

ĐẠI HỌC BÁCH KHOA HÀ NỘI
KHOA TOÁN-TIN



**BÀI GIỮA KÌ
SUY LUẬN THỐNG KÊ**

Giảng viên hướng dẫn: **TS. Nguyễn Thị Thu Thủy**

Mã học phần: **MI3031**

Mã lớp học: **163641**

Sinh viên thực hiện: **Hà Đỗ Anh Tú**

Mã số sinh viên: **20237491**

Hà Nội - 2025

Lời mở đầu

Trước tiên, em xin gửi lời cảm ơn chân thành đến cô **Nguyễn Thị Thu Thủy** – giảng viên bộ môn **Suy luận thống kê** – người đã tận tâm giảng dạy, truyền đạt cho em những kiến thức quý báu và tiết học bổ ích về môn học Suy luận thống kê. Nhờ sự hướng dẫn tận tình của cô, em đã có nền tảng vững chắc để tiếp cận và ứng dụng các phương pháp thống kê vào việc giải quyết các vấn đề trong thực tế.

Suy luận thống kê là một môn học nền tảng quan trọng trong các lĩnh vực khoa học dữ liệu, kỹ thuật và quản lý. Môn học này không chỉ trang bị cho em các công cụ để phân tích và diễn giải dữ liệu, mà còn rèn luyện tư duy phản biện, khả năng định lượng hóa sự không chắc chắn và kỹ năng ra quyết định dựa trên bằng chứng.

Trong khuôn khổ bài giữa kỳ này, em đã vận dụng các kỹ thuật suy luận thống kê cốt lõi — bao gồm ước lượng khoảng tin cậy và kiểm định giả thuyết vào một bài toán kinh doanh thực tế. Cụ thể là phân tích dữ liệu khảo sát để đánh giá tính khả thi của một dự án F&B tiềm năng tại khu vực Đại học Bách khoa Hà Nội - là một dự án em đã áp ủ muôn làm từ lâu mà chưa có đủ cơ sở để thực hiện.

Do kiến thức và kinh nghiệm còn hạn chế, bài giữa kỳ của em chắc chắn không tránh khỏi những thiếu sót. Em rất mong nhận được ý kiến đóng góp của cô để hoàn thiện bài làm này.

Em xin chân thành cảm ơn!

Mục lục

Lời mở đầu	i
Mục lục	ii
1 GIỚI THIỆU	1
1.1 Bối cảnh và Vấn đề	1
1.2 Mục tiêu kinh doanh và Giả thuyết nghiên cứu	1
1.2.1 Xác định mốc hòa vốn (μ_0)	1
1.2.2 Mục tiêu thống kê của Bài toán	2
2 CƠ SỞ LÝ THUYẾT	3
2.1 Giới thiệu	3
2.2 Thống kê mô tả	3
2.2.1 Các đại lượng đo lường	3
2.2.2 Trực quan hóa dữ liệu	3
2.3 Ước lượng Khoảng tin cậy cho Trung bình	3
2.3.1 Nguyên lý và Ý nghĩa	4
2.3.2 Công thức KTC cho trung bình (σ chưa biết)	4
2.4 Kiểm định giả thuyết	4
2.4.1 Nguyên lý chung	4
2.4.2 Giá trị p-value (p-value)	4
2.5 Lựa chọn Tiêu chuẩn Kiểm định	5
2.5.1 Quy tắc ra quyết định	5
3 PHÂN TÍCH DỮ LIỆU VÀ KẾT QUẢ	7
3.1 Phương pháp thu thập và Mô tả Dữ liệu	7
3.1.1 Kết quả Thống kê mô tả	7
3.2 Kết quả Ước lượng Khoảng tin cậy 95%	8
3.3 Kết quả Kiểm định giả thuyết	9
3.3.1 Quyết định Thống kê	9
4 PHÂN TÍCH KẾT QUẢ VÀ RA QUYẾT ĐỊNH	10
4.1 Diễn giải Kết quả thống kê	10
4.1.1 Phân tích Ý nghĩa của Kiểm định Giả thuyết	10
4.1.2 Phân tích Ý nghĩa của Khoảng tin cậy 95%	10
4.2 Quyết định Kinh doanh: Nên kinh doanh	11
4.3 Hạn chế của Nghiên cứu và Hướng phát triển	11
4.4 Kết luận chung	12
Tài liệu tham khảo	13

Phụ lục	13
A Dữ liệu khảo sát thô	14
B Mã nguồn Phân tích (Python)	16

Chương 1

GIỚI THIỆU

1.1 Bối cảnh và Vấn đề

Với tư cách là sinh viên chuyên ngành Hệ thống thông tin quản lý tại Đại học Bách khoa Hà Nội (HUST), cùng tinh thần nhiệt huyết và mong muốn khởi nghiệp, em nhận thấy rằng thị trường dịch vụ ăn uống (F&B) xung quanh khu vực Đại học Bách khoa Hà Nội (HUST) là một trong những thị trường tiềm năng. Với quy mô tuyển sinh hàng năm lên tới hơn 10.000 sinh viên, tổng số lượng sinh viên, giảng viên và nhân viên tại trường tạo ra một nhu cầu khổng lồ, ổn định và lặp lại hàng ngày về ăn uống.

Tuy nhiên, thị trường này cũng tồn tại nhiều thách thức. Qua quan sát sơ bộ, em nhận thấy hai đặc điểm chính:

1. **Mức độ cạnh tranh cao:** Có rất nhiều nhà cung cấp dịch vụ ăn uống, từ các quán cơm bình dân, xe đẩy đồ ăn vặt, đến các chuỗi cửa hàng tiện lợi.
2. **Tính nhạy cảm về giá:** Đối tượng khách hàng chủ yếu là sinh viên, một nhóm khách hàng có ngân sách hạn chế và rất nhạy cảm với các quyết định chi tiêu.

