

**ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN KHOA HỆ THỐNG THÔNG TIN**



****

**BÀI TẬP THỰC HÀNH**

**MÔN HỌC: PHÂN TÍCH DỮ LIỆU KINH DOANH**

**LAB 01**

*Giảng viên hướng dẫn:*

***PGS.TS Nguyễn Đình Thuân***

*Sinh viên thực hiện:*

|  |  |
| --- | --- |
| ***Nguyễn Hoàng Long*** | ***- 19521788*** |
| ***Phan Phạm Quỳnh Hoa***  ***Huỳnh Minh Thư*** | ***- 19521520***  ***- 19522304*** |

*TP HCM, Ngày 05 tháng 03 năm 2022*

LỜI CẢM ƠN

Em xin gửi lời cảm ơn chân thành và sự tri ân sâu sắc đối với các thầy cô của trường Trường Đại học Công nghệ thông tin – Đại học Quốc gia TP.HCM, đặc biệt là quý thầy cô khoa Hệ thống thông tin của trường đã giúp cho chúng em trang bị các kiến thức cơ bản, các kỹ năng thực tế và tạo điều kiện để chúng em có thể hoàn thành đồ án môn học của mình.

Đặc biệt chúng em xin chân thành cảm ơn thầy Nguyễn Đình Thuân đã nhiệt tình hướng dẫn hướng dẫn, quan tâm truyền đạt những kiến thức và kinh nghiệm, trực tiếp hướng dẫn tận tình, sửa chữa và đóng góp ý kiến quý báu cho chúng em trong suốt thời gian học tập để chúng em có thể hoàn thành tốt môn học này.

Trong thời gian thực hiện, nhóm tác giả đã vận dụng những kiến thức nền tảng đã tích lũy. Từ đó, nhóm tác giả vận dụng tối đa những gì đã thu thập được để hoàn thành bài tập thực hành một cách tốt nhất. Tuy nhiên, trong quá trình thực hiện, nhóm tác giả không tránh khỏi những thiếu sót. Chính vì vậy, nhóm tác giả rất mong nhận được những sự góp ý từ phía Thầy/Cô nhằm hoàn thiện những kiến thức mà nhóm tác giả đã học tập và là hành trang để nhóm tác giả thực hiện tiếp các đề tài khác trong tương lai.

Nhóm chúng em xin chân thành cảm ơn!

***Nhóm thực hiện***

**NHẬN XÉT CỦA GIẢNG VIÊN**

………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………

**MỤC LỤC**

[I. Giải thích các đại lượng 4](#_Toc97290506)

[II. Thực hiện bài tập 1(a, b) 5](#_Toc97290507)

[1. GDP Bình quân VN 5](#_Toc97290508)

[1.1 Sử dụng Excel 5](#_Toc97290509)

[1.2 Sử dụng R 10](#_Toc97290510)

[1.3 Sử dụng Python 14](#_Toc97290511)

[2. Purchase Orders 21](#_Toc97290512)

[2.1 Sử dụng Excel 21](#_Toc97290513)

[2.2 Sử dụng R 24](#_Toc97290514)

[2.3 Sử dụng Python 28](#_Toc97290515)

[3. Computer Repair Times 36](#_Toc97290516)

[3.1 Sử dụng Excel 36](#_Toc97290517)

[3.2 Sử dụng R 39](#_Toc97290518)

[3.3 Sử dụng Python 42](#_Toc97290519)

[III. Thực hiện bài tập 1c 49](#_Toc97290520)

[1a. Null Hypothesis và Alternative Hypothesis 50](#_Toc97290521)

[1b. Type I error và Type II error 50](#_Toc97290522)

[1c. α và p-value 51](#_Toc97290523)

[IV. Bảng phân công công việc 52](#_Toc97290524)

[V. Tài liệu tham khảo 53](#_Toc97290525)

1. **Giải thích các đại lượng**
2. Count: Cho biết số lượng điểm dữ liệu hiện có trong bộ dữ liệu
3. Min: Cho biết điểm dữ liệu mang giá trị nhỏ nhất trong bộ dữ liệu
4. Max: Cho biết điểm dữ liệu mang giá trị lớn nhất trong bộ dữ liệu
5. Mean: Cho ta biết được giá trị trung bình của một tập dữ liệu số
6. Median: Là trung vị, cho ta biết giá trị nằm giữa một tập dữ liệu số đã được sắp xếp. Nó được sử dụng thay cho giá trị trung bình khi có các điểm ngoại lai trong bộ dữ liệu
7. Mode: Là điểm dữ liệu có tần suất xuất hiện nhiều nhất trong tập dữ liệu
8. Quantile: Là dạng tổng quát của median, quantile là những điểm cắt (cut points) chia tập dữ liệu thành p phần có số phần tử bằng nhau. Khi đó ta gọi các điểm này là p-quantiles
9. Range: Khoảng biến thiên của tập dữ liệu
10. Variance: Dùng để đo sự phân tán dữ liệu
11. Standard Deviation: Độ lệch chuẩn cho biết mức độ chênh lệch của các điểm dữ liệu so với giá trị trung bình (mean)
12. Coefficient of Variation: Cho biết mức độ phân tán của dữ liệu xung quanh giá trị trung bình
13. Skewness: Độ lệch của phân phối, là thước đo độ đối xứng. Một kết quả dương cho biết một phân phối lệch về bên phải. Kết quả âm chỉ ra một phân phối lệch về bên trái
14. Kurtosis: Dùng để mô tả phân phối xác suất của các điểm dữ liệu ngẫu nhiên so với phân phối chuẩn. Kurtosis dương chỉ ra một phân phối tương đối đạt đỉnh. Kurtosis âm cho thấy sự phân bố tương đối bằng phẳng
15. **Thực hiện bài tập 1(a, b)**
16. **GDP Bình quân VN**
    1. **Sử dụng Excel**

