**ĐẠI HỌC HUẾ**

# KHOA KỸ THUẬT VÀ CÔNG NGHỆ



🙠🙟🕮🙝🙢

****

**BÁO CÁO**

**ĐỒ ÁN**

**NĂM HỌC 2021-2022**

**Giảng viên hướng dẫn: Nguyễn Đình Hoa Cương**

**Lớp: KHDL&TTNT**

|  |
| --- |
| Số phách  *(Do hội đồng chấm thi ghi)* |

**Thừa Thiên Huế, ngày 2 tháng 6 năm 2022**

**ĐẠI HỌC HUẾ**

# KHOA KỸ THUẬT VÀ CÔNG NGHỆ



🙠🙟🕮🙝🙢

****

(MẪU BÌA PHỤ)

**BÁO CÁO**

**ĐỒ ÁN**

**NĂM HỌC 2021-2022**

**Giảng viên hướng dẫn: Nguyễn Đình Hoa Cương**

**Lớp: KHDL&TTNT**

**Sinh viên thực hiện: Trần Đăng Minh Vũ**

*(ký tên và ghi rõ họ tên)*

|  |
| --- |
| Số phách  *(Do hội đồng chấm thi ghi)* |

**Thừa Thiên Huế, ngày 2 tháng 6 năm 2022**

Mục lục

[*Chương 1: Cơ Sở lý thuyết:* 4](#_Toc105783971)

[1. Thống kê là gì: 4](#_Toc105783972)

[2. Thống kê trong Python 4](#_Toc105783973)

[*Chương 2: Tiến hành thống kê:* 6](#_Toc105783974)

[PHẦN 1: THỐNG KÊ MÔ TẢ: 6](#_Toc105783975)

[PHẦN 2: THỐNG KÊ SUY DIỄN: 22](#_Toc105783976)

[Chương 3: Tài liệu tham khảo: 38](#_Toc105783977)

[Chương 4: Kiểm tra đạo văn: 39](#_Toc105783978)

# *Chương 1: Cơ Sở lý thuyết:*

1. **Thống kê là gì:**

Thống kê là một dạng phân tích toán học sử dụng các mô hình, sự biểu diễn và tóm tắt định lượng cho một tập hợp dữ liệu thực nghiệm hoặc nghiên cứu thực tế nhất định nhằm phục vụ cho quá trình phân tích, dự đoán và quyết định.

2 phương pháp thống kê được sử dụng trong đồ án bao gồm:

* Thống kê mô tả
* Thống kê suy diễn

1. Thống kê mô tả:

Thống kê mô tả được sử dụng để mô tả những đặc tính cơ bản của dữ liệu thu thập được từ nghiên cứu thực nghiệm qua các cách thức khác nhau. Thống kê mô tả và thống kê suy luận cùng cung cấp những tóm tắt đơn giản về mẫu và các thước đo. Cùng với phân tích đồ họa đơn giản, chúng tạo ra nền tảng của mọi phân tích định lượng về số liệu. Để hiểu được các hiện tượng và ra quyết định đúng đắn, cần nắm được các phương pháp cơ bản của mô tả dữ liệu

1. Thống kê suy diễn:

- Thống kê suy diễn là công cụ mà nhà thống kê sử dụng để rút ra kết luận về các đặc điểm của một quần thể từ các đặc điểm của mẫu và quyết định mức độ chắc chắn của chúng về độ tin cậy của các kết luận đó.

- Dựa trên kích thước mẫu và sự phân bố của dữ liệu mẫu, các nhà thống kê có thể tính toán xác suất để các thống kê đo lường xu hướng trung tâm, độ biến thiên, phân bố và mối quan hệ giữa các đặc điểm trong mẫu dữ liệu, cung cấp bức tranh chính xác về các thông số tương ứng của toàn bộ tổng thể từ đó mẫu được rút ra.

- Thống kê suy luận được sử dụng để khái quát hóa về các nhóm lớn, chẳng hạn như ước tính nhu cầu trung bình đối với một sản phẩm bằng cách khảo sát mẫu thói quen mua hàng của người tiêu dùng hoặc để cố gắng dự đoán các sự kiện trong tương lai, chẳng hạn như dự đoán lợi nhuận trong tương lai của một loại chứng khoán hoặc tài sản về lợi nhuận trong một khoảng thời gian mẫu.

1. **Thống kê trong Python**
2. Giới thiệu về ngôn ngữ lập trình python:

Python là một ngôn ngữ lập trình bậc cao phục vụ cho những mục đích lập trình đa năng,

do Guido van Rossum tạo ra và ra mắt vào năm 1991.

Python có ưu điểm mạnh là dễ đọc, dễ học và dễ nhớ. Python là ngôn ngữ có hình thức rất sáng sủa, cấu trúc rất rõ ràng, rất thuận tiện cho người mới học lập trình.

Cấu trúc của Python còn có thể cho phép người sử dụng viết mã lệnh với số lần gõ phím tối thiểu.

1. Giới thiệu về các thư viện:

* **Thư viện Pandas**

- Pandas là một Thư viện Python mã nguồn mở cung cấp công cụ phân tích và thao tác dữ liệu hiệu suất cao bằng cách sử dụng các cấu trúc dữ liệu mạnh mẽ của nó. Cái tên Pandas có nguồn gốc từ từ Panel Data - một Kinh tế lượng từ dữ liệu Đa chiều.

- Năm 2008, nhà phát triển Wes McKinney bắt đầu phát triển gấu trúc khi cần công cụ phân tích dữ liệu hiệu suất cao, linh hoạt.

- Trước Pandas, Python chủ yếu được sử dụng để trộn và chuẩn bị dữ liệu. Nó có rất ít đóng góp vào việc phân tích dữ liệu. Gấu trúc đã giải quyết vấn đề này. Sử dụng Pandas, chúng ta có thể thực hiện năm bước điển hình trong quá trình xử lý và phân tích dữ liệu, bất kể nguồn gốc của dữ liệu - tải, chuẩn bị, thao tác, lập mô hình và phân tích.

- Python with Pandas được sử dụng trong nhiều lĩnh vực bao gồm các lĩnh vực học thuật và thương mại bao gồm tài chính, kinh tế, thống kê, phân tích,…

* **Thư viên Scipy**

- SciPy, một thư viện khoa học cho Python là một thư viện mã nguồn mở, được cấp phép BSD cho toán học, khoa học và kỹ thuật.

- SciPy là một tập hợp các thuật toán toán học và các hàm tiện lợi được xây dựng dựa trên phần mở rộng NumPy của Python. Nó bổ sung sức mạnh đáng kể cho phiên Python tương tác bằng cách cung cấp cho người dùng các lệnh và lớp cấp cao để thao tác và trực quan hóa dữ liệu.

