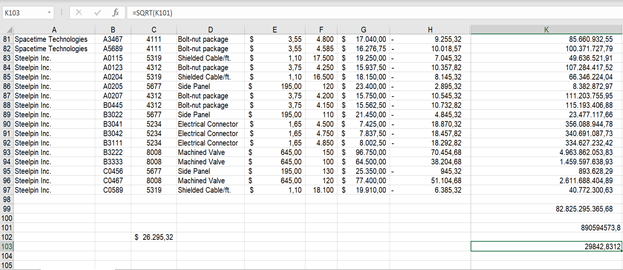


* **So sánh kết quả với hàm trong excel**

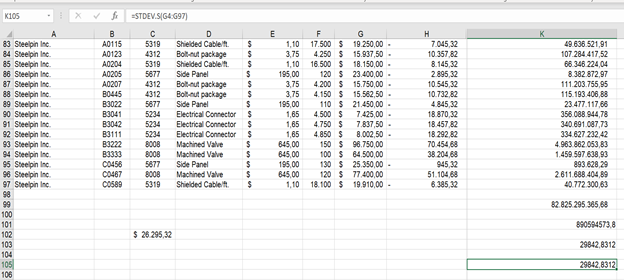


* **Công thức (For a sample)**



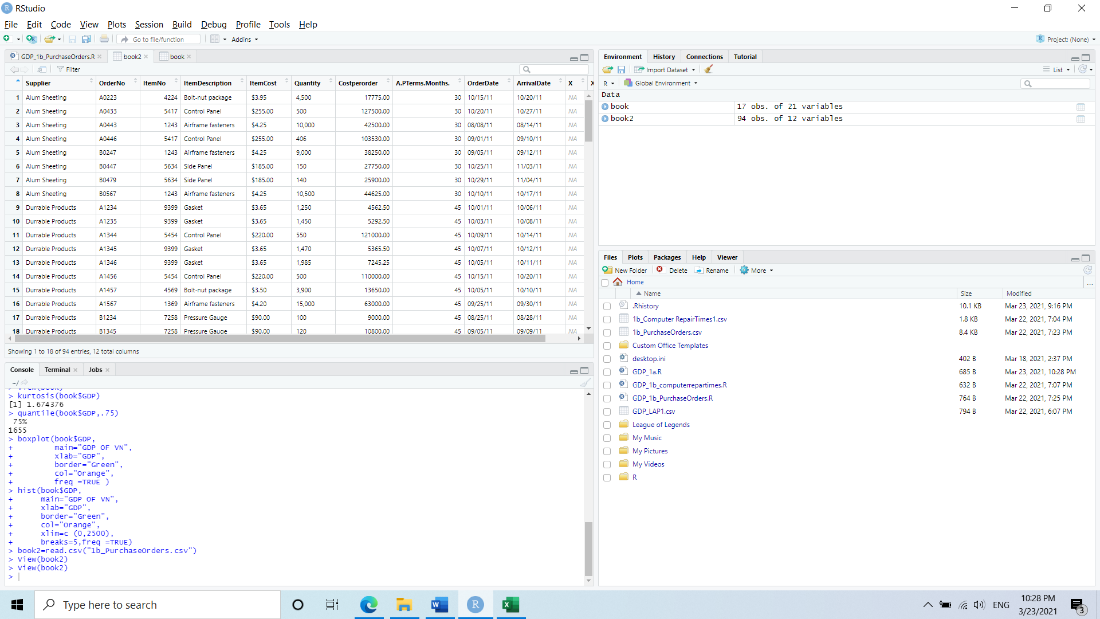


* **So sánh kết quả với hàm trong excel**

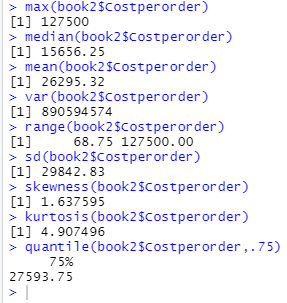


**Ngôn ngữ R**

**Import dữ liệu từ excel sang R**



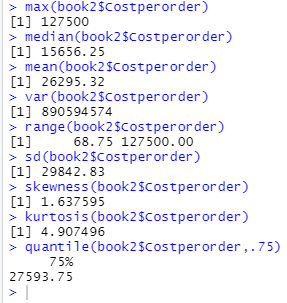
* **Hàm Max**



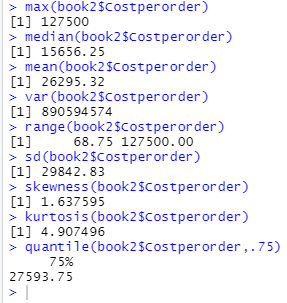
* **Hàm Min**

Đang chèn hình ảnh...