Tính toán các giá trị: Count, Min, Max, Mean, Median, Mode, Quantile, Range, Variance, Standard Deviation, Coefficient of Deviation, Skewness, Kurtosis

* Count sử dụng hàm COUNT(value1, value2….)
* Min sử dụng hàm MIN(number1, [number2], …)
* Max sử dụng hàm MAX(number1, [number2], …)
* Mean = MIN(number1, [number2], …) / COUNT(value1, value2….)
* Median sử dụng hàm MEDIAN(number1, [number2], …)
* Mode sử dụng hàm MODE(number1, [number2], …)
* Range = MAX – MIN
* Quantile sử dụng hàm QUARTILE.INC(array,quart) với:
  + - * quart = 1 tứ phân vị thứ nhất
      * quart = 2 giá trị trung vị
      * quart = 3 tứ phân vị thứ ba
* Variance áp dụng công thức
* Standard Deviation được tính bằng căn bậc hai của phương sai.
* Coefficient of Deviation được tính bằng
* Skewness áp dụng công thức
* Kurtosis áp dụng công thức -

Kết quả các giá trị cần tính:

Graphical user interface, application, table

Description automatically generated

Dùng chức năng Visualization: Histogram và Box Plot

* **Histogram**

Sử dụng chức năng Histogram trong Data Analysis. Trước hết phải tạo một mảng chứa các “bin” – là các điểm phân cách giá trị.

Giá trị GDP có Min=302 và Max=2215 nên có thể chia GDP thành các nhóm với khoảng cách là 100. Bin lần lượt là 300, 400, 500, …, 2300.

Vào Data =>Data Analysis => Histogram

Graphical user interface, text, application

Description automatically generated

Tại hộp thoại Histogram:

**Input Range:** Vùng chứa các giá trị của đại lượng cần vẽ

**Bin Range:** Vùng phân phối thành lớp

**Output Range:** Điểm đầu của kết quả cần hiển thị

Chọn Input Range, Bin Range và Output range tương ứng và nhấn OK

Graphical user interface

Description automatically generated

Bảng giá trị và Histogram tương ứng với bộ dữ liệu GDP:

Chart, bar chart

Description automatically generated

* Box Plot

Graphical user interface, application, table

Description automatically generatedVào Insert =>Insert Statistic Chart Chọn Box and Whisker

Chọn Select Data

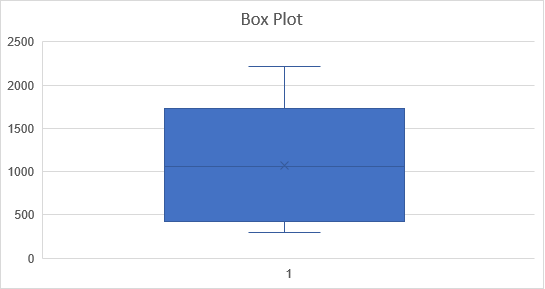
Tại hộp thoại Select Data Source

**Chart data range:** chọn vùng chứa các giá trị của đại lượng cần vẽ

Graphical user interface

Description automatically generated

Nhấn OK. Ta được Box Plot tương ứng với bộ dữ liệu GDP

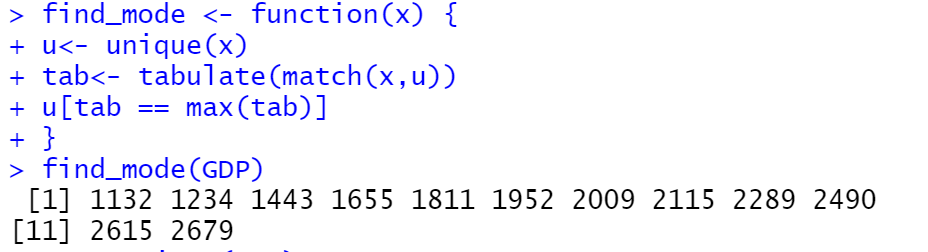


* 1. **Sử dụng R**

Tính toán các giá trị: Count, Min, Max, Mean, Median, Mode, Quantile, Range, Variance, Standard Deviation, Coefficient of Deviation, Skewness, Kurtosis

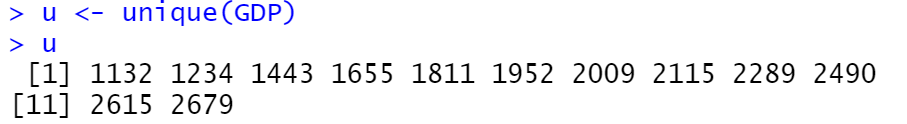
|  |  |
| --- | --- |
|  | GDP |
| Mean | 1952 |
| Median | 1980.5 |
| Mode | NA |
| Std. Deviation | 518.3912 |
| Coefficient of Deviation | 0.2655693 |
| Variance | 268729.5 |
| Skewness | -0.1245471 |
| Kurtosis | -1.426998 |
| Range | 1547 |
| Minimum | 1132 |
| Maximum | 2679 |
| Count | 12 |