- Lợi ích bổ sung của việc dựa trên SciPy trên Python là điều này cũng tạo ra một ngôn ngữ lập trình mạnh mẽ có sẵn để sử dụng trong việc phát triển các chương trình phức tạp và các ứng dụng chuyên biệt. Các ứng dụng khoa học sử dụng SciPy được hưởng lợi từ việc phát triển các mô-đun bổ sung trong nhiều ngóc ngách của bối cảnh phần mềm bởi các nhà phát triển trên toàn thế giới. Mọi thứ từ lập trình song song đến các chương trình con và lớp cơ sở dữ liệu và web đều đã được cung cấp cho lập trình viên Python. Tất cả sức mạnh này đều có sẵn ngoài các thư viện toán học trong SciPy.

* **Thư viện Matplotlyb**

- Matplotlib là một trong những gói Python phổ biến nhất được sử dụng để trực quan hóa dữ liệu. Nó là một thư viện đa nền tảng để tạo các đồ thị 2D từ dữ liệu trong các mảng.

- Matplotlib được viết bằng Python và sử dụng NumPy, phần mở rộng toán học số của Python. Nó cung cấp một API hướng đối tượng giúp nhúng các lô trong các ứng dụng sử dụng bộ công cụ GUI Python như PyQt, WxPythonotTkinter. Nó có thể được sử dụng trong Python và IPython shell, máy tính xách tay Jupyter và các máy chủ ứng dụng web.

- Matplotlib có một giao diện thủ tục được đặt tên là Pylab, được thiết kế giống với MATLAB, một ngôn ngữ lập trình độc quyền được phát triển bởi MathWorks. Matplotlib cùng với NumPy có thể được coi là mã nguồn mở tương đương với MATLAB.

- Matplotlib ban đầu được viết bởi John D. Hunter vào năm 2003. Phiên bản ổn định hiện tại là 2.2.0 được phát hành vào tháng 1 năm 2018.

# *Chương 2: Tiến hành thống kê:*

## PHẦN 1: THỐNG KÊ MÔ TẢ:

* Phần phân tích thống kê mô tả viết các mô tả về tập dữ liệu Inc\_Exp\_Data.
* Trình bày các kết quả gồm các phép tính thống kê mô tả và các hình vẽ liên quan đến tập dữ liệu Inc\_Exp\_Data.
* Trình bày nhận xét rút ra được của về kết quả phân tích thống kê mô tả.

Tiến hành thống kê mô tả:

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

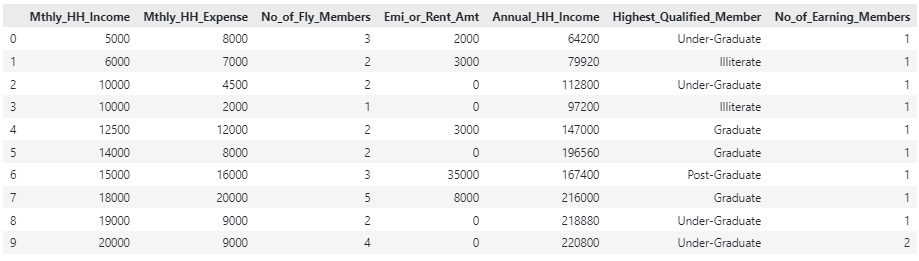
* **NHẬP DỮ LIỆU:**

*►Import các thư viện sử dụng để tiến hành thống kê mô tả:*



*►Đọc file CSV vào dataframe bằng thư viện Pandas:*

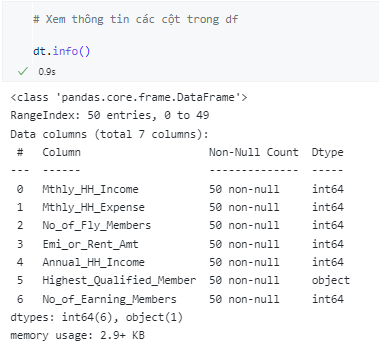


►*Sau khi đọc file CSV, in ra 10 hàng đầu của dataframe để kiểm tra các trường dữ liệu của dataframe:*

→Diễn giải ý nghĩa của các trường dữ liệu:

|  |  |
| --- | --- |
| Tên cột: | Ý nghĩa cột: |
| Mthly\_HH\_Income | Monthly Household Income: Thu nhập gia đình hàng tháng |
| Mthly\_HH\_Expense | Monthly Household Expense: Chi tiêu gia đình hàng tháng |
| No\_of\_Fly\_Members | Number of Family Members: Số thành viên trong gia đình |
| Emi\_or\_Rent\_Amt | Monthly EMI (Equated Monthly Installment) or Rent Amount: Khoản trả định kì. |
| Annual\_HH\_Income | Annual Household Income: Thu nhập hộ gia đình hàng năm |
| Highest\_Qualified\_Member | Highest Qualified Member: Trình độ học vấn của thành viên có trình độ cao nhất trong gia đình |
| No\_of\_Earning\_Members | Number of Earning Members: Số thành viên có thu nhập |

*►Dùng df.info( ) để hiển thị kích thước của dataframe:*



* **TIẾN HÀNH PHÂN TÍCH**
* *2 cột dữ liệu (Mthly\_HH\_Income) và (Mthly\_HH\_Expense) phản ánh thu chi của các hộ gia đình.*

Để tránh dữ liệu về thu nhập và chi tiêu phân bố không đều, ta tiến hành phân nhóm các hộ gia đình theo mức thu nhập, chi tiêu thông qua việc thêm vào 2 cột cho dataframe: (Type\_In) biểu thị mức thu nhập, (Type\_Out) biểu thị mức chi tiêu.

Quy ước:

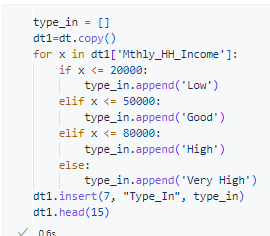
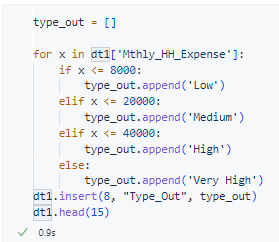
(Type\_In)