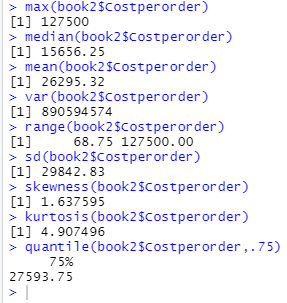
* **Hàm Median**



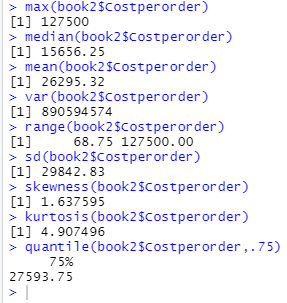
* **Hàm Mean**



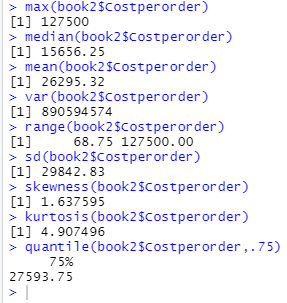
* **Hàm Variance**



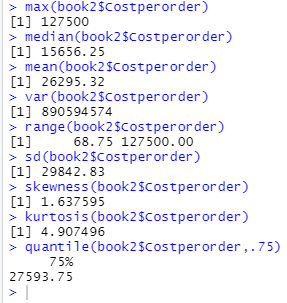
* **Hàm Range**



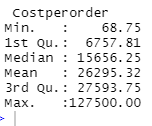
* **Hàm Standard Deviation**

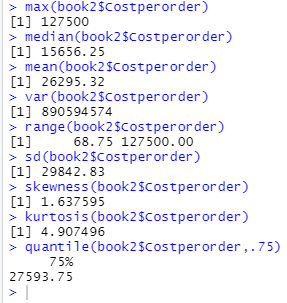


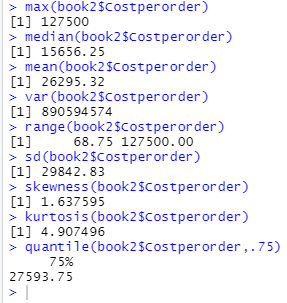
* **Hàm Quantile**



* **Hàm Summary**



* **Hàm Skewness**
* 
* **Hàm Kurtosis**



**Câu 2:**

**Dùng chức năng Visualization: Histogram, Box Plot hãy hiển thị các giá trị sau**

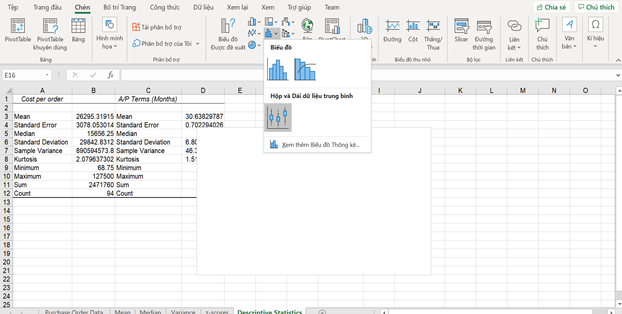
**-** **Box Plot: Min, Max, Mean, Median, Mode, Quantile, Range**

**-** **Histogram: Count, Mode, Variance, Standard Deviation, Coefficient of Deviation, Skewness, Kurtosis.**

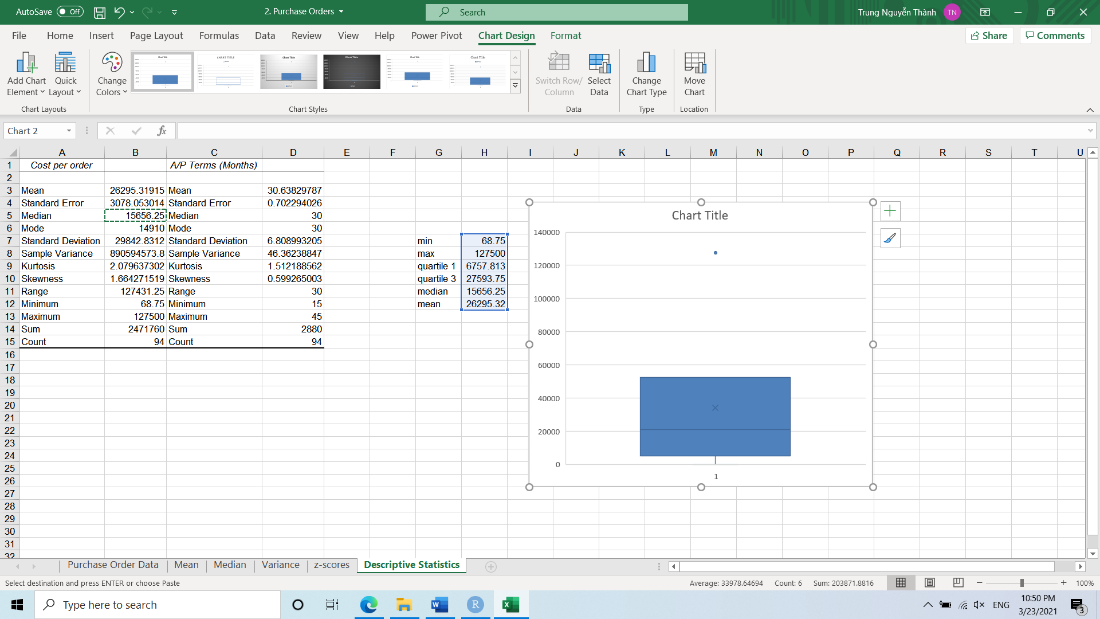
* **Box Plot**
* **Box Plot trong MS Excel**
* **Các bước thực hiện**
* Trước tiên tính các giá trị

