TỔNG LIÊN ĐOÀN LAO ĐỘNG VIỆT NAM

**TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

**KHOA CÔNG NGHỆ THÔNG TIN**



**BÀI TẬP QUÁ TRÌNH 2**

**XÁC SUẤT THỐNG KÊ – CÔNG NGHỆ THÔNG TIN**

*Người hướng dẫn*: **TS NGUYỄN LÂM**

*Người thực hiện*: **TỪ GIA BẢO – 522H0087**

Lớp **: 22H50201**

Khoá  **: 26**

**THÀNH PHỐ HỒ CHÍ MINH, NĂM 2023**

LỜI CẢM ƠN

*“Để hoàn thành tiểu luận này, em xin gửi lời cảm ơn chân thành đến:*

*Ban giám hiệu trường Đại Học Tôn Đức Thắng vì đã tạo điều kiện về cơ sở vật chất với hệ thống thư viện hiện đại, đa dạng các loại sách, tài liệu thuận lợi cho việc tìm kiếm, nghiên cứu thông tin.*

*Xin cảm ơn giảng viên bộ môn – Thầy Nguyễn Lâm đã giảng dạy tận tình, chi tiết để em có đủ kiến thức và vận dụng chúng vào bài tiểu luận này.*

*Do chưa có nhiều kinh nghiệm làm để tài cũng như những hạn chế về kiến thức, trong bài tiểu luận chắc chắn sẽ không tránh khỏi những thiếu sót. Rất mong nhận được sự nhận xét, ý kiến đóng góp, phê bình từ phía Cô để bài tiểu luận được hoàn thiện hơn.*

*Lời cuối cùng, em xin kính chúc cô nhiều sức khỏe, thành công và hạnh phúc.”*

**ĐỒ ÁN ĐƯỢC HOÀN THÀNH**

**TẠI TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

Tôi xin cam đoan đây là sản phẩm đồ án của riêng tôi và được sự hướng dẫn của TS Nguyễn Lâm. Các nội dung nghiên cứu, kết quả trong đề tài này là trung thực và chưa công bố dưới bất kỳ hình thức nào trước đây. Những số liệu trong các bảng biểu phục vụ cho việc phân tích, nhận xét, đánh giá được chính tác giả thu thập từ các nguồn khác nhau có ghi rõ trong phần tài liệu tham khảo.

Ngoài ra, trong đồ án còn sử dụng một số nhận xét, đánh giá cũng như số liệu của các tác giả khác, cơ quan tổ chức khác đều có trích dẫn và chú thích nguồn gốc.

**Nếu phát hiện có bất kỳ sự gian lận nào tôi xin hoàn toàn chịu trách nhiệm về nội dung đồ án của mình.** Trường đại học Tôn Đức Thắng không liên quan đến những vi phạm tác quyền, bản quyền do tôi gây ra trong quá trình thực hiện (nếu có).

*TP. Hồ Chí Minh, ngày tháng năm*

*Tác giả*

*(ký tên và ghi rõ họ tên)*

*Từ Gia Bảo*

PHẦN XÁC NHẬN VÀ ĐÁNH GIÁ CỦA GIẢNG VIÊN

**Phần xác nhận của GV hướng dẫn**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Tp. Hồ Chí Minh, ngày tháng năm

(kí và ghi họ tên)

**Phần đánh giá của GV chấm bài**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Tp. Hồ Chí Minh, ngày tháng năm

(kí và ghi họ tên)

TÓM TẮT

Bài báo cáo này trình bài những nội dung tìm hiểu về các hàm mà giáo viên hướng dẫn đưa ra cụ thể là nêu ra mục đích sử dụng của các hàm, các biến, tham số liên quan và các ví dụ về việc sử dụng hàm.