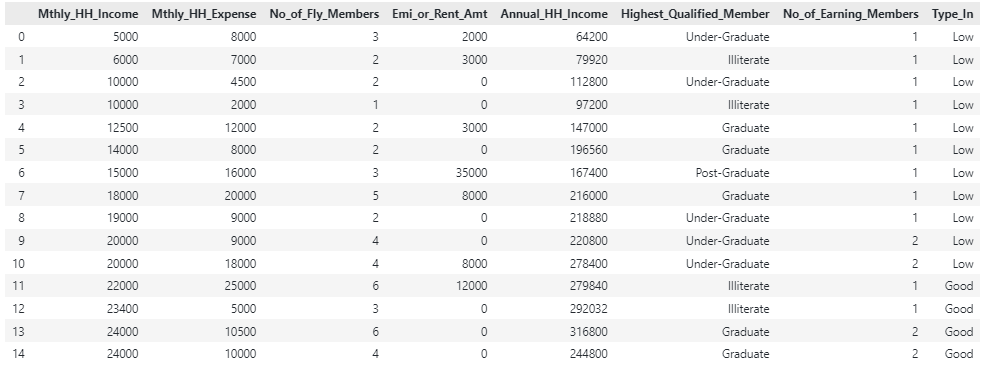
* ≤ 20,000$ → Thu nhập thấp (Low)
* ≤ 50,000$ → Thu nhập tốt (Good)
* ≤ 80,000$ → Thu nhập cao (High)
* > 80,000$ → Thu nhập rất cao (Very High)

(Type\_Out)

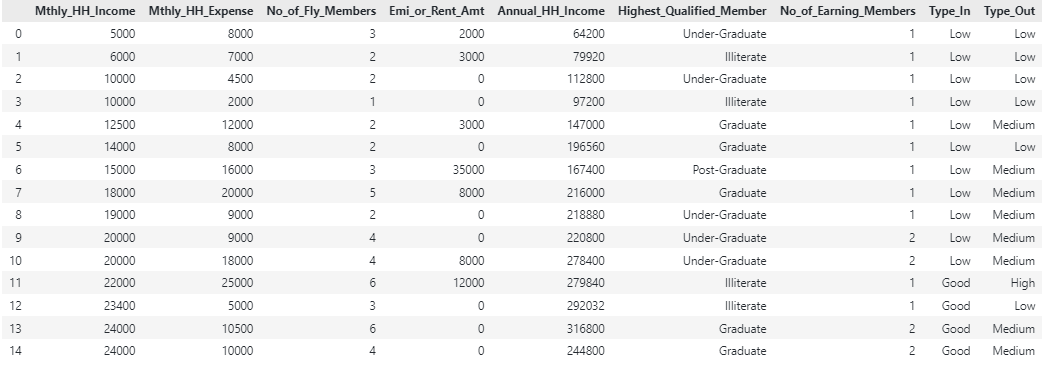
* ≤ 8,000$ → Chi tiêu thấp (Low)
* ≤ 20,000$ → Chi tiêu trung bình (Medium)
* ≤ 40,000$ → Chi tiêu cao (High)
* > 40,000$ → Chi tiêu rất cao (Very High)

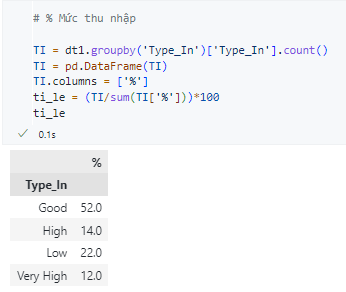
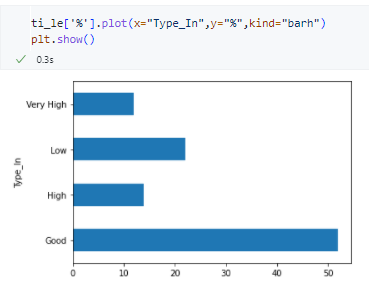
*►Tiến hành thêm dữ liệu:*



*►Type In*:

*►Type Out*:

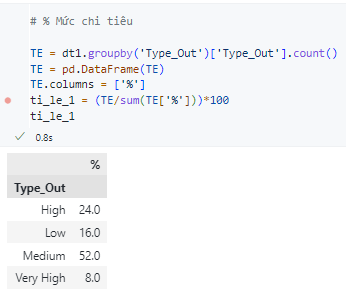
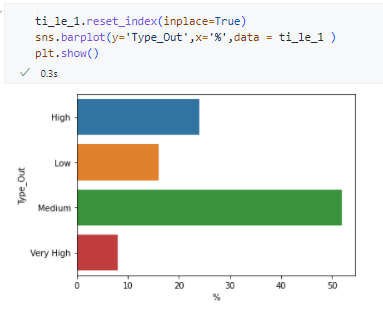
*►Bảng thống kê phần trăm hộ gia đình thông qua mức thu nhập*



**→Biểu đồ:**

* ***Qua bảng thống kê và biểu đồ thể hiện, ta thấy được hộ gia đình có thu nhập tốt chiếm tỉ lệ cao nhất (52%), hộ gia đình có thu nhập rất cao chiếm tỉ lệ thấp nhất (12%).***

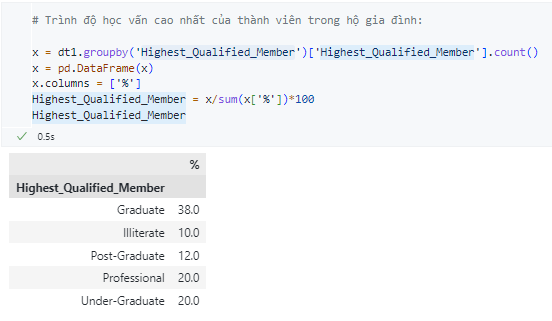
*►Bảng thống kê phần trăm hộ gia đình thông qua mức chi tiêu*

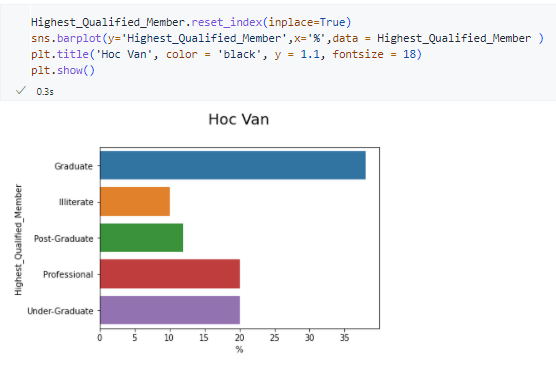


**→Biểu đồ:**

* ***Qua bảng thống kê và biểu đồ thể hiện, ta thấy được hộ gia đình có mức chi tiêu trung bình chiếm tỉ lệ cao nhất (52%), hộ gia đình có chi tiêu rất cao chiếm tỉ lệ thấp nhất (8%).***
* **Nhận xét về 2 biểu đồ trên cho ta thấy được mức thu nhập hàng tháng sẽ quyết định chi tiêu của các hộ gia đình, với thu nhập tốt có thể ổn định mức chi tiêu ở ngưỡng trung bình, với thu nhập cao sẽ sống với mức chi tiêu cao hơn,…**
* *Cột dữ liệu (Highest\_Qualified\_Member) thể hiện trình độ học vấn cao nhất của thành viên trong hộ gia đình:*