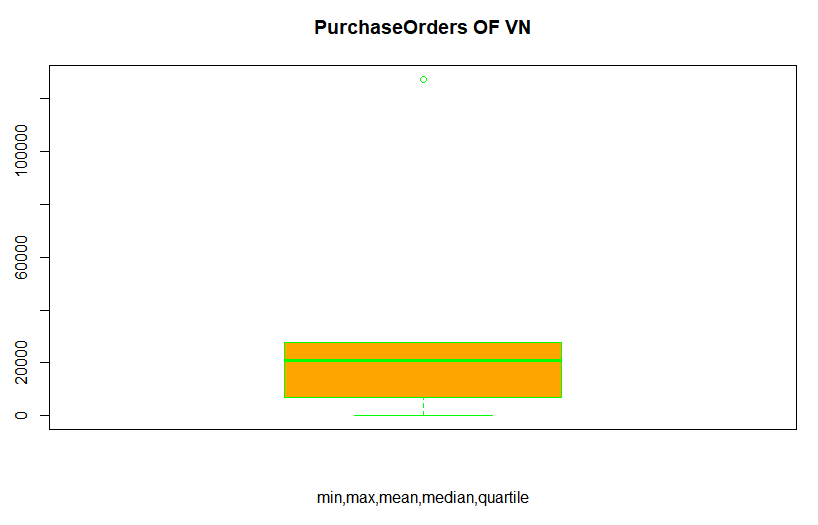
|  |  |  |
| --- | --- | --- |
| **Tên hàm** | **Thao tác** | **Giá trị** |
| Min | =Min (G4:G97) | $68,75 |
| Quartile 1 | =Quartile (G4:G97, 1) | 6757, 8125 |
| Median | =Median (G4:G97) | 15656,25 |
| Quartile 3 | = Quartile (G4:R97, 3) | 27593,75 |
| Max | =Max (G4:G97) | $127500 |

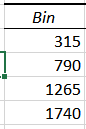
* Chọn tab Chèn trong MS Excel -> Chọn Box Plot (Hộp và dải dữ liệu trung bình)



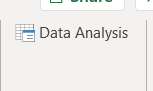
* **Bảng kết quả**



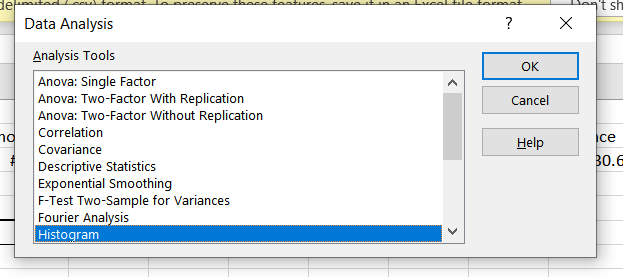
* **Box Plot trong R**
* 
* **Histogram**
* **Histogram trong MS Excel**
* Trước tiên xác định các giá trị Bin



* Trong MS Excel, vào tab Data chọn Data Analysis



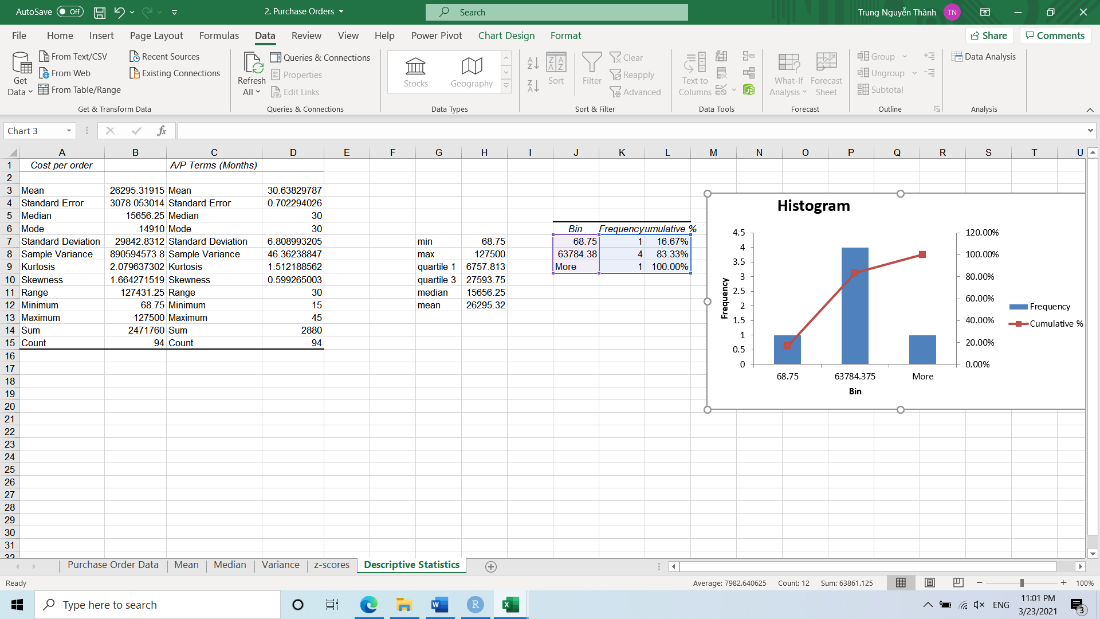
* Trong tab Data Analysis -> chọn Histogram -> chọn OK