**MỤC LỤC**

[Phần I: Averages and measures of central location 6](#_Toc141955552)

[1.Hàm mean(): 6](#_Toc141955553)

[2.Hàm fmean(): 6](#_Toc141955554)

[3.Hàm geometric\_mean(): 7](#_Toc141955555)

[4.Hàm harmonic\_mean(): 7](#_Toc141955556)

[5.Hàm media(): 8](#_Toc141955557)

[6.Hàm median\_low(): 9](#_Toc141955558)

[7.Hàm median\_high(): 9](#_Toc141955559)

[8.Hàm median\_grouped(): 10](#_Toc141955560)

[9.Hàm mode(): 10](#_Toc141955561)

[10.Hàm multimode(): 11](#_Toc141955562)

[11.Hàm quantiles(): 11](#_Toc141955563)

[Phần II: Measures of spread 11](#_Toc141955564)

[12.Hàm pstdev(): 11](#_Toc141955565)

[13.Hàm pvariance(): 12](#_Toc141955566)

[14.Hàm stdev(): 12](#_Toc141955567)

[15.Hàm variance(): 13](#_Toc141955568)

[16.Hàm convariance(): 13](#_Toc141955569)

[17.Hàm correlation(): 13](#_Toc141955570)

[18.Hàm linear\_regression(): 14](#_Toc141955571)

[Phần III: NormalDist objects 14](#_Toc141955572)

[19.Hàm NormalDist(): 14](#_Toc141955573)

## Phần I: Averages and measures of central location

## 1. Hàm mean():

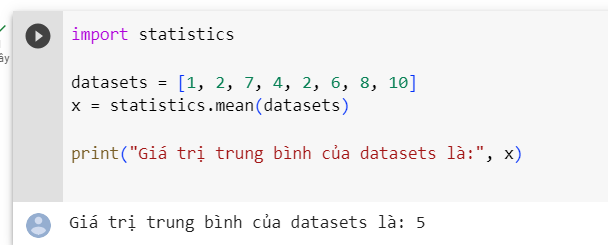
**Công dụng:** Hàm mean() trong module statistics của Python được sử dụng để tính giá trị trung bình số học của các số trong danh sách.

**Cú pháp:** statistics**.mean(**data)

Trong đó, data là một chuỗi hoặc một iterable. Giá trị trung bình được tính bằng cách chia tổng của các data point cho số data point.

**Lưu ý:** Nếu mục data bị bỏ trống sẽ gây ra lỗi StatisticsError.

**Examples:**



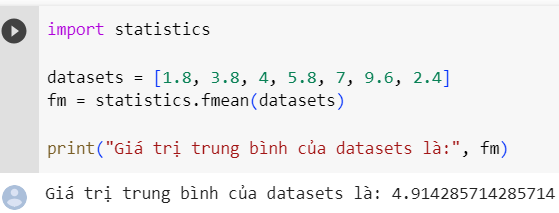
## 2. Hàm fmean():

**Công dụng:** Hàm fmean() trong module statistics của Python cũng được sử dụng để tính giá trị trung bình số học của các số trong danh sách. Tuy nhiên, hàm này có cú pháp khác so với hàm mean(). Hàm fmean() sẽ chuyển tất cả các data point trong danh sách thành kiểu dữ liệu float trước khi tính giá trị trung bình.

**Cú pháp:** statistics.**fmean**(data, weights=None)

Lưu ý: Nếu thêm trường trọng số (weights), thì trọng số phải có cùng độ dài với data (chỉ có ở các version sau 3.8 trở đi).

**Examples:**

****

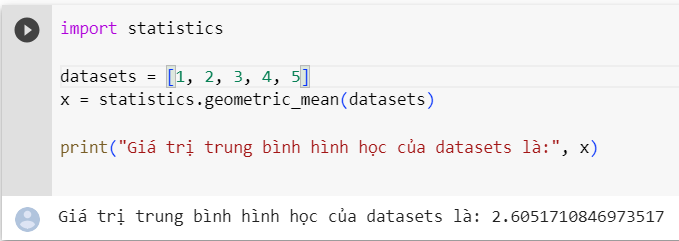
## 3. Hàm geometric\_mean():

**Công dụng:** Hàm geometric\_mean() trong module statistics của Python được sử dụng để tính trung bình hình học của một danh sách các số.

**Cú pháp:** statistics.**geometric\_mean**(data)

Trong đó, data là một chuỗi hoặc một iterable. Giá trị trung bình hình học được tính bằng cách lấy tích của các data point và lấy căn bậc n của số lượng data point.