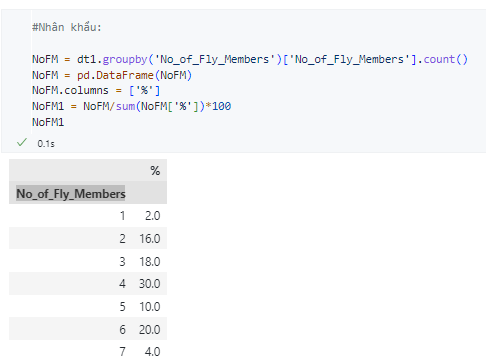
►*Tiến hành thống kê loại trình độ học vấn theo mức %:*



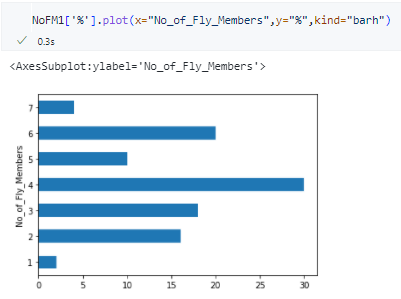
**→Biểu đồ:** 

* ***Qua bảng thống kê và biểu đồ thể hiện, ta thấy hộ gia đình có thành viên với trình độ học vấn cao nhất là Graduate (Tốt nghiệp đại học, cao đẳng) chiếm tỉ lệ lớn, trong khi đó hộ gia đình không có thành viên với trình độ học vấn (Illiterate, mù chữ) chiếm tỉ trọng thấp nhất.***
* *Cột dữ liệu (No\_of\_Fly\_Members) cho biết số nhân khẩu của các hộ gia đình:*

►*Tiến hành thống kê theo mức %:*

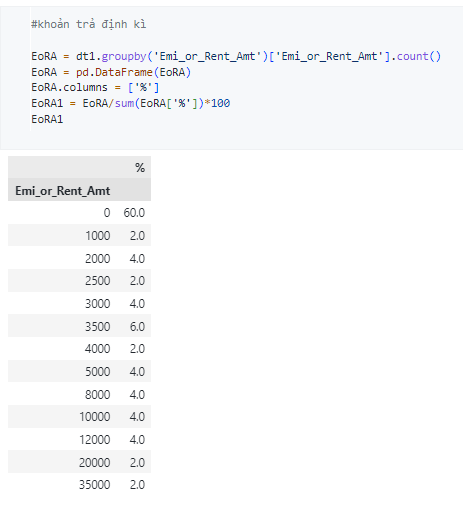


**→Biểu đồ:**

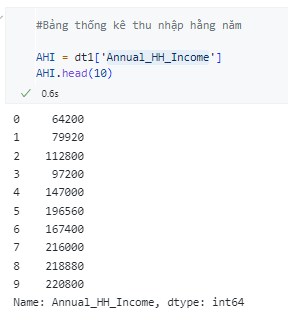


* ***Qua bảng thống kê và biểu đồ thể hiện, ta thấy hộ gia đình có 4 thành viên chiếm tỉ lệ lớn nhất.***
* *Cột dữ liệu (Emi\_or\_Rent\_Amt) cho biết khoản phí phải nộp thêm định kì:*

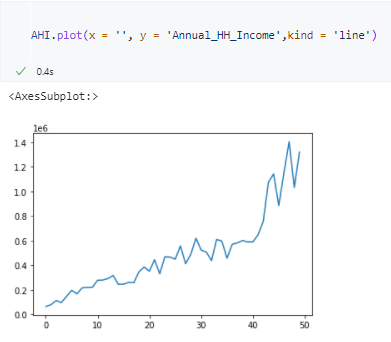
*►Tiến hành thống kê theo mức %:*



* *Thống kê thu nhập hàng năm (Annual\_HH\_Income):*

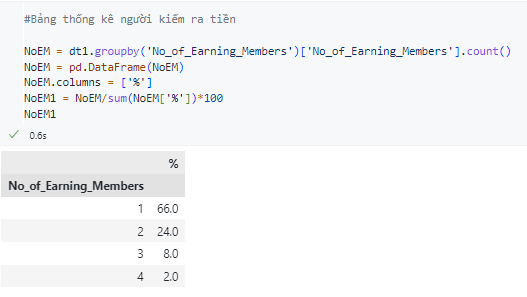
►

**→Biểu đồ:**

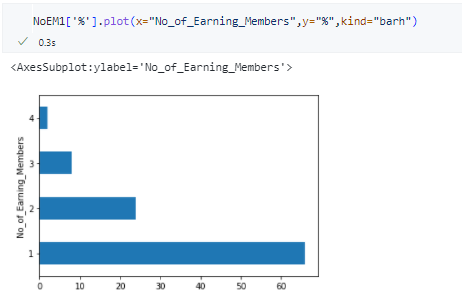


* *Cột dữ liệu (No\_of\_Earning\_Members) cho biết số thành viên kiếm ra tiền trong gia đình:*

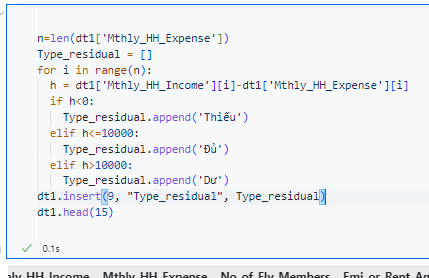
*►Tiến hành thống kê theo mức %:*

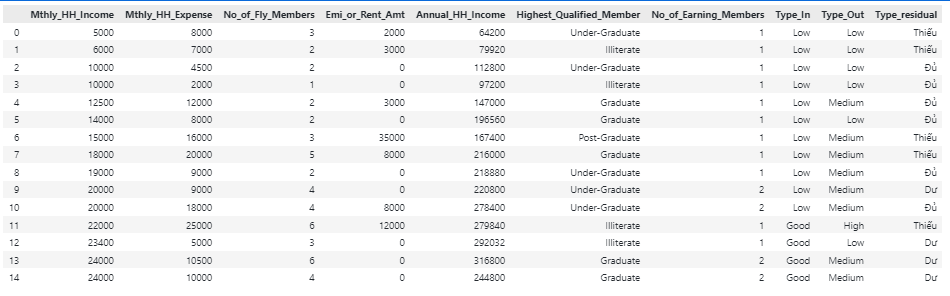


**→Biểu đồ:**

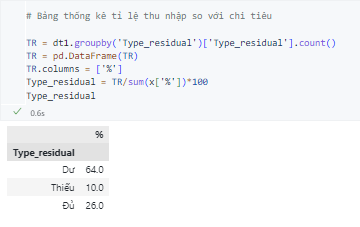


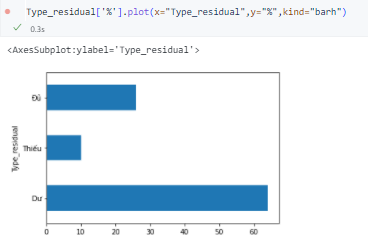
* *Tiến hành khảo sát tình trạng kinh tế của 1 gia đình bằng cách so sánh thu (Type\_In) và chi (Type\_Out) hàng tháng:*