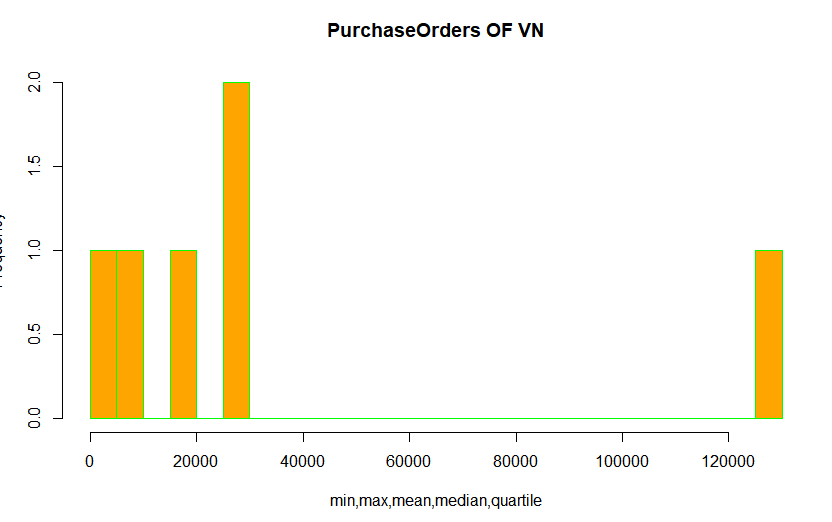
* Nhập vào các trường giá trị:

+ Input range: $H$7:$H$12

+ Bin range: $J$7:$J$8



* **Histogram trong ngôn ngữ R**



Bài 1b:

Phát biểu bài toán:

Tập tin dữ liệu cho ta biết được thông tin về đối tượng dữ liệu và số lần sửa máy tính mà ta cần để ta thống kê thông tin

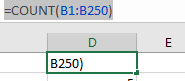
Với tập tin dữ liệu: **Computer Campair Time**

Dùng công cụ trực quan hóa dữ liệu (Data Visualization) và Thống kê mô tả (Descriptive Statistics) thực hiện yêu cầu sau:

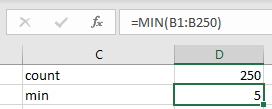
**Sử dụng MS Excel và ngôn ngữ R với số liệu về GDP bình quân của VN từ 2010 đến 2016 hãy tính và giải thích ý nghĩa của các giá trị: Count, Min, Max, Mean, Median, Mode, Quantile, Range, Mode, Variance, Standard Deviation, Coefficient of Deviation, Skewness, Kurtosis.**