**Examples:**



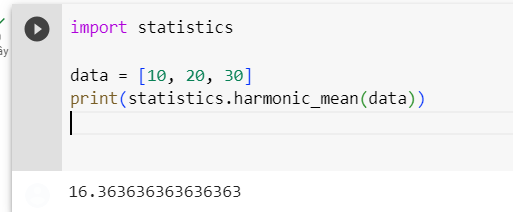
## 4. Hàm harmonic\_mean():

**Công dụng:** Hàm harmonic\_mean() trong module statistics được sử dụng để tính giá trị trung bình điều hòa của các số trong danh sách. Giá trị trung bình điều hòa được tính bằng cách lấy nghịch đảo của giá trị trung bình số học của các nghịch đảo của dữ liệu. Nó được tính như sau: Nếu bạn có bốn giá trị (a, b, c và d) - nó sẽ tương đương với 4 / (1/a + 1/b + 1/c + 1/d).

**Cú pháp:** statistics.**harmonic\_mean**(data)

Trong đó data là danh sách các số thực.

**Examples:**



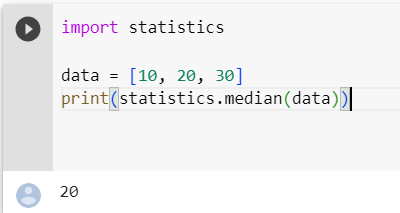
## 5. Hàm media():

**Công dụng:** Hàm median() trong module statistics được sử dụng để trả về giá trị giữa của dữ liệu số trong danh sách. Giá trị giữa được tính bằng cách sắp xếp dữ liệu theo thứ tự tăng dần và lấy giá trị ở giữa nếu số phần tử là lẻ hoặc lấy trung bình của hai giá trị ở giữa nếu số phần tử là chẵn.

**Cú pháp:** statistics.**median(**data)

Trong đó data là danh sách các số thực.

**Examples:**

****

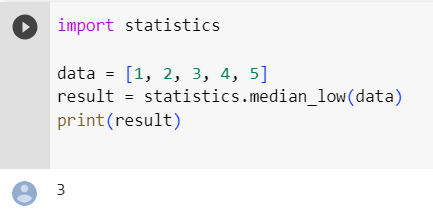
## 6. Hàm median\_low():

**Công dụng:** Hàm median\_low() trong module statistics trả về giá trị trung vị thấp nhất của dãy số đã cho. Nếu số lượng phần tử trong dãy là số chẵn thì hàm sẽ trả về giá trị ở vị trí giữa hai giá trị trung tâm.

**Cú pháp:** statistics.**median\_low**(data)

Trong đó, data là dãy số cần tính giá trị trung vị.

**Examples:**



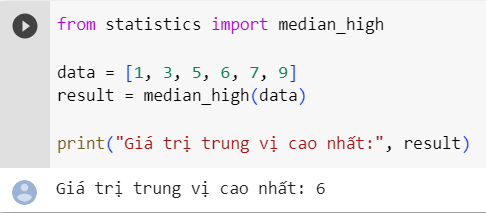
## 7. Hàm median\_high():

**Công dụng:** được sử dụng để tính giá trị trung vị cao nhất (giá trị ở giữa khi sắp xếp dãy số) của một tập hợp dữ liệu số.

**Cú pháp:** statistics**.median\_high**(data)

Trong đó data là tập hợp dữ liệu số (có thể là danh sách, tuple hoặc một loại dữ liệu dãy số khác)

**Examples:**

****

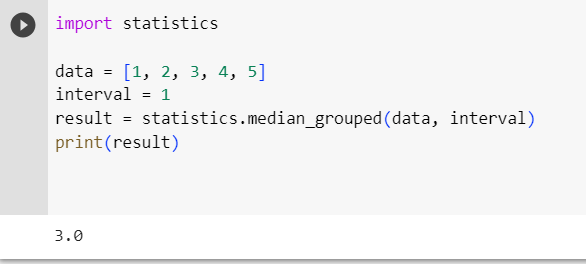
## 8. Hàm median\_grouped():

**Công dụng:** Hàm median\_grouped() trong module statistics tính giá trị trung vị của dãy số đã cho bằng phương pháp nhóm.

**Cú pháp:** statistics.**median\_grouped(**data, interval)

Trong đó, data là dãy số cần tính giá trị trung vị và interval là kích thước của mỗi nhóm.