►*Thêm cột Type\_residual vào dataframe, ý nghĩa tượng trưng cho tình trạng kinh tế của hộ gia đình, tính bằng hiệu của thu và chi:*



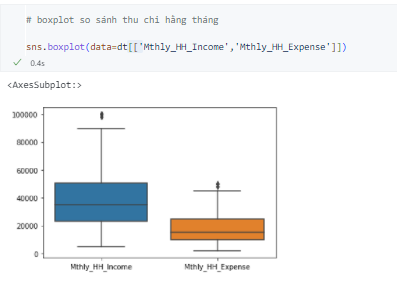
►*Thống kê theo %:*



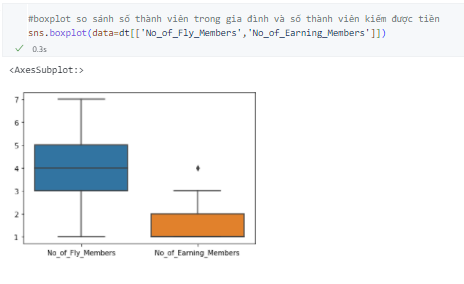
**→Biểu đồ:**

* ***Qua bảng thống kê và biểu đồ thể hiện, ta thấy hộ gia đình có có kinh tế dư dả chiếm tỉ lệ cao nhất.***
* *Các biểu đồ khác:*

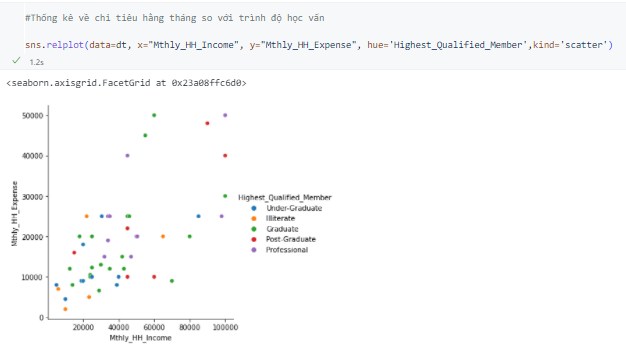
*►Boxplot so sánh thu chi hằng tháng*



*► Boxplot so sánh số thành viên trong gia đình và số thành viên kiếm ra tiền*

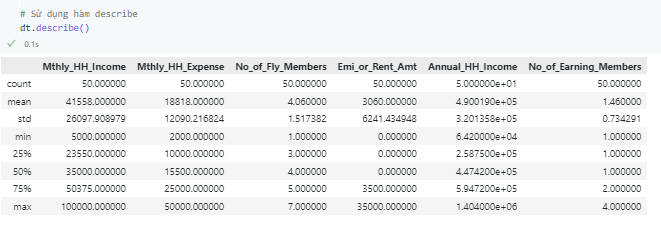


*► Biểu đồ chi tiêu hằng tháng so với trình độ học vấn:*

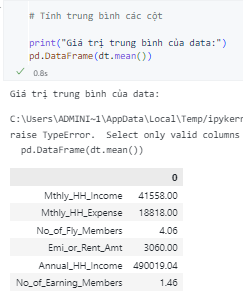


* ***Qua biểu đồ ta nhận thấy chi tiêu không qua phụ thuộc vào trình độ học vấn.***
* **XU HƯỚNG TẬP TRUNG CỦA DỮ LIỆU:**

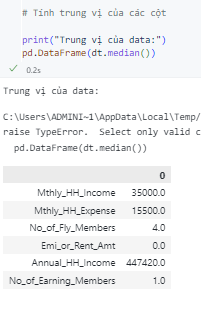
*Sử dụng hàm describe() để có một bảng tóm tắt những thống kê cơ bản của dữ liệu. Hàm tính toán các giá trị như số lượng cá thể khảo sát, trung bình, độ lệch chuẩn, min, max, bách phân vị, …*



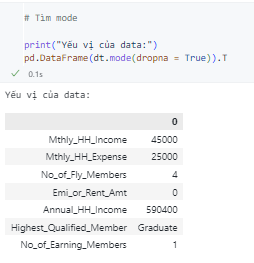
*►Trung bình:*



*►Trung vị:*

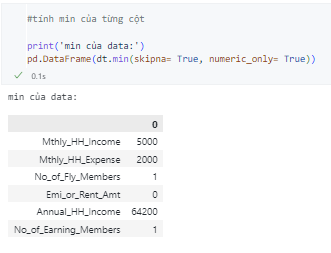


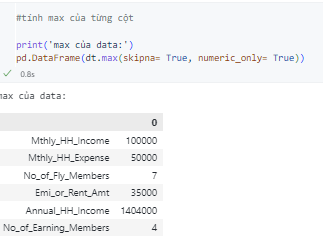
*►Yếu vị:*



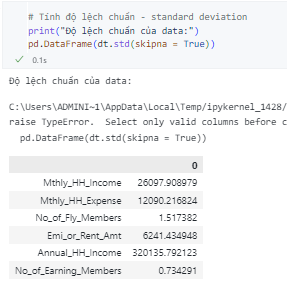
* **ĐỘ PHÂN TÁN CỦA DỮ LIỆU:**

*►Tính min max:*

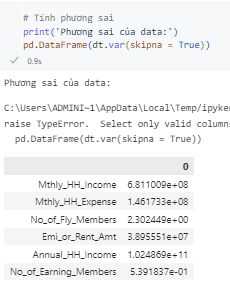




*►Tính độ lệch chuẩn:*



*►Tính phương sai:*



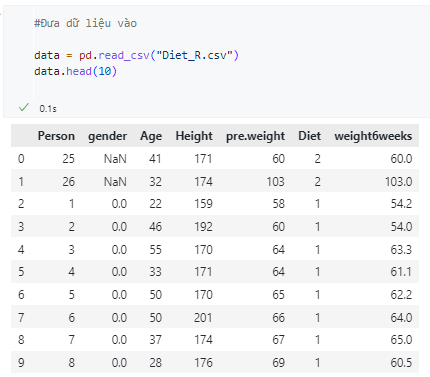
## PHẦN 2: THỐNG KÊ SUY DIỄN:

* Phần phân tích thống kê suy diễn trình bày quy trình phân tích ANOVA một chiều, phát biểu câu hỏi nghiên cứu (research question), phát biểu giả thuyết nghiên cứu, tiến hành kiểm định, kết quả kiểm định, và kết luận về kết quả kiểm định.
* Có thể them vào các hình vẽ (nếu cần thiết)