***Hàm find\_mode:*** tìm các giá trị có tần số xuất hiện nhiều nhất trong dữ liệu. Được tính bằng công thức:

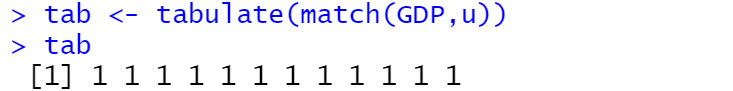


*Trong đó:*

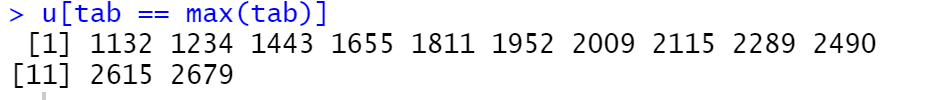
* Unique(x) để loại bỏ giá trị trùng trong vector x
* Match(x,u) để trả về một vector có vị trí xuất hiện đầu tiên của vector x trong vector u, nếu phần tử của x không tồn tại trong vector u thì trả về giá trị NA.
* Tabulate() để đếm tần số xuất hiện của từng phần tử trong vector.







* Tìm giá trị có tần số lớn nhất: kiểm tra vị trí trong vector u có giá trị bằng giá trị lớn nhất trong tab.



Các hàm ***Min, max, mean, median, var, Std. Deviation(sd) :***

Graphical user interface, text, application

Description automatically generated

Hàm ***range*** được tính bằng công thức ***max(x) – min(x)***

Hàm ***cv*** *(Coefficient of Deviation)* được tính bằng công thức ***sd(x)/mean(x).***

Graphical user interface, text, application

Description automatically generated

Để gọi hàm ***skewness và kurtosis,*** trước tên cần *install* package *e1071* và *require (e1071).*

Graphical user interface, text, application

Description automatically generated

Dùng chức năng Visualization: Histogram và Box Plot

* Histogram: Cú pháp *hist(x,main=”Ten bieu do”, y ,col=”yellow”)*

Chart, histogram

Description automatically generated

* Box Plot: Cú pháp *boxplot(x,main=”Ten bieu do”, y ,col=”orange”)*

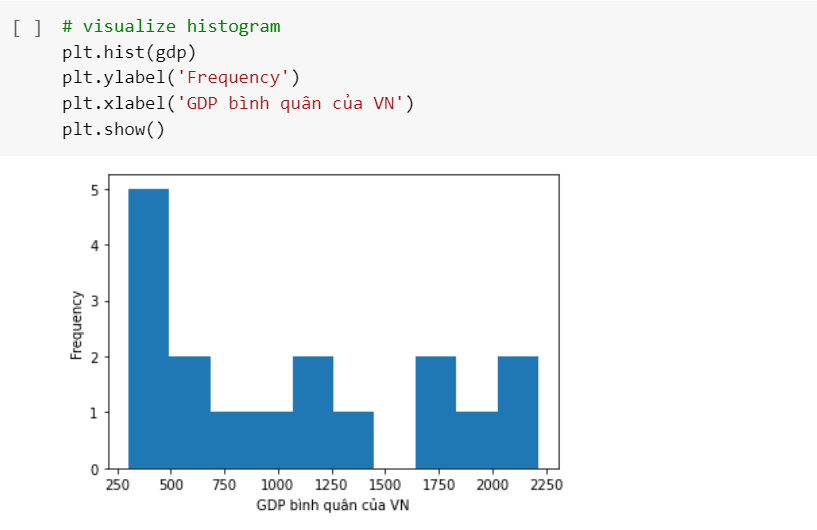
Chart, box and whisker chart

Description automatically generated

* 1. **Sử dụng Python**

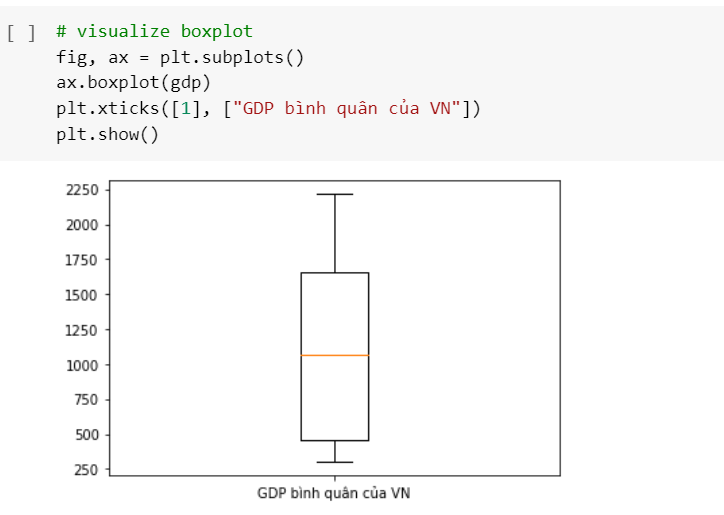
Sử dụng Matplotlib thực hiện visualize Histogram và Boxplot