**Examples:**

****

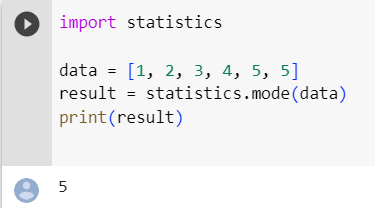
## 9. Hàm mode():

**Công dụng:** Hàm mode() trong module statistics trả về giá trị xuất hiện nhiều nhất trong dãy số đã cho.

**Cú pháp:** statistics.**mode**(data)

Trong đó, data là dãy số ,tập hợp cần tìm giá trị xuất hiện nhiều nhất.

**Examples:**

****

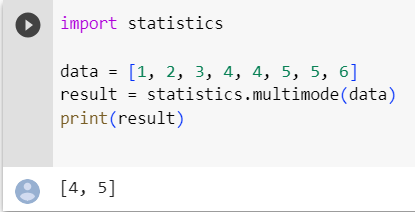
## 10. Hàm multimode():

**Công dụng:** Hàm multimode() trong module statistics trả về một danh sách chứa các giá trị xuất hiện nhiều nhất trong dãy số đã cho.

**Cú pháp:** statistics.**multimode**(data)

Trong đó, data là dãy số cần tìm giá trị xuất hiện nhiều nhất.

**Examples:**



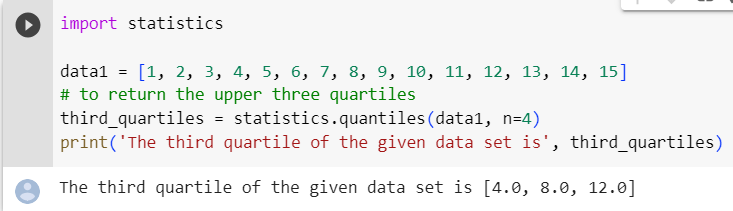
## 11. Hàm quantiles():

**Công dụng:** Hàm statistics.quantiles() trong Python được sử dụng để trả về các phân vị tương ứng với số n trong dữ liệu có thể lặp lại

**Cú pháp:** Statistics.**quantiles**(data, \*, n=4, method = ‘exclusive’)

Đặt n thành 4 cho quartiles (mặc định). Đặt n thành 10 cho deciles. Đặt n thành 100 cho phần trăm cung cấp 99 điểm cắt để phân tách data thành 100 nhóm có kích thước bằng nhau. Để có kết quả có ý nghĩa, số lượng điểm dữ liệu trong dữ liệu phải lớn hơn n .

**Examples:**

****

# Phần II: Measures of spread

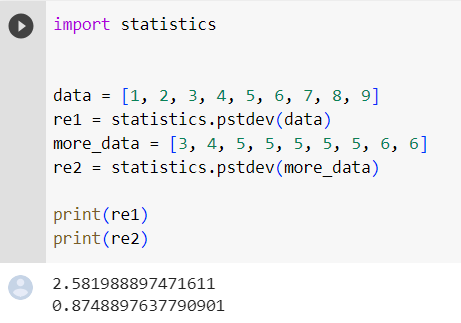
## 12. Hàm pstdev():

**Công dụng:** Trả về độ lệch chuẩn tổng thể (căn bậc hai của phương sai tổng thể).

**Cú pháp:** statistics.**pstdev**(data, mu=None)

Trong đó, data là dãy số , nếu đối số thứ hai tùy chọn mu được đưa ra, thì đó thường là giá trị trung bình của dữ liệu. Nếu nó bị thiếu hoặc None(mặc định), giá trị trung bình cộng sẽ được tính tự động.

**Examples:**

****

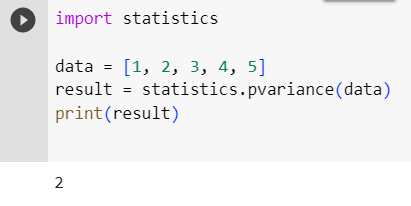
## 13. Hàm pvariance():

**Công dụng:** Để tính phương sai từ toàn bộ tổng thể..

**Cú pháp:** statistics.**pvariance**(data)

Trong đó, data là dãy số cần tính phương sai. Nếu đối số thứ hai tùy chọn mu được đưa ra, thì đó thường là giá trị trung bình của dữ liệu. Nếu nó bị thiếu hoặc None(mặc định), giá trị trung bình cộng sẽ được tính tự động.