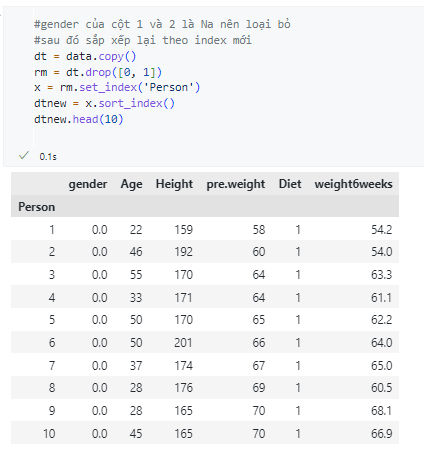
►Tiến hành import các thư việc cần thiết:



►Đọc dữ liệu vào dataframe:



►Vì dữ liệu còn giá trị trống, nên dùng hàm drop để loại bỏ các giá trị trống khỏi dataframe:



►Tạo các hàm:



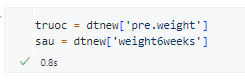
Tiến hành thống kê suy diễn:

1. *Kiểm định T Test:*

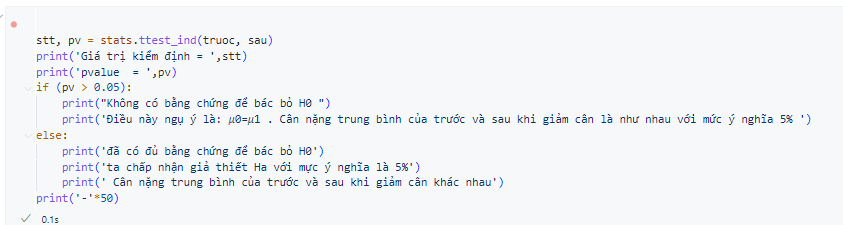
- Vậy chúng ta có giả thiết không và đối thiết như sau:

* H0: Cân nặng trung bình của mọi người trước và sau khi ăn kiêng là như nhau.
* H1: Cân nặng trung bình của mọi người trước và sau khi ăn kiêng khác nhau.

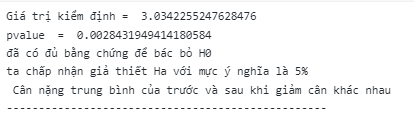
►Trích xuất dữ liệu của cột “pre.weight” và “weight6weeks”.



►Tiến hành kiểm định T Test:

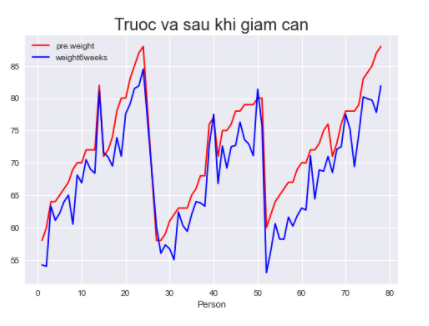


►Kết quả:



* + ***Như vậy với mức tin cậy là 5% chúng ta có đủ tự tin để khẳng định rằng sau khi thực hiện phương pháp giảm cân là có hiệu quả.***

►Biểu đồ trước và sau giảm cân:



1. *Kiểm định ANOVA:*
   1. **Câu hỏi nghiên cứu 1: Có sự khác biệt hiệu quả giữa các phương pháp giảm cân hay không?**

- Xây dựng giả thiết:

* H0: Các phương pháp giảm cân cho hiệu quả là như nhau.
* H1: Các phương pháp giảm cân cho hiệu quả khác nhau.

Để kiểm chứng giả thiết chúng ta cần trích xuất dữ liệu để phân tích, phân loại dữ liệu và tiến hành phân tích ANOVA.

►Lấy 2 cột Diet và lostweigt để tiến hành phân tích:



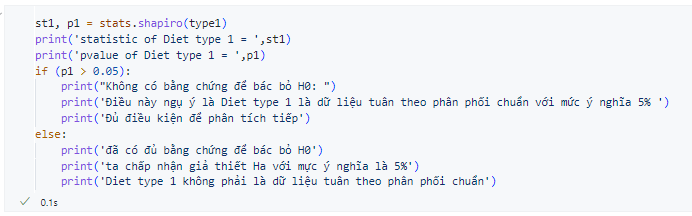
- Để thực hiện kiểm định ANOVA thì dữ liệu phải thỏa mãn các yêu cầu sau:

* Dữ liệu phải tuân theo phân phối chuẩn. (1)
* Dữ liệu đồng nhất về phương sai . (2)
* Các cá thể phải được lấy độc lập với nhau. (3)

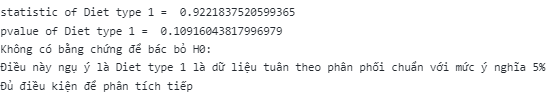
***(1)► để kiểm tra tính chất đầu tiên của dữ liệu, ta sử dụng kiểm định Shapiro-Wilk:***

**VỚI DIET TYPE 1:**

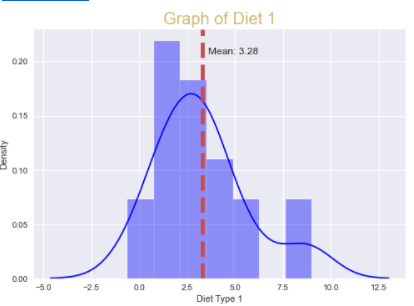
* H0: Diet Type 1 là dữ liệu tuân theo phân phối chuẩn
* H1: Diet Type 1 không phải là dữ liệu tuân theo phân phối chuẩn.

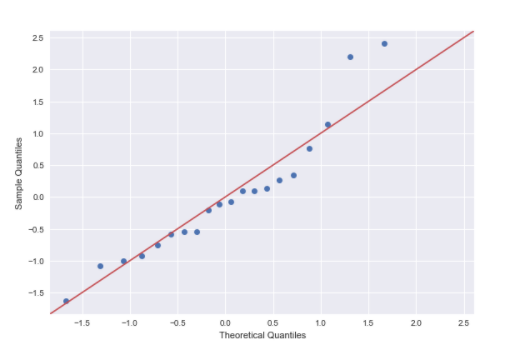


►Kết quả:



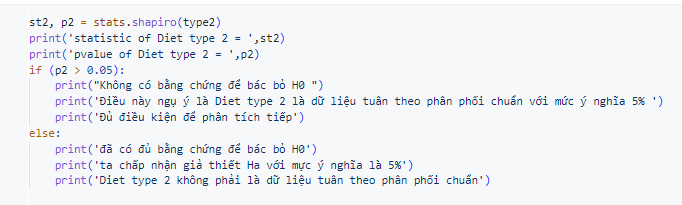
►Histogram và biểu đồ đường:



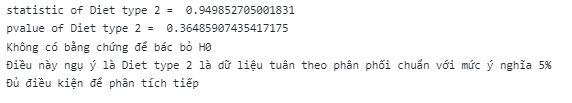


**VỚI DIET TYPE 2:**

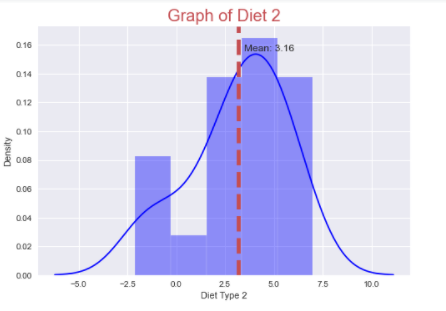
* H0: Diet Type 2 là dữ liệu tuân theo phân phối chuẩn
* H1: Diet Type 2 không phải là dữ liệu tuân theo phân phối chuẩn.