Điều này dẫn đến một nhận định rằng, bất kỳ một mô hình F&B mới nào muốn thành công và chiếm lĩnh thị phần trong khu vực này đều phải lấy **ưu thế về giá** làm yếu tố cạnh tranh bắt buộc.

1.2 Mục tiêu kinh doanh và Giả thuyết nghiên cứu

Với bối cảnh trên, em đang xem xét một ý tưởng khởi nghiệp: xây dựng một mô hình F&B tinh gọn (ví dụ: một quầy bán đồ ăn mang đi) với khả năng tối ưu hóa chi phí vận hành để cung cấp các suất ăn chất lượng với mức giá cạnh tranh nhất.

Tuy nhiên, để mô hình này có lãi, một bản kế hoạch kinh doanh sơ bộ (dựa trên chi phí thuê mặt bằng, nguyên vật liệu, nhân công) đã được vạch ra. Phân tích này chỉ ra rằng, dự án **chỉ khả thi về mặt tài chính nếu tổng chi tiêu trung bình hàng tháng cho ăn uống của sinh viên (trung bình tổng thể μ) vượt qua một mốc hòa vốn tối thiểu**.

1.2.1 Xác định mốc hòa vốn (μ_0)

Dựa trên các tính toán chi phí, mô hình F&B giả định này chỉ có thể tồn tại và bắt đầu sinh lãi nếu quy mô thị trường đủ lớn. Em xác định mốc an toàn tối thiểu cho chi tiêu trung bình của một sinh viên là:

$$\mu_0 = 2.0 \text{ triệu VNĐ/tháng}$$

Nếu mức chi tiêu trung bình thực tế của toàn bộ sinh viên HUST (μ) thấp hơn hoặc bằng con số này, thị trường được coi là không đủ cung tiền để một mô hình mới (không kể các yếu tố khách quan trong kinh tế) có thể thành công.

1.2.2 Mục tiêu thống kê của Bài toán

Mục tiêu của bài giັງ kì này là sử dụng các công cụ suy luận thống kê để kiểm định giả thuyết kinh doanh trên. Cụ thể, em sẽ tiến hành một cuộc khảo sát trên mẫu sinh viên HUST để kiểm định cặp giả thuyết sau:

- **Giả thuyết H_0** : $\mu = 2.0$

(Phát biểu: Chi tiêu trung bình thực tế bằng 2.0 triệu. Thị trường không khả thi, quyết định: Hủy bỏ dự án).

- **Đối thuyết H_1** : $\mu > 2.0$

(Phát biểu: Chi tiêu trung bình thực tế lớn hơn 2.0 triệu. Thị trường khả thi, quyết định: Tiếp tục triển khai dự án).

Bài toán này sẽ được kiểm định ở mức ý nghĩa $\alpha = 0.05$.

Chương 2

CƠ SỞ LÝ THUYẾT

2.1 Giới thiệu

Chương này trình bày cơ sở lý thuyết về suy luận thống kê được sử dụng để phân tích dữ liệu. Các phương pháp này là nền tảng để từ dữ liệu mẫu, em có thể đưa ra các kết luận về tổng thể và ra quyết định kinh doanh.

Các kỹ thuật chính được sử dụng bao gồm: Thống kê mô tả, Ước lượng khoảng tin cậy và Kiểm định giả thuyết thống kê.

2.2 Thống kê mô tả

Thống kê mô tả là bộ các phương pháp nhằm mục đích tóm tắt và trình bày các đặc điểm chính của một tập dữ liệu. Trước khi thực hiện bất kỳ suy luận nào, việc hiểu rõ dữ liệu mẫu là bước bắt buộc.

2.2.1 Các đại lượng đo lường

- **Đo lường xu hướng trung tâm:** Cho biết giá trị "diễn hình" của dữ liệu. Các đại lượng phổ biến là **Trung bình mẫu** (\bar{x}) và **Trung vị** (Med).
- **Đo lường độ phân tán:** Cho biết mức độ lan rộng hoặc biến động của dữ liệu. Đại lượng quan trọng nhất là **Độ lệch chuẩn mẫu hiệu chỉnh** (s).

2.2.2 Trực quan hóa dữ liệu

Các biểu đồ trực quan giúp phát hiện các xu hướng, quy luật phân phối và các giá trị bất thường.

- **Biểu đồ Histogram:** Giúp nhận diện hình dạng phân phối của dữ liệu (ví dụ: chuẩn, lệch trái, lệch phải).
- **Biểu đồ Hộp (Boxplot):** Hiệu quả trong việc tóm tắt 5 vị trí (Min, Q1, Med, Q3, Max) và phát hiện các giá trị ngoại lai.

2.3 Ước lượng Khoảng tin cậy cho Trung bình

Trong thực tế, ta không thể biết giá trị chính xác của trung bình tổng thể μ (ví dụ: chỉ tiêu trung bình *thực sự* của toàn bộ sinh viên HUST). Thay vào đó, ta dùng dữ liệu mẫu để xây dựng một "khoảng" mà ta tin rằng nó chứa μ .

2.3.1 Nguyên lý và Ý nghĩa

Một **khoảng tin cậy** với độ tin cậy $1 - \alpha$ là một khoảng ngẫu nhiên $(\hat{\theta}_L, \hat{\theta}_U)$ được xây dựng từ mẫu. Trước khi lấy mẫu, xác suất để khoảng này "bắt" được tham số θ (là μ) đúng bằng $1 - \alpha$.

Điễn giải: Khi ta nói "Khoảng tin cậy 95%", điều đó có nghĩa là nếu ta lặp lại quy trình lấy mẫu và xây dựng khoảng này 100 lần, thì trung bình sẽ có 95 khoảng "bắt" trùng giá trị μ thật. Đó là sự tin cậy vào **phương pháp**, chứ không phải xác suất cho một khoảng cụ thể. [1]

2.3.2 Công thức KTC cho trung bình (σ chưa biết)

Trong báo cáo giữa kì này, em không biết độ lệch chuẩn của tổng thể (σ). Do đó, em sử dụng độ lệch chuẩn mẫu hiệu chỉnh (s) và **phân phối t-Student**.

Công thức khoảng tin cậy $1 - \alpha$ cho trung bình tổng thể μ là:

$$\bar{x} - t_{\alpha/2,n-1} \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{\alpha/2,n-1} \frac{s}{\sqrt{n}}$$

Hoặc viết gọn là:

$$\bar{x} \pm t_{\alpha/2,n-1} \frac{s}{\sqrt{n}}$$

2.4 Kiểm định giả thuyết

Kiểm định giả thuyết là một quy trình thống kê chuẩn mực dùng để ra quyết định giữa hai giả thuyết đối lập nhau (chấp nhận hay bác bỏ một nhận định) dựa trên bằng chứng từ dữ liệu mẫu.

2.4.1 Nguyên lý chung

Quy trình luôn bắt đầu với việc phát biểu hai giả thuyết:

- **Giả thuyết 0 (H_0):** Là giả thuyết ban đầu. Đây là giả thuyết ta sẽ mặc định là đúng cho đến khi có bằng chứng đủ mạnh để bác bỏ nó.
- **Đối thuyết (H_1):** Là điều ta muốn chứng minh (ví dụ: có sự thay đổi, có sự khác biệt, hoặc giá trị lớn hơn/nhỏ hơn một mốc nào đó).

Mức ý nghĩa (α): Là xác suất tối đa mà ta chấp nhận mắc **Sai lầm loại I** (bác bỏ H_0 trong khi H_0 đúng). Trong kinh doanh và nghiên cứu, $\alpha = 0.05$ (5%) thường được sử dụng làm ngưỡng tiêu chuẩn.

2.4.2 Giá trị p-value (p-value)

p-value là khái niệm cốt lõi để ra quyết định. Nó là xác suất, *giả sử H_0 là đúng*, ta quan sát được một kết quả mẫu "lạ" bằng hoặc "lạ" hơn kết quả mà ta đã thu thập được.

- *p-value* nhỏ (ví dụ: $p < 0.05$): Bằng chứng mẫu rất "lạ", khó có thể xảy ra nếu H_0 đúng. Ta có bằng chứng mạnh để **bác bỏ H_0** và chấp nhận H_1 .
- *p-value* lớn (ví dụ: $p \geq 0.05$): Bằng chứng mẫu là "bình thường", hoàn toàn có thể xảy ra nếu H_0 đúng. Ta **không đủ bằng chứng để bác bỏ H_0** . [2]

2.5 Lựa chọn Tiêu chuẩn Kiểm định

Đây là bước quan trọng nhất trong việc áp dụng lý thuyết vào thực tế. Dựa trên các kiến thức đã học (Chương 3), em có bảng tổng hợp các trường hợp kiểm định giả thuyết cho kỳ vọng μ như sau:

Bảng 2.1: Các trường hợp kiểm định giả thuyết về kỳ vọng μ (với $H_0 : \mu = \mu_0$).

Điều kiện thông tin	Thống kê kiểm định	Phân phối (khi H_0 đúng)
1. Đã biết phương sai σ^2	$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$	Chuẩn tắc $N(0, 1)$
2. Chưa biết σ^2 , mẫu lớn ($n \geq 40$)	$Z \approx \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$	Xấp xỉ $N(0, 1)$
3. Chưa biết σ^2 , mẫu chưa đủ lớn ($n < 40$)	$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$	Student $t(n - 1)$

Áp dụng vào đề tài: Trong báo cáo này, em thực hiện khảo sát thực tế nên **chưa biết phương sai tổng thể σ^2** . Mặc dù kích thước mẫu $n = 35$ có thể coi là đủ lớn để dùng xấp xỉ Z (Trường hợp 2), nhưng theo lý thuyết, trong thực tế, để đảm bảo độ chính xác và an toàn cao nhất về mặt thống kê, báo cáo lựa chọn sử dụng **Trường hợp 3 (Kiểm định t một mẫu)**. [2]

Thống kê kiểm định được sử dụng là:

$$T_0 = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

Giá trị T_0 này tuân theo phân phối t-Student với $n - 1$ bậc tự do.

Quy trình ra quyết định: Vì mục tiêu là kiểm tra xem chi tiêu trung bình có **lớn hơn** 2.0 triệu hay không, em sử dụng phép kiểm định một phía bên phải:

- $H_0 : \mu = 2.0$ (Thị trường không khả thi)
- $H_1 : \mu > 2.0$ (Thị trường khả thi)

Một phần lý thuyết quan trọng nữa cho việc kiểm định và ra quyết định, là mặc dù giả thuyết H_0 được nêu bằng dấu " $=$ ", nhưng nó được hiểu là bao gồm bất kỳ giá trị nào của μ không được chỉ định bởi đối thuyết [2]. Do đó, khi ta bác bỏ H_0 , ta không chỉ bác bỏ giá trị cụ thể 2.0 triệu, mà còn bác bỏ toàn bộ tập hợp các giá trị μ mà H_0 bao hàm ($\mu \leq 2.0$).

2.5.1 Quy tắc ra quyết định

Để đảm bảo tính chặt chẽ và khách quan, báo cáo sử dụng song song hai phương pháp tiếp cận tương đương nhau được trình bày trong giáo trình [2]:

Phương pháp 1: Tiếp cận cỗ điển (Miền bác bỏ)

Phương pháp này dựa trên việc xác định trước một miền giá trị được gọi là miền bác bỏ W_α dựa trên mức ý nghĩa α .

- **Xác định giá trị tối hạn:** Với mức ý nghĩa α và bậc tự do $df = n - 1$, ta tìm giá trị tối hạn $t_{\alpha,n-1}$ từ bảng phân phối Student sao cho $P(T > t_{\alpha,n-1}) = \alpha$.
- **Xác định miền bắc bỏ:** Đối với bài toán kiểm định phía phải ($H_1 : \mu > \mu_0$), miền bắc bỏ nằm ở đuôi phải của phân phối:

$$W_\alpha = (t_{\alpha,n-1}; +\infty)$$

- **Quy tắc quyết định:**

- Nếu $t_0 \in W_\alpha$ (tức là $t_0 > t_{\alpha,n-1}$): Bắc bỏ H_0 .
- Nếu $t_0 \notin W_\alpha$: Chưa đủ cơ sở bác bỏ H_0 .

Phương pháp 2: Tiếp cận p-giá trị (p-value)

Phương pháp này đánh giá sức mạnh của bằng chứng mẫu bằng cách tính xác suất.

- **Định nghĩa:** P-giá trị là xác suất để thống kê T nhận giá trị lớn hơn hoặc bằng giá trị quan sát t_0 , với giả định H_0 đúng.
- **Công thức xác định:** Đối với kiểm định phía phải ($H_1 : \mu > \mu_0$):

$$\text{P-value} = P(T > t_0)$$

trong đó T là biến ngẫu nhiên tuân theo phân phối Student với $n - 1$ bậc tự do.

- **Quy tắc quyết định:**

- Nếu P-value $< \alpha$: Bắc bỏ H_0 (Kết quả có ý nghĩa thống kê).
- Nếu P-value $\geq \alpha$: Chưa đủ cơ sở bác bỏ H_0 .

Trong báo cáo này, với mục tiêu kiểm định $H_1 : \mu > 2.0$ (kiểm định phía phải), miền bắc bỏ sẽ là $W_\alpha = (t_{\alpha,n-1}; +\infty)$.

Chương 3

PHÂN TÍCH DỮ LIỆU VÀ KẾT QUẢ

3.1 Phương pháp thu thập và Mô tả Dữ liệu

Để thực hiện mục tiêu nghiên cứu, em đã giả định sẽ có một cuộc khảo sát đã được tiến hành trên một mẫu ngẫu nhiên gồm $n = 35$ sinh viên đang theo học tại Đại học Bách khoa Hà Nội. Dữ liệu thu thập bao gồm các thông tin định danh (Họ tên; Khoa, trường; Năm học) và biến định lượng chính của nghiên cứu: "Chi tiêu hàng tháng cho việc ăn uống" (đơn vị: triệu VND).

Toàn bộ dữ liệu thu được trình bày chi tiết tại Phụ lục B. Phân tích trong chương này sẽ tập trung vào biến **ChiTieu**.

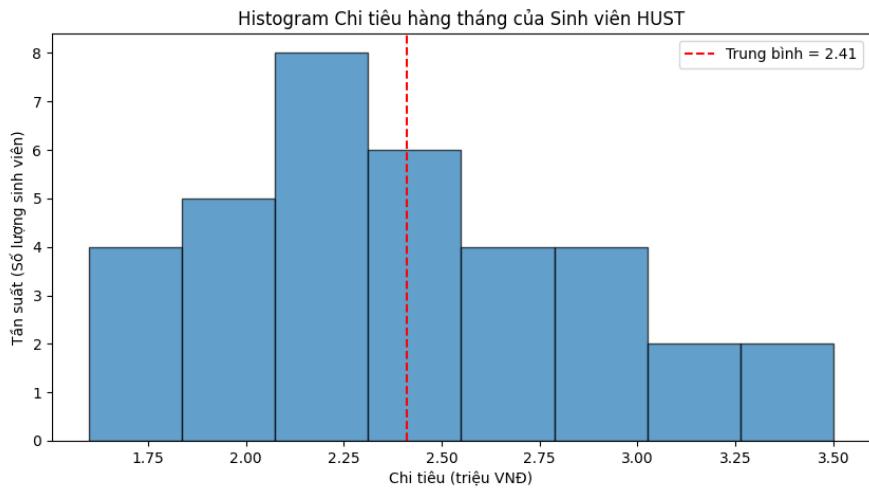
3.1.1 Kết quả Thống kê mô tả

Các giá trị thống kê mô tả cơ bản của mẫu được tính toán bằng thư viện **Pandas** trong Python và được trình bày trong Bảng 3.1.

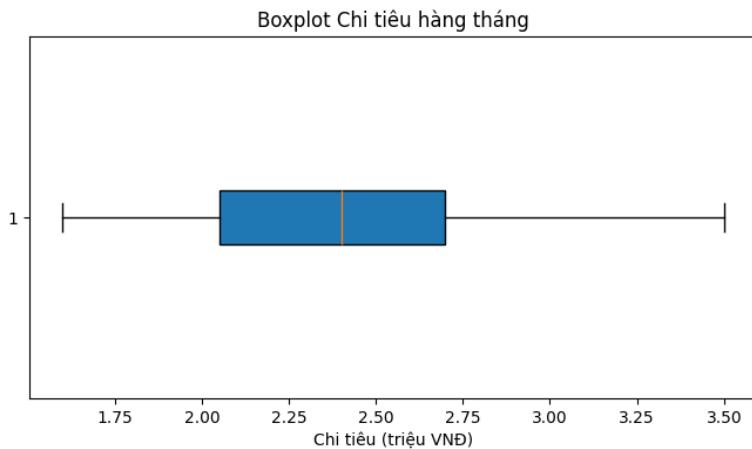
Bảng 3.1: Các giá trị thống kê mô tả của dữ liệu chi tiêu.

Đại lượng	Giá trị
Kích thước mẫu (n)	35
Trung bình mẫu (\bar{x})	2.411 triệu VND
Độ lệch chuẩn mẫu hiệu chỉnh (s)	0.482 triệu VND
Trung vị (Median)	2.400 triệu VND
Tứ phân vị thứ nhất (Q1)	2.050 triệu VND
Tứ phân vị thứ ba (Q3)	2.700 triệu VND
Giá trị nhỏ nhất (Min)	1.600 triệu VND
Giá trị lớn nhất (Max)	3.500 triệu VND

Để trực quan hóa phân phối của dữ liệu chi tiêu, biểu đồ Histogram (Hình 3.1) và biểu đồ Hộp (Hình 3.2) được sử dụng.



Hình 3.1: Histogram phân phối chi tiêu hàng tháng của sinh viên HUST.



Hình 3.2: Boxplot biểu diễn phân phối chi tiêu.

Nhận xét sơ bộ: Biểu đồ Histogram cho thấy dữ liệu phân phối tương đối đối xứng, có tâm phân phối tập trung quanh mốc 2.4 triệu đồng. Biểu đồ Boxplot cũng xác nhận điều này và không cho thấy sự xuất hiện của các giá trị ngoại lai rõ rệt.

3.2 Kết quả Ước lượng Khoảng tin cậy 95%

Sử dụng các giá trị thống kê mô tả đã tính ($\bar{x} = 2.411$, $s = 0.482$, $n = 35$) và công thức khoảng tin cậy cho trung bình tổng thể μ khi σ chưa biết, ta có:

Với bậc tự do $n - 1 = 34$, giá trị tối hạn $t_{\alpha/2, 34}$ cho độ tin cậy 95% là $t_{0.025, 34} \approx 2.032$.

Kết quả tính toán (em sử dụng `scipy.stats.t.interval`) cho ra **Khoảng tin cậy 95% cho μ** là:

$$(2.246, 2.577)$$

Diễn giải: Kết quả này có nghĩa là, với độ tin cậy 95%, chúng ta có thể kết luận rằng chi tiêu trung bình **thực sự** (trung bình tổng thể μ) cho việc ăn uống của toàn bộ sinh viên HUST nằm trong khoảng từ 2.246 triệu VND đến 2.577 triệu VND mỗi tháng.

3.3 Kết quả Kiểm định giả thuyết

Đây là phần phân tích để ra quyết định kinh doanh. Em thực hiện kiểm định t-test một mẫu, một phía (phía phải) với mốc hòa vốn $\mu_0 = 2.0$ triệu VND. Lý giải thêm cho việc sử dụng kiểm định một phía bên phải là vì mục tiêu của em là xác định xem chi tiêu trung bình có **lớn hơn** mốc 2.0 triệu hay không, từ đó đánh giá tính khả thi của thị trường.

- **Giả thuyết H_0 :** $\mu = 2.0$ (*Chi tiêu trung bình bằng 2.0 triệu, thị trường không khả thi*).
- **Đối thuyết H_1 :** $\mu > 2.0$ (*Chi tiêu trung bình cao hơn 2.0 triệu, thị trường khả thi*).
- **Mức ý nghĩa:** $\alpha = 0.05$.

Kết quả tính toán từ mẫu được trình bày trong Bảng 3.2.

Bảng 3.2: Kết quả kiểm định t-test một mẫu.

Đại lượng	Giá trị
Giá trị kiểm định (H_0)	$\mu_0 = 2.0$
Thống kê kiểm định (T_0)	5.051
Bậc tự do	34
P-value (một phía)	0.00000738

3.3.1 Quyết định Thống kê

Ta có thể đưa ra kết luận dựa trên hai phương pháp tương đương:

Phương pháp 1: Tiếp cận theo phương pháp cỗ diễn

Với mức ý nghĩa $\alpha = 0.05$ và bậc tự do 34, giá trị tối hạn cho kiểm định phía phải là $t_{\alpha,34} \approx 1.691$. Miền bác bỏ là:

$$W_\alpha = (1.691; +\infty)$$

Quan sát thấy giá trị thống kê kiểm định $T_0 = 5.051$ rơi vào miền bác bỏ ($5.051 > 1.691$). Do đó, ta bác bỏ H_0 .

Phương pháp 2: Tiếp cận theo p-value

Giá trị p-value tính được là 0.00000738. Vì $p\text{-value} < \alpha$ ($0.00000738 < 0.05$), ta có bằng chứng thống kê rất mạnh để bác bỏ H_0 .

Chương 4

PHÂN TÍCH KẾT QUẢ VÀ RA QUYẾT ĐỊNH

Trong Chương 3, em đã tiến hành các phân tích thống kê trên bộ dữ liệu khảo sát. Kết quả tính toán cho thấy trung bình mẫu $\bar{x} = 2.411$ triệu VND, khoảng tin cậy 95% là (2.246, 2.577) và giá trị p-value cho kiểm định $H_1 : \mu > 2.0$ là 0.00000738.

Chương này sẽ tập trung vào việc diễn giải các con số thống kê này trong bối cảnh thực tế của bài toán kinh doanh đã đặt ra: "Liệu thị trường F&B tại HUST có đủ tiềm năng (với $\mu > 2.0$ triệu VND) để triển khai một mô hình kinh doanh cạnh tranh về giá hay không?".

4.1 Diễn giải Kết quả thống kê

Kết quả tính toán không chỉ là những con số, chúng cung cấp những bằng chứng cụ thể để trả lời cho câu hỏi nghiên cứu.

4.1.1 Phân tích Ý nghĩa của Kiểm định Giả thuyết

Kết quả quan trọng nhất của phân tích là từ phép kiểm định t-test một phía:

Giá trị p-value = 0.00000738

Theo định nghĩa, p-value là xác suất quan sát được một giá trị trung bình mẫu "lạ" như $\bar{x} = 2.411$ (hoặc lạ hơn), *nếu giả sử rằng giả thuyết H_0 là đúng* (tức là μ thực sự chỉ là 2.0 triệu).

0.00000738 là một con số gần như bằng không. Điều này có nghĩa là: nếu chi tiêu trung bình thực sự của sinh viên HUST chỉ là 2.0 triệu, thì việc em khảo sát 35 sinh viên và ra được con số trung bình lên tới 2.411 triệu là một sự kiện "cực kỳ hiếm", gần như không thể xảy ra do ngẫu nhiên.

Vì $p\text{-value}(0.00000738) << \alpha(0.05)$, em có một bằng chứng thống kê rất mạnh mẽ để **bắc bỏ giả thuyết H_0** .

Nói cách khác, em có thể kết luận với độ tin cậy cao rằng nhận định "chi tiêu trung bình của sinh viên HUST lớn hơn 2.0 triệu VND" là đúng.

4.1.2 Phân tích Ý nghĩa của Khoảng tin cậy 95%

Kiểm định giả thuyết chỉ trả lời "Có" hoặc "Không" (lớn hơn 2.0 hay không). Khoảng tin cậy sẽ cho chúng ta biết "Lớn hơn bao nhiêu?".

Kết quả tính toán cho ra **Khoảng tin cậy 95% cho μ là (2.246, 2.577) triệu VNĐ**.

Việc diễn giải khoảng này cung cấp hai thông tin giá trị cho quyết định kinh doanh:

- Xác nhận kết quả kiểm định:** Toàn bộ khoảng tin cậy (từ 2.246 đến 2.577) đều nằm **trên** mốc hòa vốn 2.0 triệu. Điều này cung cấp thêm cho quyết định bác bỏ H_0 .
- Cung cấp biên an toàn:** Đây là giá trị quan trọng nhất. Phân tích chỉ ra rằng, ngay cả ở kịch bản "xấu nhất" (cận dưới của khoảng tin cậy), mức chi tiêu trung bình thực tế (μ) cũng

đã là 2.246 triệu VND. Con số này cao hơn mốc hòa vốn 2.0 triệu của chúng ta một khoảng 0.246 triệu (tương đương 12.3%).

Biên an toàn 12.3% này là một tín hiệu tích cực, cho thấy dự án có khả năng chống chịu được các biến động nhỏ về chi phí hoặc sai số trong dự báo ban đầu.

4.2 Quyết định Kinh doanh: Nên kinh doanh

Dựa trên các diễn giải thống kê ở trên, em đã có thể đưa ra khuyến nghị trực tiếp cho bài toán kinh doanh đã đặt ra ở Chương 1:

Quyết định: Nên kinh doanh.

Lý do là vì cả hai phương pháp suy luận đều cho kết quả đồng thuận:

- **Kiểm định t-test** khẳng định (với p-value ≈ 0) rằng thị trường có mức chi tiêu trung bình lớn hơn 2.0 triệu VND.
- **Khoảng tin cậy 95%** chỉ ra rằng mức chi tiêu này không chỉ lớn hơn, mà còn lớn hơn một khoảng an toàn (ít nhất là 12.3%).

Mặc dù bối cảnh thị trường sinh viên HUST rất nhạy cảm về giá, nhưng dữ liệu cho thấy tổng ngân sách hàng tháng mà sinh viên dành cho ăn uống là đủ lớn. Điều này cho phép một mô hình F&B mới, dù cạnh tranh về giá (ví dụ: bán suất ăn 25 nghìn đồng cho tới 30 nghìn đồng), vẫn có đủ "dung lượng thị trường" để tồn tại.

4.3 Hạn chế của Nghiên cứu và Hướng phát triển

Việc ra quyết định này dựa trên một nghiên cứu thống kê và cần được xem xét cùng với các hạn chế của nó.

- **Tính đại diện của mẫu:** Mẫu $n = 35$ là tương đối nhỏ so với quy mô khoảng hơn 10.000 sinh viên nhập học hàng năm. Hơn nữa, phương pháp khảo sát khả thi (Ví dụ như em định gửi form khảo sát) có thể dẫn đến "thiên kiến chọn mẫu", khi chỉ những người năng động hoặc có thói quen chi tiêu nhất định trả lời tử tế.
- **Dữ liệu tự khai báo:** Con số sinh viên cung cấp là "ước tính". Họ có thể không nhớ chính xác hoặc cố tình báo cáo sai lệch.
- **Giả định của mô hình:** Phép kiểm định t-test yêu cầu dữ liệu xấp xỉ chuẩn hoặc cỡ mẫu đủ lớn. Mặc dù $n = 35$ được xem là đủ để áp dụng Định lý Giới hạn Trung tâm (CLT), một cỡ mẫu lớn hơn sẽ cho kết quả đáng tin cậy hơn.
- **Yếu tố bên ngoài không được xem xét:** Nghiên cứu chỉ tập trung vào chi tiêu trung bình mà không xem xét các yếu tố khác như xu hướng tiêu dùng, sự cạnh tranh từ các mô hình F&B hiện có, hoặc các yếu tố kinh tế vĩ mô có thể ảnh hưởng đến chi tiêu của sinh viên. Ngoài ra, nghiên cứu chưa phân tích sự khác biệt chi tiêu giữa các nhóm sinh viên (theo Khoa, năm học, giới tính,...), điều này có thể cung cấp những hiểu biết sâu sắc hơn về thị trường mục tiêu nhưng không được đề cập đến trong khuôn khổ bài báo cáo giữa kì môn Suy luận thống kê.

Hướng phát triển: Trước khi đầu tư một số vốn lớn, em dự tính sẽ làm nghiên cứu này thực tế nhưng cần được mở rộng với một mẫu lớn hơn (ví dụ $n > 200$) và sử dụng phương pháp **lấy mẫu phân tầng** (stratified sampling) — đảm bảo thu thập đủ dữ liệu từ các Khoa/Trường khác nhau (Toán-Tin, CNTT, Kinh tế,...) và các năm học khác nhau (Năm 1 đến Năm 4) để có một bức tranh toàn cảnh chính xác nhất về thị trường.

4.4 Kết luận chung

Báo cáo này được thực hiện nhằm giải quyết một bài toán kinh doanh thực tế: đánh giá tính khả thi của một mô hình F&B cạnh tranh về giá tại thị trường sinh viên HUST, dựa trên mốc chi tiêu trung bình tối thiểu là $\mu_0 = 2.0$ triệu VND/tháng.

Để trả lời câu hỏi này, một cuộc khảo sát trên $n = 35$ sinh viên đã được tiến hành. Dữ liệu thu thập được đã qua các bước xử lý, thống kê mô tả và trực quan hóa để đảm bảo tính hợp lệ.

Sử dụng các phương pháp suy luận thống kê cốt lõi đã trình bày trong Chương 2, em đã thực hiện hai phép phân tích chính:

1. **Ước lượng Khoảng tin cậy 95%** cho chi tiêu trung bình, kết quả thu được là **(2.246, 2.577) triệu VND**.
2. **Kiểm định giả thuyết** $H_1 : \mu > 2.0$, cho kết quả **p-value** 0.00000738.

Cả hai kết quả phân tích đều đồng thuận. Giá trị p-value (nhỏ hơn $\alpha = 0.05$) đã cung cấp bằng chứng mạnh mẽ để **bắc bỏ giả thuyết** H_0 . Đồng thời, khoảng tin cậy 95% cũng xác nhận rằng mức chi tiêu trung bình thực tế không chỉ lớn hơn 2.0 triệu mà còn vượt trên một khoảng an toàn đáng kể (ít nhất là 12.3%).

Kết luận cuối cùng của nghiên cứu là: **Thị trường khả thi**. Quyết định **Tiếp tục triển khai** được khuyến nghị cho dự án F&B.

Tài liệu tham khảo

- [1] Nguyễn Thị Thu Thủy. *Chương 2: Ước lượng tham số*. Slide bài giảng môn Suy luận thống kê (MI3031). Hà Nội: Đại học Bách khoa Hà Nội (FaMI), 2025.
- [2] Nguyễn Thị Thu Thủy. *Chương 3: Kiểm định giả thuyết thống kê*. Slide bài giảng môn Suy luận thống kê (MI3031). Hà Nội: Đại học Bách khoa Hà Nội (FaMI), 2025.

Chương A

Dữ liệu khảo sát thô

Bảng A.1: Bảng dữ liệu khảo sát thô (n=35).

Họ Tên	Năm học	Khoa / Trường	Quê quán	Chi tiêu
Nguyễn Văn An	Năm 4	Khoa Toán-Tin	Hải Phòng	2.4
Trần Thị Bình	Năm 2	Trường Kinh tế	Dà Nẵng	1.9
Lê Minh Cường	Năm 3	Trường Kinh tế	Hà Nội	2.8
Phạm Thu Hà	Năm 3	Trường Công nghệ thông tin và Truyền thông	Hải Dương	2.2
Hoàng Văn Dũng	Năm 4	Khoa Toán-Tin	TP. Hồ Chí Minh	3.4
Dặng Thị Em	Năm 3	Trường Công nghệ thông tin và Truyền thông	Bắc Ninh	2.5
Vũ Minh Hải	Năm 3	Trường Kinh tế	Hà Nam	2.3
Bùi Thu Hằng	Năm 2	Khoa Vật lý	Thái Bình	2.1
Hồ Văn Hùng	Năm 2	Khoa Toán-Tin	Hải Phòng	1.8
Ngô Thị Lan	Năm 4	Trường Kinh tế	Nghệ An	3.1
Dương Văn Kiên	Năm 3	Khoa Vật lý	Dà Nẵng	2.6
Lý Thị Mai	Năm 4	Trường Công nghệ thông tin và Truyền thông	Vĩnh Phúc	2.0
Phan Văn Nam	Năm 2	Trường Kinh tế	Hà Nội	2.7
Dỗ Thu Nga	Năm 4	Khoa Toán-Tin	Nghệ An	2.2
Trịnh Văn Quang	Năm 2	Khoa Vật lý	Thanh Hóa	2.3
Dào Thị Oanh	Năm 3	Trường Kinh tế	Hải Dương	1.6
Mai Văn Phong	Năm 3	Trường Công nghệ thông tin và Truyền thông	Thái Bình	2.4
Nguyễn Thị Quyên	Năm 4	Khoa Toán-Tin	Nam Định	2.9
Võ Văn Sáng	Năm 2	Trường Kinh tế	Lào Cai	2.1
Lâm Thị Tâm	Năm 3	Trường Công nghệ thông tin và Truyền thông	Hà Nội	1.9
Trương Văn Tuấn	Năm 4	Khoa Toán-Tin	Lạng Sơn	2.7
Huỳnh Thị Uyên	Năm 2	Khoa Vật lý	Hải Phòng	1.9
Phạm Văn Việt	Năm 3	Trường Công nghệ thông tin và Truyền thông	Hải Dương	2.5

Tiếp tục ở trang sau

Bảng A.1 – (tiếp theo)

Họ Tên	Năm học	Khoa / Trường	Quê quán	Chi tiêu
Bùi Thị Xuân	Năm 4	Khoa Toán-Tin	Nghệ An	2.2
Đoàn Văn Ý	Năm 2	Trường Kinh tế	Hà Nội	3.0
Lại Thị Thu	Năm 4	Trường Công nghệ thông tin và Truyền thông	Vĩnh Phúc	1.8
Giang Văn Trung	Năm 3	Khoa Vật lý	Thanh Hóa	3.2
Nguyễn Minh Châu	Năm 2	Trường Kinh tế	Nam Định	2.5
Vương Tuấn Kiệt	Năm 3	Trường Công nghệ thông tin và Truyền thông	Quảng Ninh	3.5
Hà Thị Yến	Năm 2	Khoa Toán-Tin	Hà Nội	2.6
Lê Quang Vinh	Năm 4	Khoa Vật lý	Lạng Sơn	1.7
Ngô Bảo Long	Năm 2	Trường Kinh tế	Thái Nguyên	2.3
Trần Hữu Phước	Năm 3	Trường Công nghệ thông tin và Truyền thông	Nam Định	2.9
Hoàng Thảo My	Năm 4	Khoa Vật lý	Dà Nẵng	2.4
Dinh Văn Toàn	Năm 3	Khoa Toán-Tin	Vĩnh Phúc	2.0

Chương B

Mã nguồn Phân tích (Python)

Dưới đây là toàn bộ mã nguồn Python được sử dụng trong Jupyter Notebook để đọc dữ liệu, tính toán các đại lượng thống kê, vẽ biểu đồ và thực hiện kiểm định.

```
1 import pandas as pd
2 from scipy import stats
3 import matplotlib.pyplot as plt
4 import numpy as np
5
6 # 1. Doc du lieu tu file CSV
7 filename = 'khao_sat_chi_tieu.csv'
8 df = pd.read_csv(filename)
9
10 # 2. Chon cot du lieu 'ChiTieu' de phan tich
11 data = df['ChiTieu']
12
13 # 3. Tinh cac gia tri thong ke co ban
14 n = len(data)
15 x_bar = data.mean()
16 s = data.std()
17 median = data.median()
18 print(f"Kich thuoc mau (n): {n}")
19 print(f"Trung binh mau (x_bar): {x_bar}")
20 print(f"Do lech chuan mau (s): {s}")
21
22 # 4. Ve bieu do
23 # Histogram
24 plt.figure(figsize=(10, 5))
25 plt.hist(data, bins=8, edgecolor='black', alpha=0.7)
26 plt.title('Histogram Chi tieu hang thang')
27 plt.xlabel('Chi tieu (trieu VND)')
28 plt.ylabel('Tan suat')
29 plt.axvline(x_bar, color='red', linestyle='--', label=f'Trung binh = {x_bar:.2f}')
30 plt.legend()
31 plt.savefig('images/histogram.png')
32 # Boxplot
33 plt.figure(figsize=(8, 4))
34 plt.boxplot(data, vert=False, patch_artist=True)
35 plt.title('Boxplot Chi tieu hang thang')
36 plt.xlabel('Chi tieu (trieu VND)')
37 plt.savefig('images/boxplot.png')
38
39 # 5. Uoc luong Khoang Tin Cay 95%
40 alpha = 0.05
41 df = n - 1 # Bac tu do
42 ci_low, ci_high = stats.t.interval(confidence=1-alpha,
43                                     df=df,
44                                     loc=x_bar,
```

```

45                                     scale=stats.sem(data))
46 print(f"KTC 95% cho mu la: ({ci_low}, {ci_high})")
47
48 # 6. Kiem dinh Gia thuyet
49 mu_0 = 2.0 # Gia thuyet H0
50 t_statistic, p_value = stats.ttest_1samp(data, mu_0, alternative='greater')
51 print(f"Thong ke T_0: {t_statistic}")
52 print(f"P-value (one-tailed): {p_value}")
53 if p_value < alpha:
54     print("Ket luan: Bac bo H0.")
55 else:
56     print("Ket luan: Khong du bang chung bac bo H0.")

```

Listing B.1: Mã nguồn Python cho Chương 3