**Examples:**

****

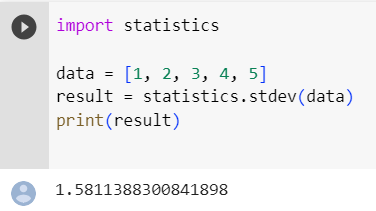
## 14. Hàm stdev():

**Công dụng:** Trả về độ lệch chuẩn mẫu (căn bậc hai của phương sai mẫu).

**Cú pháp:** statistics.**stdev**(data)

Trong đó, data là dãy số , nếu đối số thứ hai tùy chọn mu được đưa ra, thì đó thường là giá trị trung bình của dữ liệu. Nếu nó bị thiếu hoặc None(mặc định), giá trị trung bình cộng sẽ được tính tự động.

**Examples:**

****

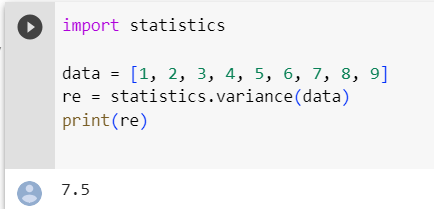
## 15. Hàm variance():

**Công dụng:** Để ước tính phương sai từ một mẫu.

**Cú pháp:** statistics.**variance**(data, xbar=None)

Trong đó, data là dãy số cần tính phương sai. Nếu đối số thứ hai tùy chọn xbar được cung cấp, thì đó phải là giá trị trung bình của dữ liệu . Nếu nó bị thiếu hoặc None(mặc định), giá trị trung bình sẽ được tính tự động.

**Examples:**



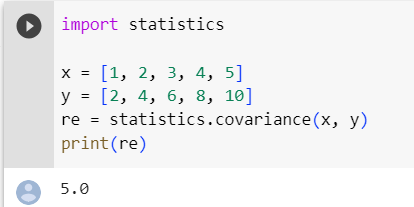
## 16. Hàm convariance():

**Công dụng:** Hàm statistics.covariance() trong Python được sử dụng để tính đồng phương sai của hai dữ liệu đầu vào. Đồng phương sai là một độ đo về mối quan hệ hướng của hai biến ngẫu nhiên trong thống kê.

**Cú pháp:** statistics.**covariance**(x, y)

Trong đó x, y là dãy số cần tính đồng phương sai và có thể lặp lại.

**Examples:**

****

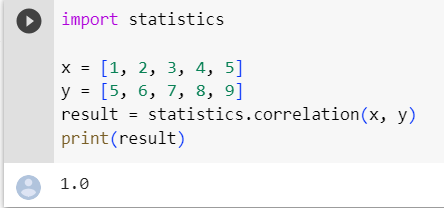
## 17. Hàm correlation():

**Công dụng:** Hàm correlation() trong module statistics tính hệ số tương quan giữa hai dãy số.

**Cú pháp:** statistics.**correlation**(x, y)

Trong đó, x và y là hai dãy số cần tính hệ số tương quan.

**Examples:**

****

## 18. Hàm linear\_regression():

**Công dụng:** Trả về độ dốc và chặn các tham số hồi quy tuyến tính đơn giản được ước tính bằng cách sử dụng các bình phương nhỏ nhất thông thường.

**Cú pháp:** statistics.**linear\_regression**(x, y, /, \*, proportional=False)

Cả hai đầu vào phải có cùng chiều dài (không nhỏ hơn hai) và biến độc lập x không thể là hằng số;

**Examples:**

# Phần III: NormalDist objects

## 19. Hàm NormalDist():

**Công dụng:** Hàm NormalDist() trong module statistics tính giá trị hàm mật độ xác suất của phân phối chuẩn.

**Cú pháp:** statistics.NormalDist(x, mean, standard\_dev)

Trong đó, x là giá trị cần tính hàm mật độ xác suất, mean là giá trị trung bình của phân phối chuẩn và standard\_dev là độ lệch chuẩn của phân phối chuẩn

**Examples:**