1. Histogram



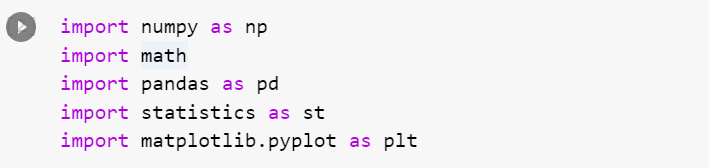
***Histogram GDP bình quân của VN***

2. Boxplot



***Boxplot GDP bình quân của VN***

3. Tính toán các đại lượng



***Khai báo các thư viện cần thiết***



***Đọc dữ liệu sử dụng pandas***

Ảnh có chứa văn bản

Mô tả được tạo tự động

***Đại lượng Count***

Ảnh có chứa văn bản

Mô tả được tạo tự động

***Đại lượng Min***

Ảnh có chứa văn bản

Mô tả được tạo tự động

***Đại lượng Max***

Ảnh có chứa văn bản

Mô tả được tạo tự động

***Đại lượng Mean***

Ảnh có chứa văn bản

Mô tả được tạo tự động

***Đại lượng Median***

Ảnh có chứa văn bản

Mô tả được tạo tự động

***Đại lượng Mode***

Ảnh có chứa văn bản

Mô tả được tạo tự động

***Đại lượng Quantile***

Ảnh có chứa văn bản

Mô tả được tạo tự động

***Đại lượng Range***

Ảnh có chứa văn bản

Mô tả được tạo tự động

***Đại lượng Variance***

Ảnh có chứa văn bản

Mô tả được tạo tự động

***Đại lượng Standard Deviation***

Ảnh có chứa văn bản

Mô tả được tạo tự động

***Đại lượng Coefficient of Variation***

Ảnh có chứa văn bản

Mô tả được tạo tự động

***Đại lượng Skewness***

Ảnh có chứa văn bản

Mô tả được tạo tự động

***Đại lượng Kurtosis***

1. **Purchase Orders**
   1. **Sử dụng Excel**

Tính toán các giá trị: Count, Min, Max, Mean, Median, Mode, Quantile, Range, Variance, Standard Deviation, Coefficient of Deviation, Skewness, Kurtosis. Tương tự mục 1.1.1 ta được kết quả như sau:

Graphical user interface, application, table, Excel

Description automatically generated

Graphical user interface, application, table, Excel

Description automatically generated

Dùng chức năng Visualization: Histogram và Box Plot

* Histogram

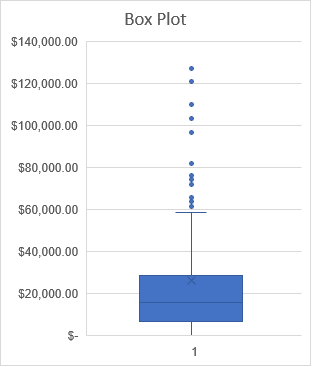
Giá trị Cost per order có Min=68.75 và Max=127,500.00 nên có thể chia Cost per order thành các nhóm với khoảng cách là 5000. Bin lần lượt là 0, 5000, 10000, …, 130000.

Bảng giá trị và Histogram tương ứng với bộ dữ liệu:

**Chart, histogram

Description automatically generated**

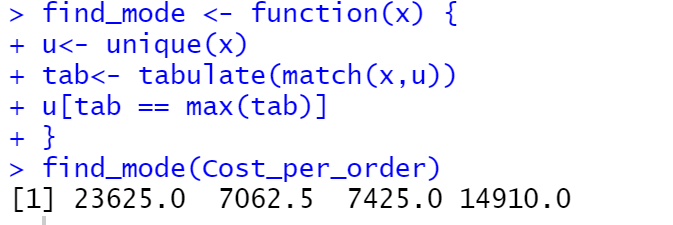
* Box Plot



* 1. **Sử dụng R**

Tính toán các giá trị: Count, Min, Max, Mean, Median, Mode, Quantile, Range, Variance, Standard Deviation, Coefficient of Deviation, Skewness, Kurtosis

|  |  |
| --- | --- |
|  | Cost per Order |
| Mean | 26295.32 |
| Median | 15656.25 |
| Mode | NA |
| Std. Deviation | 29842.83 |
| Coefficient of Deviation | 1.13491 |
| Variance | 890594574 |
| Skewness | 1.611533 |
| Kurtosis | 1.803636 |
| Range | 127431.2 |
| Minimum | 68.75 |
| Maximum | 127500 |
| Count | 94 |



Graphical user interface, text, application

Description automatically generated

A picture containing application

Description automatically generated

A picture containing diagram

Description automatically generated

Text

Description automatically generated

A picture containing graphical user interface

Description automatically generated

Dùng chức năng Visualization: Histogram và Box Plot

* Histogram: Cú pháp tương tự *hist(x,main=”Ten bieu do”, y ,col=”green”),* thu được kết quả như sau:

Chart, histogram

Description automatically generated

* Box Plot: Cú pháp *boxplot(x,main=”Ten bieu do”, y ,col=”orange”),* thu được kết quả như sau:

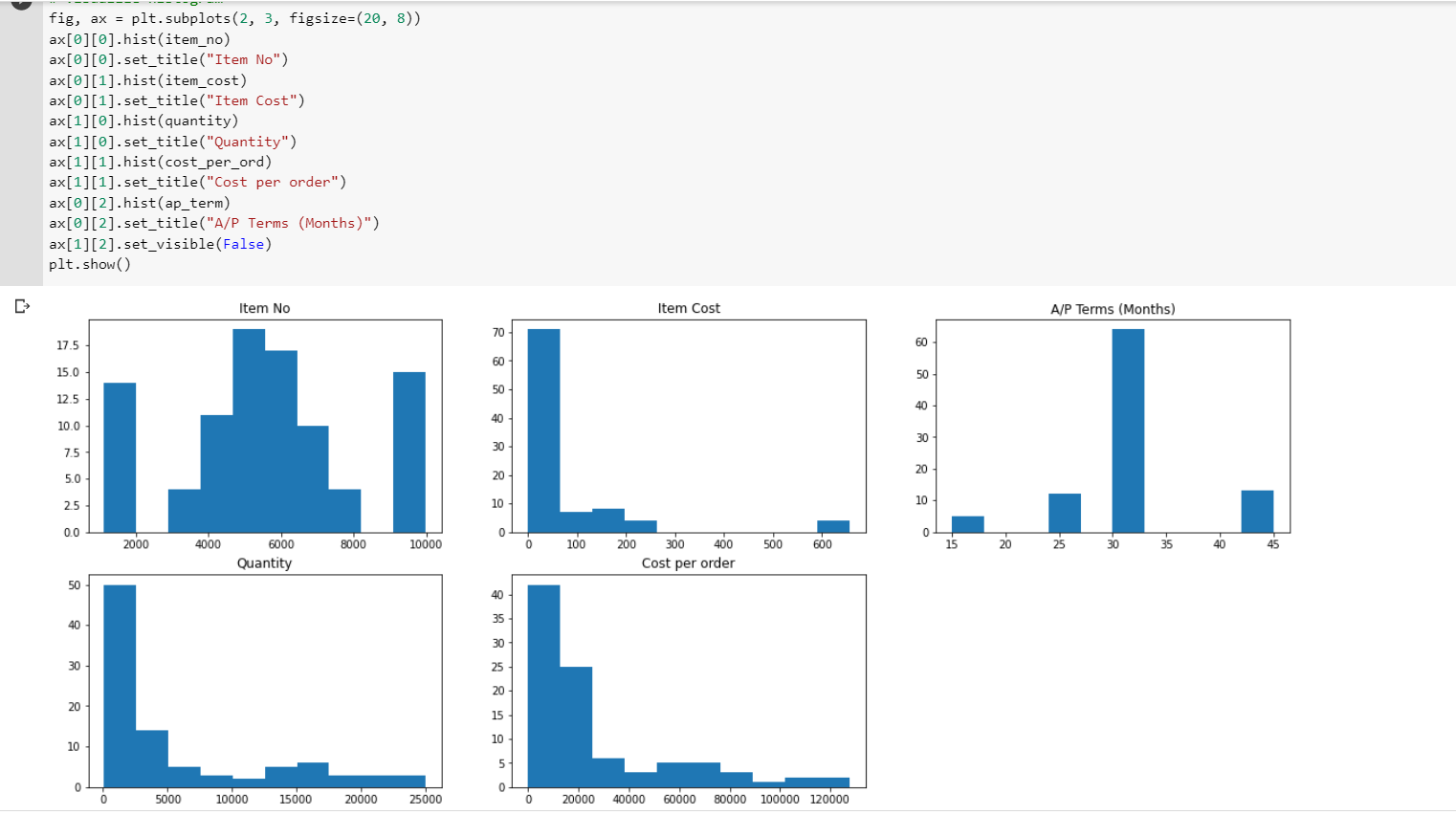
Chart, diagram, box and whisker chart

Description automatically generated

* 1. **Sử dụng Python**

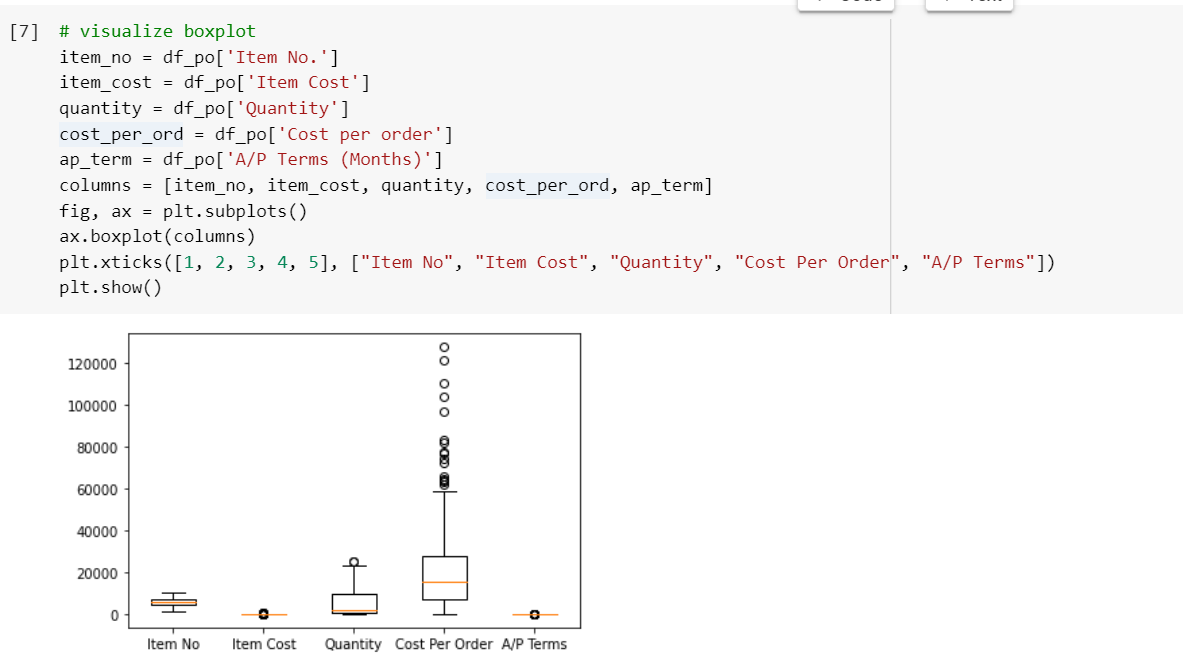
Sử dụng Matplotlib thực hiện visualize Histogram và Boxplot

1. Histogram



***Histogram các cột mang giá trị số***

2. Boxplot



***Boxplot các cột mang giá trị số***

3. Tính toán các đại lượng

Ảnh có chứa văn bản

Mô tả được tạo tự động

***Đọc dữ liệu sử dụng pandas***

Ảnh có chứa văn bản

Mô tả được tạo tự động

***Đại lượng Count***

Ảnh có chứa văn bản

Mô tả được tạo tự động

***Đại lượng Min***

Ảnh có chứa văn bản

Mô tả được tạo tự động

***Đại lượng Max***

Ảnh có chứa văn bản

Mô tả được tạo tự động

***Đại lượng Mean***

Ảnh có chứa văn bản

Mô tả được tạo tự động

***Đại lượng Median***

Ảnh có chứa văn bản

Mô tả được tạo tự động

***Đại lượng Mode***

Ảnh có chứa văn bản

Mô tả được tạo tự động

***Đại lượng Quantile với p = 0.2***

Ảnh có chứa văn bản

Mô tả được tạo tự động

***Đại lượng Range***

Ảnh có chứa văn bản

Mô tả được tạo tự động

***Đại lượng Variance***

Ảnh có chứa văn bản

Mô tả được tạo tự động

***Đại lượng Standard Deviation***

Ảnh có chứa văn bản

Mô tả được tạo tự động

***Đại lượng Coefficient of Variation***

Ảnh có chứa văn bản

Mô tả được tạo tự động

***Đại lượng Skewness***

Ảnh có chứa văn bản

Mô tả được tạo tự động

***Đại lượng Kurtosis***

1. **Computer Repair Times**
   1. **Sử dụng Excel**

Tính toán các giá trị: Count, Min, Max, Mean, Median, Mode, Quantile, Range, Variance, Standard Deviation, Coefficient of Deviation, Skewness, Kurtosis. Tương tự mục 1.1.1 ta được kết quả như sau:

Graphical user interface, table, Excel

Description automatically generated

Graphical user interface, application, table, Excel

Description automatically generated

Dùng chức năng Visualization: Histogram và Box Plot

* Histogram

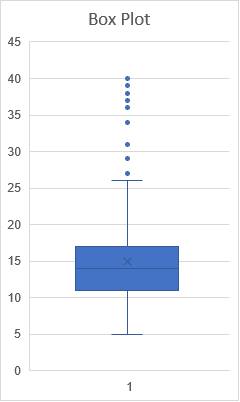
Giá trị Days có Min=5 và Max=40 nên có thể chia Days thành các nhóm với khoảng cách là 2. Bin lần lượt là 4, 6, 8, …, 10.

Bảng giá trị và Histogram tương ứng với bộ dữ liệu:

**Chart, bar chart, histogram

Description automatically generated**

* Box Plot



* 1. **Sử dụng R**

Tính toán các giá trị: Count, Min, Max, Mean, Median, Mode, Quantile, Range, Variance, Standard Deviation, Coefficient of Deviation, Skewness, Kurtosis

|  |  |
| --- | --- |
|  | Repair Time (Days) |
| Mean | 14.912 |
| Median | 14 |
| Mode | NA |
| Std. Deviation | 5.958378 |
| Coefficient of Deviation | 0.3995693 |
| Variance | 35.50227 |
| Skewness | 1.674987 |
| Kurtosis | 3.918314 |
| Range | 35 |
| Minimum | 5 |
| Maximum | 40 |
| Count | 250 |

Graphical user interface, text

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application

Description automatically generated

Dùng chức năng Visualization: Histogram và Box Plot

* Histogram: Cú pháp tương tự 1.2.2 *hist(x,main=”Ten bieu do”, y ,col=”yellow”),* thu được kết quả như sau:

Chart, histogram

Description automatically generated

* Box Plot: Cú pháp tương tự *boxplot(x,main=”Ten bieu do”, y ,col=”orange”),* thu được kết quả:

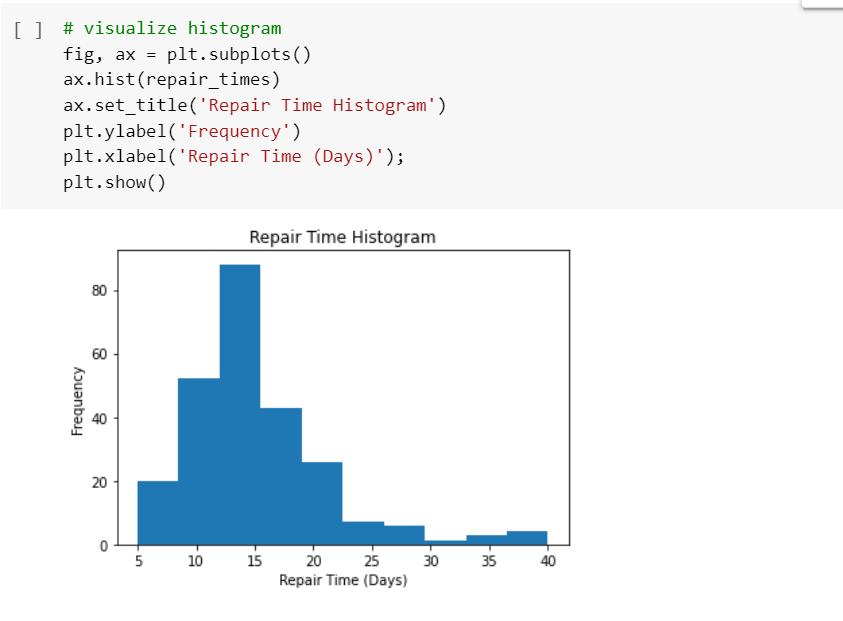
Chart, box and whisker chart

Description automatically generated

* 1. **Sử dụng Python**

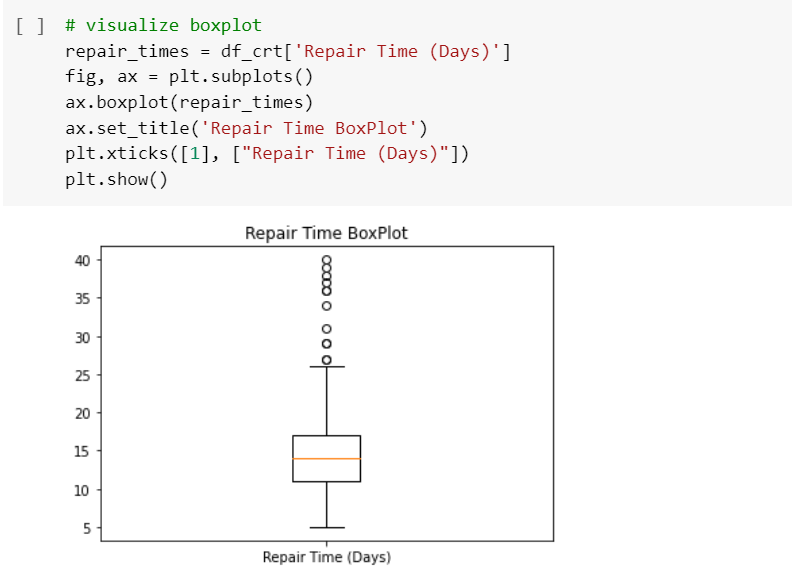
Sử dụng Matplotlib thực hiện visualize Histogram và Boxplot

1. Histogram

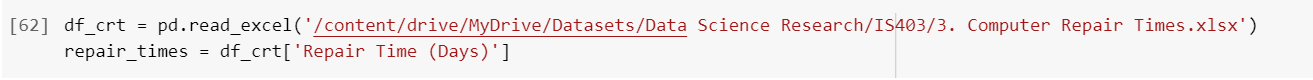


***Histogram Repair Time (Days)***

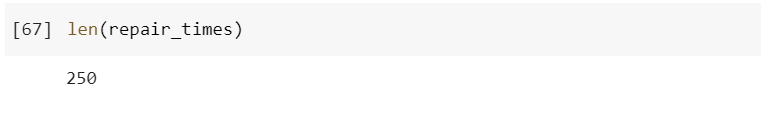
2. Boxplot



***Boxplot Repair Time (Days)***



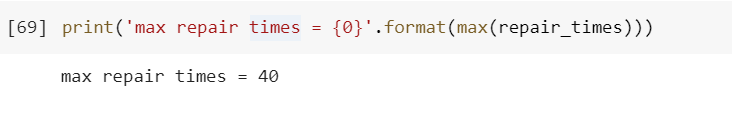
***Đọc dữ liệu sử dụng pandas***



***Đại lượng Count***



***Đại lượng Min***



***Đại lượng Max***

Ảnh có chứa văn bản

Mô tả được tạo tự động

***Đại lượng Mean***

Ảnh có chứa văn bản

Mô tả được tạo tự động

***Đại lượng Median***

Ảnh có chứa văn bản

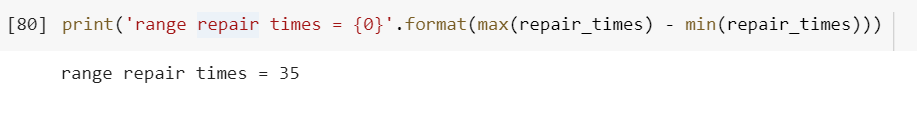
Mô tả được tạo tự động

***Đại lượng Mode***

Ảnh có chứa văn bản

Mô tả được tạo tự động

***Đại lượng Quantile với p = 0.2***



***Đại lượng Range***

Ảnh có chứa văn bản

Mô tả được tạo tự động

***Đại lượng Variance***

Ảnh có chứa văn bản

Mô tả được tạo tự động

***Đại lượng Standard Deviation***

Ảnh có chứa văn bản

Mô tả được tạo tự động

***Đại lượng Coefficient of Variation***

Ảnh có chứa văn bản

Mô tả được tạo tự động

***Đại lượng Skewness***

Ảnh có chứa văn bản

Mô tả được tạo tự động

***Đại lượng Kurtosis***

1. **Thực hiện bài tập 1c**

Xét tình huống tiệm bánh kem HTCL mới mua một máy nướng bánh mới. Ông chủ mong muốn bánh nướng ra phải có khối lượng là 30g

Câu hỏi đặt ra: Vậy những cái bánh nướng mới này có đạt được yêu cầu của ông không?

Chúng ta lấy 50 mẫu bánh sau khi nướng ra cân và chạy 1 sample T test. Phương pháp này sẽ giúp ta tìm ra đáp án

**1a. Null Hypothesis và Alternative Hypothesis**

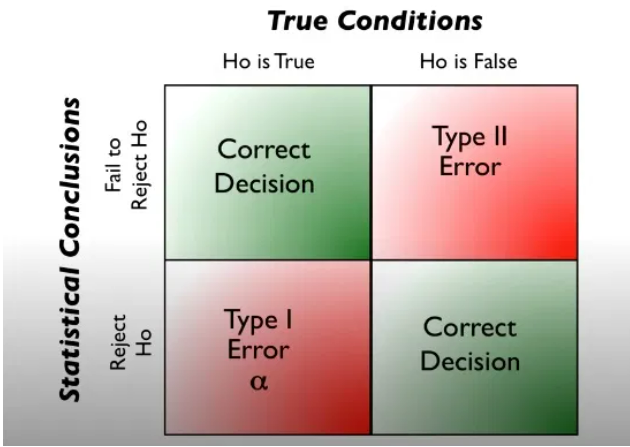
Giả thiết Ho – Null Hypothesis

Giả thiết Ha – Alternative Hypothesis

Trong đó:

* Ho là mệnh đề giả định rằng các giá trị đều bằng nhau (hay được hiểu rằng không có sự khác biệt đáng kể giữa các mẫu). Nhìn vào tình huống trên, Ho là những chiếc bánh nướng có khối lượng bằng 30g
* Ha là mệnh đề thay thế cho Ho (hay còn gọi là Reject Ho), nghĩa là giả định rằng có sự khác biệt đáng kể giữa các mẫu. Nhìn vào tình huống trên, Ha là những chiếc bánh nướng có khối lượng khác 30g

**1b. Type I error và Type II error**



* Type I Error: Là lỗi do bác bỏ giả thuyết Ho (Null hypothesis) trong khi nó đúng. Xét ví dụ, lỗi loại 1 tương ứng với bác bỏ giả thuyết chiếc bánh nướng có khối lượng 30g nhưng thực tế là nó có khối lượng 30g. Lỗi loại 1 liên quan tới hệ số ***α***
* Type II Error: Là lỗi do chấp nhận giả thuyết Ho (Reject Null Hypothesis) trong khi nó sai. Xét ví dụ, lỗi loại 2 tương ứng với chấp nhận giả thuyết chiếc bánh nướng có khối lượng 30g nhưng thực tế là nó có khối lượng khác 30g. Lỗi loại 2 liên quan tới hệ số **b**

**1c. α và p-value**

Về bản chất, p-value và α cùng là mức ý nghĩa (level of significance). Điểm khác nhau là, α được chọn trước, còn p-value thì do tính toán mà ra. Khi kiểm nghiệm một giả thuyết, chúng ta **chọn một mức ý nghĩa (α)**, và sau đó tính toán số liệu thống kê dựa trên dữ liệu để lấy được một **mức ý nghĩa quan sát được (p-value)**. Cuối cùng, ta so sánh 2 mức này (α và p-value) để đưa ra quyết định bác bỏ giả thuyết hoặc không bác bỏ giả thuyết. Mức ý nghĩa **α** còn mang ý nghĩa là chúng ta sẵn sàng chấp nhận mức rủi ro rằng chúng ta đã sai khi bác bỏ Ho.

Xét lại ví dụ trên, p-value là giá trị mà ta sẽ nhận được sau khi chạy T-test và dùng nó để kết luận, còn α là giá trị ta đặt trước với công thức như sau

Ảnh có chứa bàn

Mô tả được tạo tự động

1. **Bảng phân công công việc**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Nguyễn Hoàng Long | Phan Phạm Quỳnh Hoa | Huỳnh Minh Thư |
| Xây dựng Template báo cáo | X |  |  |
| Tìm hiểu và sử dụng Excel |  |  | X |
| Tìm hiểu và sử dụng R |  | X |  |
| Tìm hiểu và sử dụng Python | X |  |  |
| Làm bài tập 1a,b (Sử dụng Excel) |  |  | X |
| Làm bài tập 1a,b (Sử dụng R) |  | X |  |
| Làm bài tập 1a,b (Sử dụng Python) | X |  |  |
| Làm bài tập 1c (phần a) |  | X |  |
| Làm bài tập 1c (phần b) |  |  | X |
| Làm bài tập 1c (phần c) | X |  |  |

1. **Tài liệu tham khảo**

[1] <https://ngocminhtran.com/2019/06/10/machine-learning-mot-vai-khai-niem-co-ban-trong-thong-ke-statistic/>

[2] <https://stackabuse.com/calculating-variance-and-standard-deviation-in-python/>

[3] <https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.skew.html>

[4] <https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.kurtosis.html>

[5] <https://nguyenvantien0405.files.wordpress.com/2018/03/hd_excel_xstk_161223.pdf>

[6] <https://corporatefinanceinstitute.com/resources/excel/functions/>

[7] <https://www.careerlink.vn/cam-nang-viec-lam/kien-thuc-kinh-te/null-hypothesis-la-gi-khac-gi-voi-alternative-hypothesis>

[8] <https://vietquality.vn/hypothesis-testing-la-gi/>

[9] <https://tuihoctaichinh.com/quant-alpha-vs-p-value/>

[10] <https://en.wikipedia.org/wiki/Type_I_and_type_II_errors>