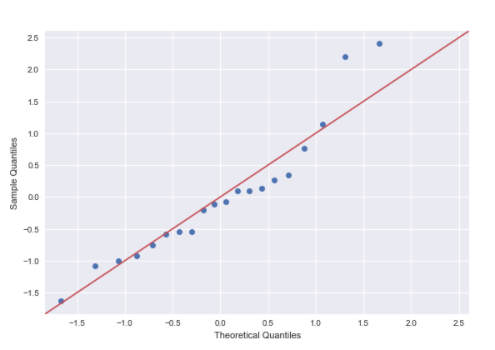


►Kết quả:



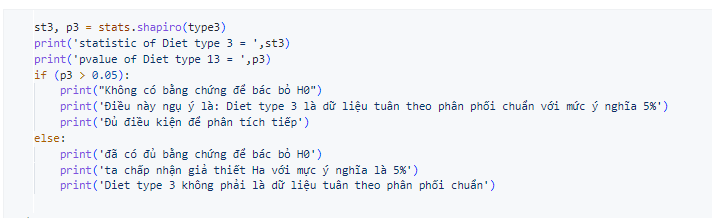
►Histogram và biểu đồ đường:



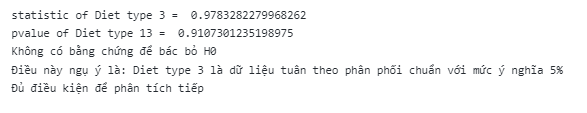


**VỚI DIET TYPE 3:**

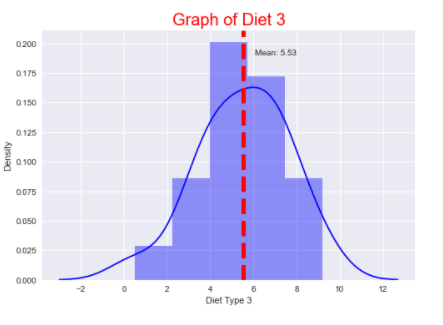
* H0: Diet Type 3 là dữ liệu tuân theo phân phối chuẩn
* H1: Diet Type 3 không phải là dữ liệu tuân theo phân phối chuẩn.

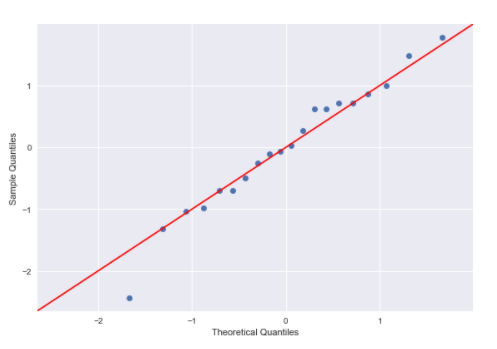


►Kết quả:



►Histogram và biểu đồ đường:





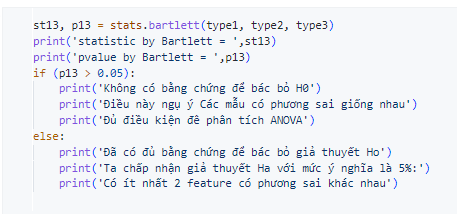
***(2)► để kiểm tra tính chất thứ 2 của dữ liệu, có nhiểu phương pháp để kiểm định như kiểm định Bartlett, kiểm định Levene, …***

- Phát biểu giả thiết:

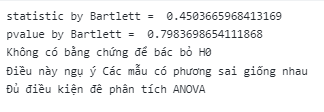
* H0: Các feature đồng nhất phương sai.
* H1: Các feature không đồng nhất phương sai.

Sử dụng kiểm định Bartlett: Phát biểu lại giả thuyết của kiểm định Barlett:

* H0: Các feature đều đồng nhất phương sai.
* H1: Có ít nhất 2 feature có phương sai khác nhau.

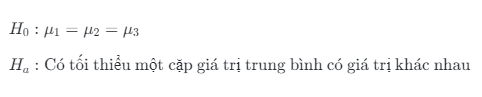


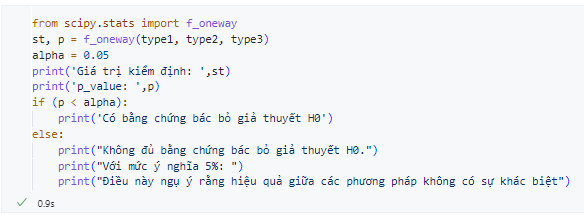
►Kết quả:



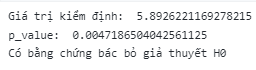
***► Tiến hành kiểm định ANOVA***

Phát biểu lại câu hỏi nghiên cứu 1:



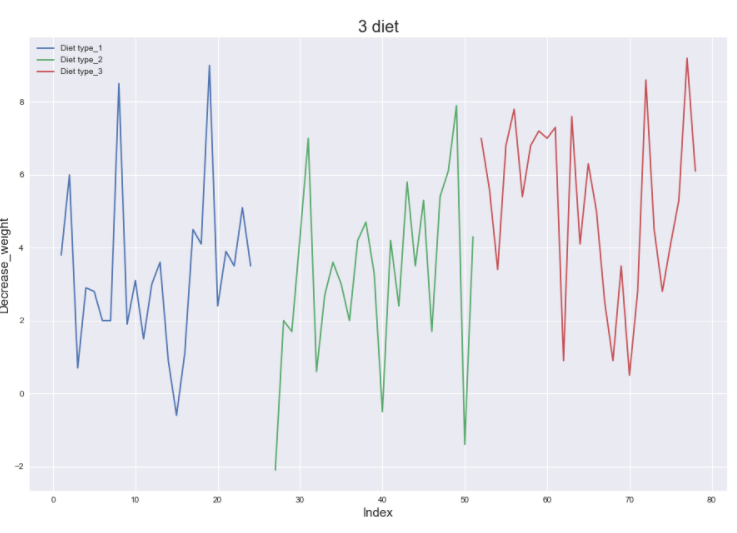


***►***Kết quả:



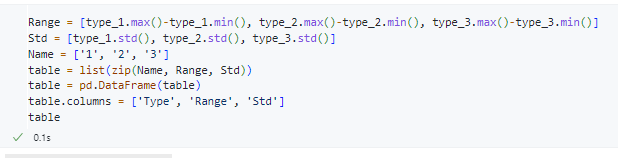
* + ***Như vậy chúng ta có đủ điều kiện để khẳng định các phương pháp giảm cân khác nhau sẽ cho ra hiệu quả khác nhau.***