**MS Excel**

* **Hàm Count**



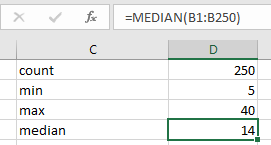
* **Hàm Min**



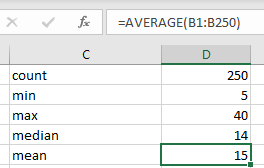
**Hàm Max**



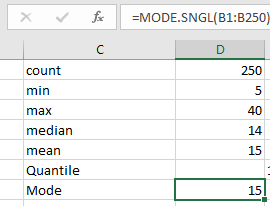
**Hàm Median**



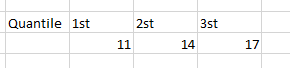
**Hàm Mean**

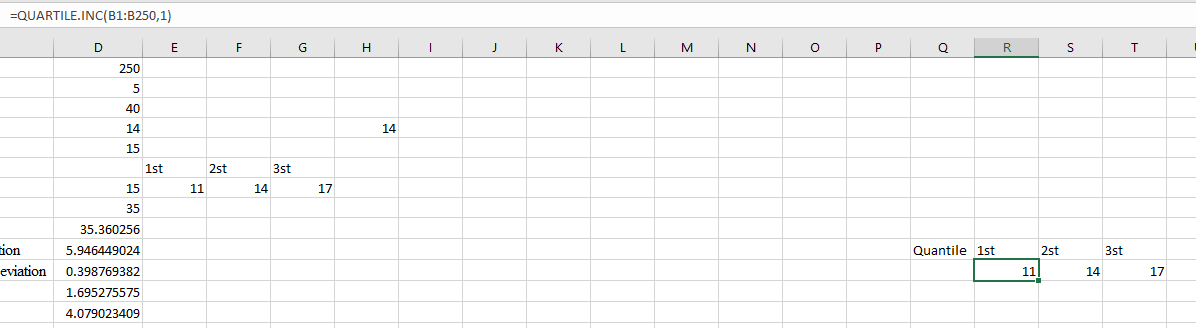


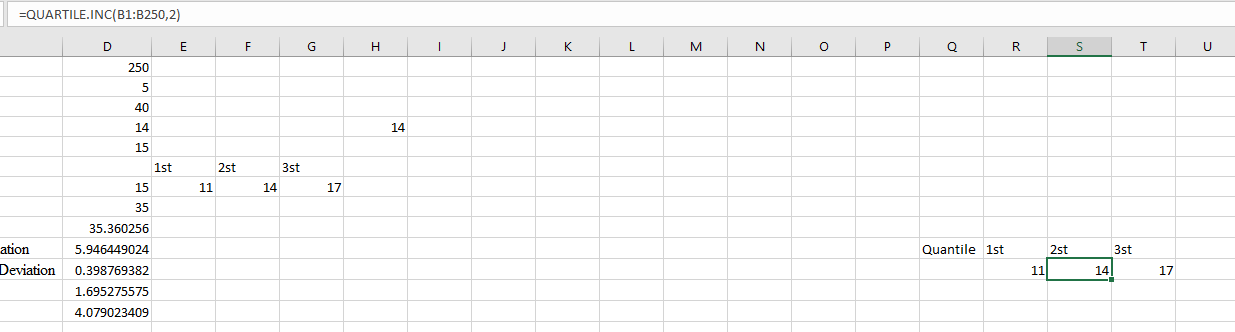
**Hàm Mode**

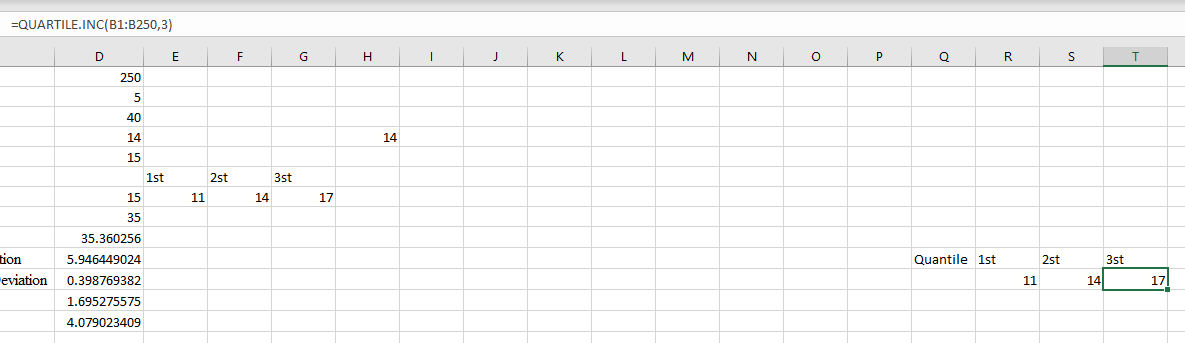


**Hàm Quantile**

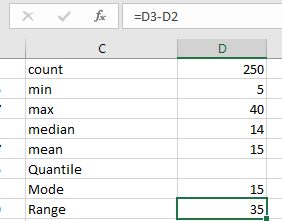




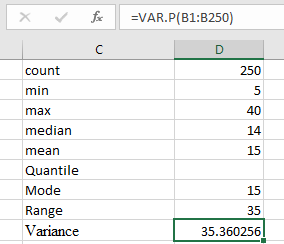




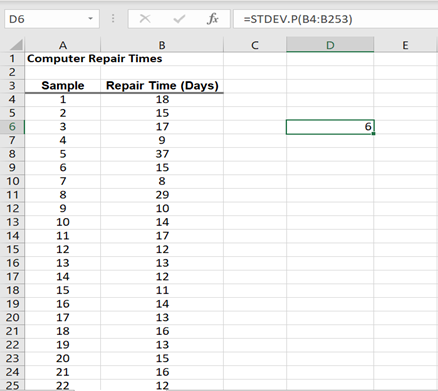
**Hàm Range**



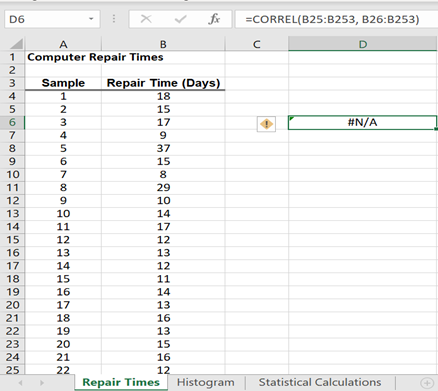
**Hàm Variance**



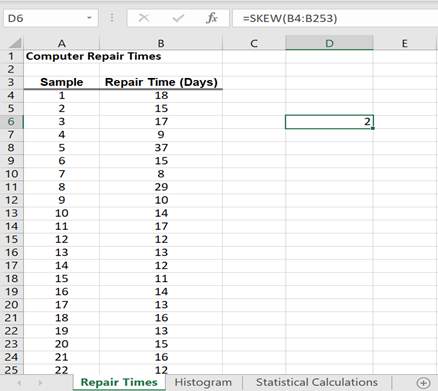
* **Hàm Standard Deviation**



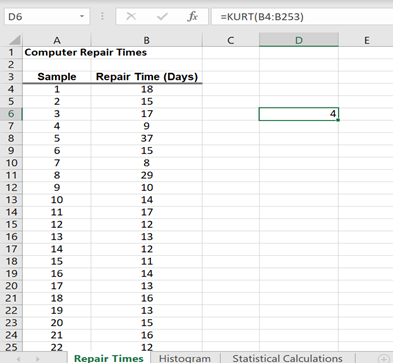
* **Hàm Coefficient of Deviation**



* **Hàm Skewness**

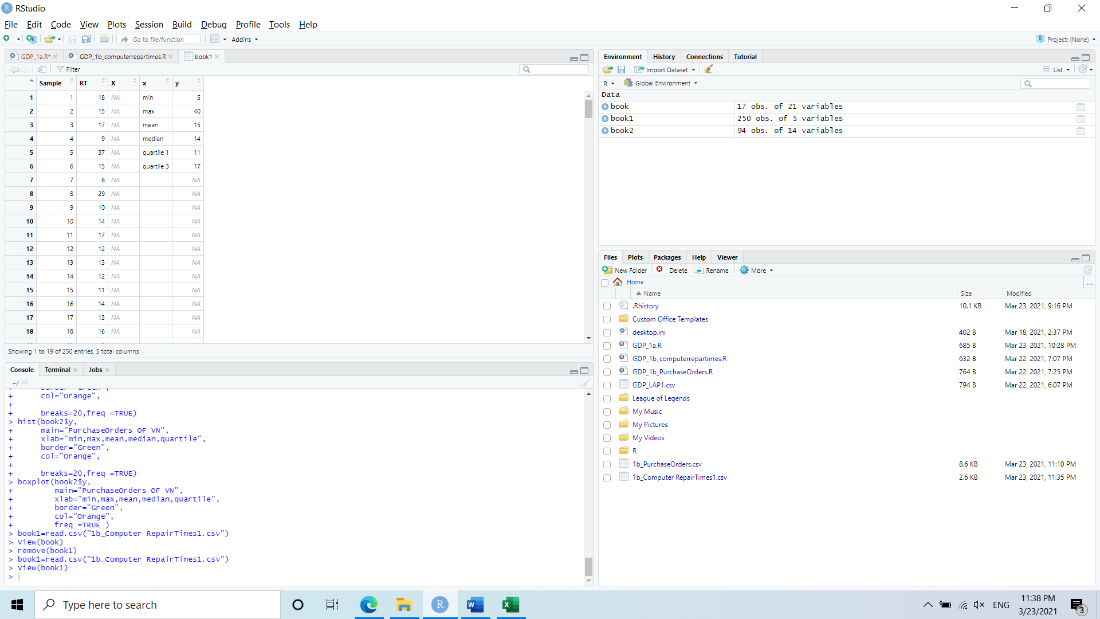


* **Hàm Kurtosis**



**Ngôn ngữ R**

**Import dữ liệu từ excel sang R**



* **Hàm Max**



* **Hàm Min**



* **Hàm Median**



* **Hàm Mean**



* **Hàm Variance**



* **Hàm Range**



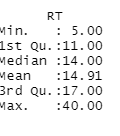
* **Hàm Standard Deviation**



* **Hàm Quantile**

Đang chèn hình ảnh...

* **Hàm Summary**



* **Hàm Skewness**
* 
* **Hàm Kurtosis**



**Câu 2:**

**Dùng chức năng Visualization: Histogram, Box Plot hãy hiển thị các giá trị sau**

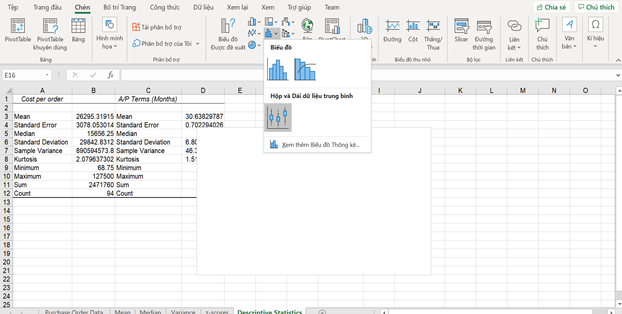
**-** **Box Plot: Min, Max, Mean, Median, Mode, Quantile, Range**

**-** **Histogram: Count, Mode, Variance, Standard Deviation, Coefficient of Deviation, Skewness, Kurtosis.**

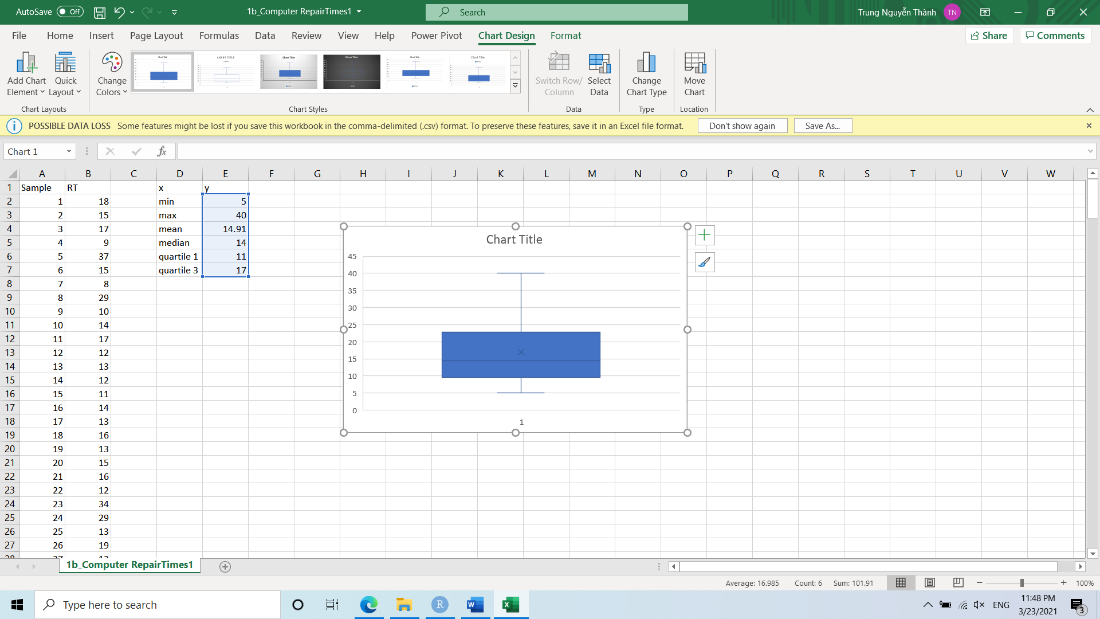
* **Box Plot**
* **Box Plot trong MS Excel**
* **Các bước thực hiện**
* Trước tiên tính các giá trị

|  |  |  |
| --- | --- | --- |
| **Tên hàm** | **Thao tác** | **Giá trị** |
| Min | =Min (B2:B251) | 5 |
| Quartile 1 | =Quartile (B2:B251, 1) | 11 |
| Median | =Median (B2:B251) | 14 |
| Quartile 3 | = Quartile (B2:B2517, 3) | 17 |
| Max | =Max (B2:B251) | 40 |
| mean | =AVERAGE(B2:B251) | 14.91 |

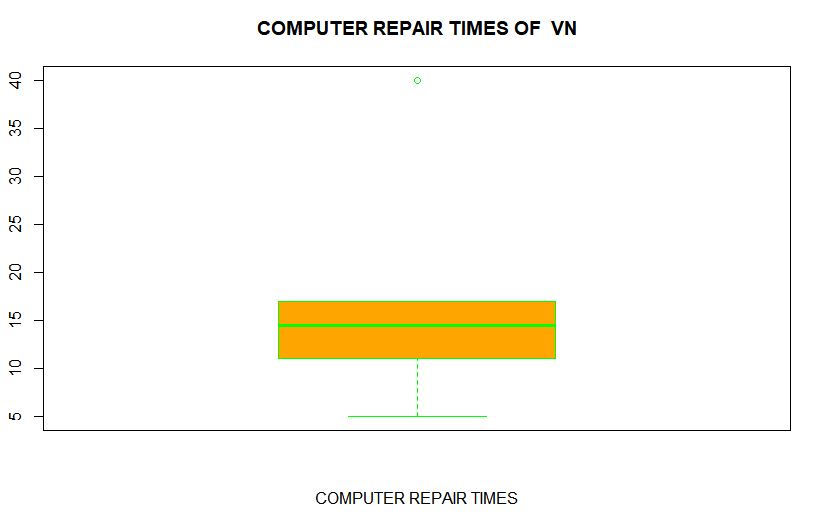
* Chọn tab Chèn trong MS Excel -> Chọn Box Plot (Hộp và dải dữ liệu trung bình)



* **Bảng kết quả**



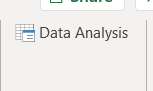
* **Box Plot trong R**



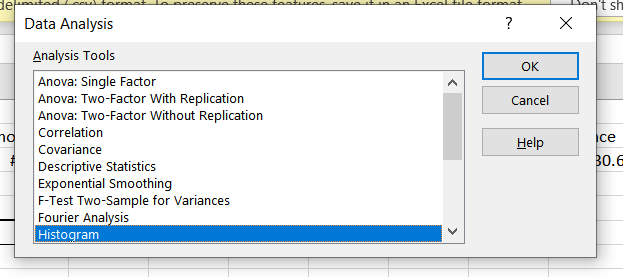
* **Histogram**
* **Histogram trong MS Excel**
* Trước tiên xác định các giá trị Bin



* Trong MS Excel, vào tab Data chọn Data Analysis



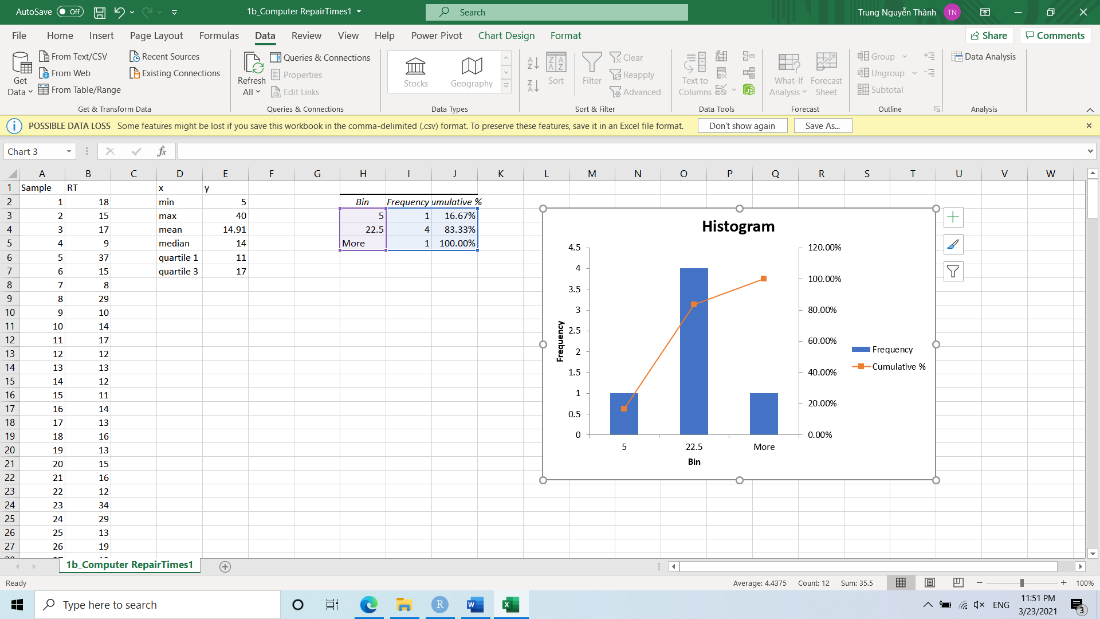
* Trong tab Data Analysis -> chọn Histogram -> chọn OK



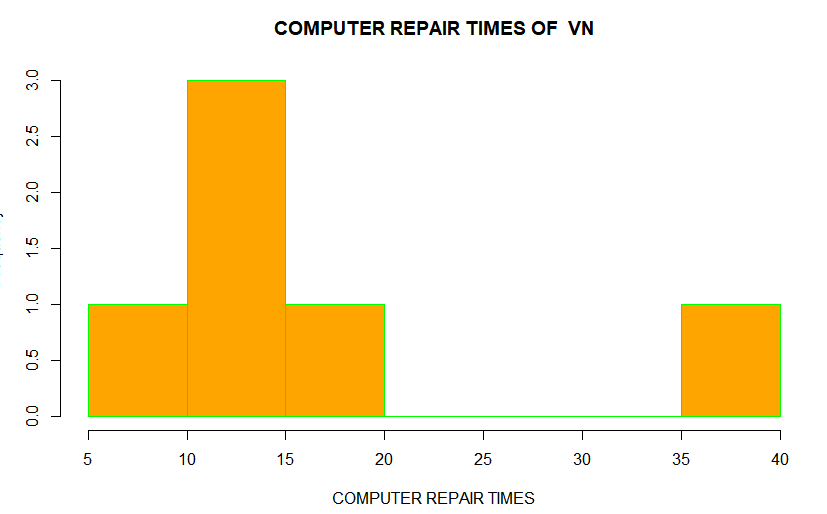
* Nhập vào các trường giá trị:

+ Input range: $E$2:$E$7

+ Bin range: $J$7:$J$8



* **Histogram trong ngôn ngữ R**



Bài 1C.

Đề bài:

Sự khác nhau giữa Predictive Analytics và Prescriptive Analytics. Cho ví dụ minh họa.

Câu trả lời:

|  |  |  |
| --- | --- | --- |
|  | Predictive Analytics | Prescriptive Analytics |
| Lí thuyết | Dựa vào dữ liệu quá khứ để dự đoán tương lai | Đề xuất giải pháp |
| Ví dụ | Sau khi xem dữ liệu các ngày 8/3 năm trước, cửa hàng hoa dựa số lượng của từng loại hoa để chuẩn bị cho 8/3 năm nay và năm sau | Sử dụng các chương trình giảm giá một số loại hoa cho một số đối tượng (vd: phụ nữ, …) vào ngày 8/3 |
| Ví dụ | Rạp chiếu phim nhận thấy rất ít khách đến rạp vào thứ 4 và thứ 5 hàng tuần. Nên thứ 4 và thứ 5 tuần này sẽ rất ít khách đến. | Tạo các chương trình giảm giá, ưu đãi, các coupon áp dụng vào ngày thứ 4 và thứ 5. |
| Ví dụ | Tiệm pizza nhận thấy rằng khách hàng rất lười ra tiệm ăn vào ngày thứ 4 hằng tuần. | Tạo chương trình ưu đãi phí ship hoặc giảm giá pizza vào ngày thứ 4. |