***►***Biểu đồ:

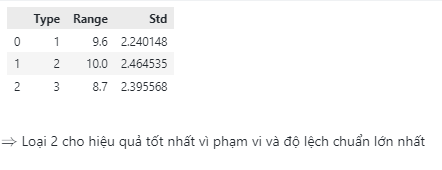


* 1. **Câu hỏi nghiên cứu 2: Phương pháp giảm cân nào cho hiệu quả tốt nhất?**

- Để trả lời được câu hỏi này chúng ta dựa vào độ phân tán của dữ liệu, tức là pham vi và độ lệch chuẩn của từng phương pháp ăn kiêng. Phương pháp nào cho độ lệch chuẩn lớn, phạm vi dữ liệu lớn tức là cho hiệu quả cao và ngược lại.



***►***Kết quả:



* 1. **Câu hỏi nghiên cứu 3: Liệu hiệu quả giảm cân có ảnh hưởng đến giới tính không?**

- Xây dựng giả thiết:

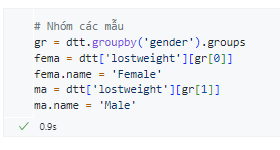
* H0: Hiệu quả giảm cân của nam và nữ là như nhau.
* H1: Hiệu quả giảm cân của nam và nữ là khác nhau.

- Để thực hiện kiểm định ANOVA thì dữ liệu phải thỏa mãn các yêu cầu sau:

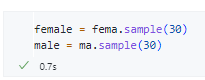
* Dữ liệu phải tuân theo phân phối chuẩn. (1)
* Dữ liệu đồng nhất về phương sai . (2)
* Các cá thể phải được lấy độc lập với nhau. (3)

Tương tự như câu hỏi nghiên cứu 1, tính chất 1 và 2 sẽ được kiểm tra để làm điều kiện tiến hành phân tích ANOVA:

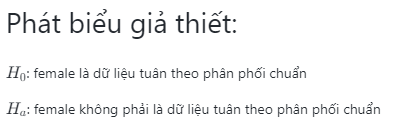
►Tiến hành nhóm các mẫu theo giới tính:



►Vì các mẫu không có kích thước bằng nhau nên lấy ngẫu nhiên 30 giá trị trong mỗi mẫu:



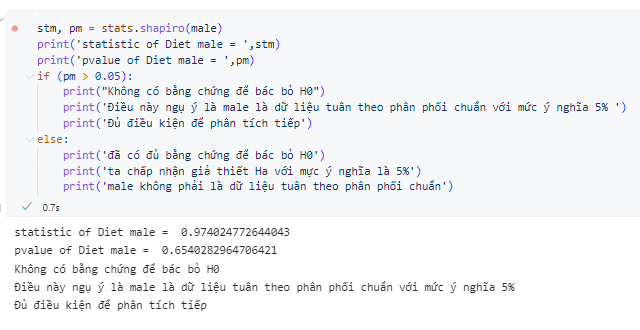
***(1)► để kiểm tra tính chất đầu tiên của dữ liệu, ta sử dụng kiểm định Shapiro-Wilk:***



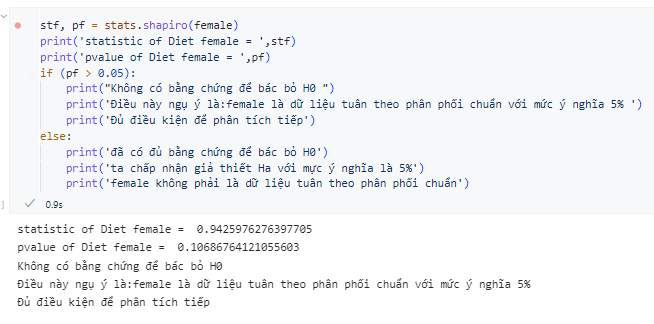


***Tương tự như kiểm định tính chất (1) của câu hỏi nghiên cứu 1, ta có các kiểm định và kết quả sau:***

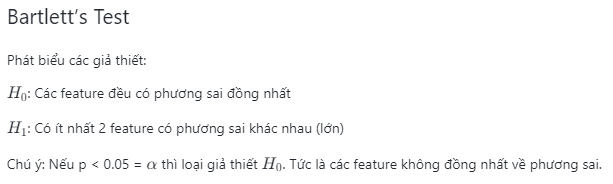
**VỚI MALE:**

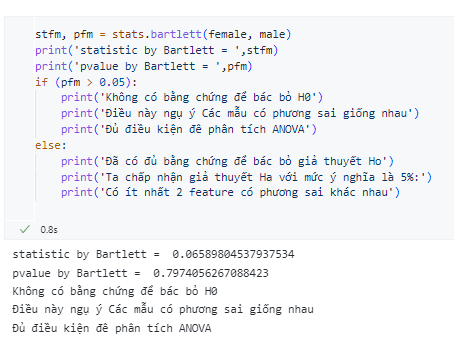


**VỚI FEMALE:**



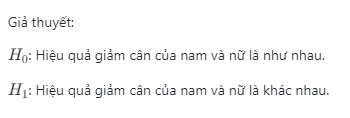
***(2)► để kiểm tra tính chất thứ 2 của dữ liệu, ta sử dụng kiểm định Barlett:***





**Tiến hành kiểm định ANOVA**

Phát biểu lại câu hỏi nghiên cứu 3:





►Kết quả:



* + ***Như vậy chúng ta có đủ điều kiện để khẳng định rằng giới tính không ảnh hưởng đến hiệu quả giảm cân.***

# CHƯƠNG 3: TÀI LIỆU THAM KHẢO:

<https://pandas.pydata.org/docs/>

<https://www.scipy.org/docs.html>

<https://matplotlib.org/>

<https://stackoverflow.com/>

<https://vietnambiz.vn/thong-ke-statistics-la-gi-phan-loai-thong-ke-2020061716451094.htm>

<https://vi.wikipedia.org/wiki/Th%E1%BB%91ng_k%C3%AA_m%C3%B4_t%E1%BA%A3>

<https://vi.wikipedia.org/wiki/Th%E1%BB%91ng_k%C3%AA#%E1%BB%A8ng_d%E1%BB%A5ng>

<https://en.wikipedia.org/wiki/Python_(programming_language)>

# CHƯƠNG 4: KIỂM TRA ĐẠO VĂN: