TỔNG LIÊN ĐOÀN LAO ĐỘNG VIỆT NAM

**TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

**KHOA CÔNG NGHỆ THÔNG TIN**



**BÁO CÁO CUỐI KÌ**

**MÔN XÁC SUẤT THỐNG KÊ CHO  
CÔNG NGHỆ THÔNG TIN**

*Người hướng dẫn*: **TS. NGUYỄN THỊ HUỲNH TRÂM**

*Người thực hiện*: **HUỲNH HOÀNG TIẾN ĐẠT – 52200023**

Lớp **: 22050201**

Khoá  **: 26**

**THÀNH PHỐ HỒ CHÍ MINH, THÁNG 12 NĂM 2023**

TỔNG LIÊN ĐOÀN LAO ĐỘNG VIỆT NAM

**TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

**KHOA CÔNG NGHỆ THÔNG TIN**



**BÁO CÁO CUỐI KÌ**

**MÔN XÁC SUẤT THỐNG KÊ CHO**

**CÔNG NGHỆ THÔNG TIN**

*Người hướng dẫn*: **TS. NGUYỄN THỊ HUỲNH TRÂM**

*Người thực hiện*: **HUỲNH HOÀNG TIẾN ĐẠT – 52200023**

Lớp **: 22050201**

Khoá  **: 26**

THÀNH PHỐ HỒ CHÍ MINH, THÁNG 12 NĂM 2023**LỜI CẢM ƠN**

Để hoàn thành đề tài, bài báo cáo cuối Học kì I năm học 2023 - 2024 môn Xác suất và thống kê ứng dụng cho Công nghệ thông tin lần này.

Lời đầu tiên, em xin gửi lời cảm ơn chân thành của mình đến Ban giám hiệu trường Đại học Tôn Đức Thắng, quý thầy cô giáo giảng viên của khoa Công nghệ thông tin vì đã tạo điều kiện tốt nhất về cơ sở vật chất và hệ thống thư viện hiện đại, đa dạng đầy đủ các loại sách giúp em có thể trau đồi kiến thức, tìm kiếm thông tin, tư liệu để hoàn thành bài Báo cáo cuối kỳ lần này. Đây thực sự là một cơ hội tuyệt vời giúp cho nghề nghiệp của em trong tương lai rộng mở hơn khi được tiếp xúc với sự hiện đại và nhiều kiến thức.

Bên cạnh đó, em cũng nhận được rất nhiều sự giúp đỡ tận tình của quý thầy cô. Thầy cô giảng viên là những người hướng dẫn và truyền cảm hứng cho em trong học tập trong suốt thời gian qua ở môi trường đại học.

Với lòng biết ơn sâu sắc và vô cùng đặc biệt, em xin gửi lời cảm ơn của mình đến TS. Nguyễn Thị Huỳnh Trâm – Giảng viên lý thuyết môn Xác suất và thống kê ứng dụng cho Công nghệ thông tin, người đã luôn đồng hành, dẫn dắt và giúp đỡ em trong việc hoàn thành bài báo cáo cuối kỳ lần này. Từ những kiến thức cô đã giảng dạy trên những giờ học để em có thể áp dụng những kiến thức đã được vào bài báo cáo cuối kỳ lần này. Một lần nữa em xin chân thành cảm ơn cô vì sự hỗ trợ của cô ạ.

Vì kiến thức bản thân còn hạn chế nên trong quá trình giải quyết vấn đề nên khi hoàn thành bài báo cáo lần này em không tránh khỏi những sai sót, em kính mong nhận được những lời nhận xét, đóng góp ý kiến từ cô ạ.

Lời cuối cùng, em xin gửi lời cảm ơn chân thành và gửi ngàn lời chúc tốt đẹp đến với quý thầy cô khi đã tạo cơ hội cho chúng em nâng cấp kiến thức trong môn học này.

**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 / chúng tôi và được sự hướng dẫn của TS.Nguyễn Thị Huỳnh Trâm;. 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 03 tháng 11 năm 2023*

*Tác giả*

*(ký tên và ghi rõ họ tên)*

A picture containing linedrawing

Description automatically generated

*Huỳnh Hoàng Tiến Đạt*

**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 gồm có **06 chương**:

**Chương 1: Mở đầu**

**Chương 2: Tổng quan**

**Chương 3: Phần 1**

**Chương 4: Phần 2**

**Chương 5: Phần 3**

**Chương 6: Phần 4**

# **MỤC LỤC**

[TÓM TẮT 6](#_Toc154664172)

[MỤC LỤC 7](#_Toc154664173)

[DANH MỤC CÁC HÌNH ẢNH 9](#_Toc154664174)

[LỜI MỞ ĐẦU 10](#_Toc154664175)

[CHƯƠNG 1 - MỞ ĐẦU 11](#_Toc154664176)

[CHƯƠNG 2 - TỔNG QUAN 12](#_Toc154664177)

[CHƯƠNG 3 - PHẦN 1 13](#_Toc154664178)

[3.1 Câu 1 13](#_Toc154664179)

[3.1.1 Yêu cầu 13](#_Toc154664180)

[3.1.2 Trả lời 13](#_Toc154664181)

[3.2 Câu 2 14](#_Toc154664182)

[3.2.1 Yêu cầu 14](#_Toc154664183)

[3.2.2 Trả lời 14](#_Toc154664184)

[3.3 Câu 3 14](#_Toc154664185)

[3.3.1 Yêu cầu 14](#_Toc154664186)

[3.3.2 Trả lời 14](#_Toc154664187)

[3.4 Câu 4 15](#_Toc154664188)

[3.4.1 Yêu cầu 15](#_Toc154664189)

[3.4.2 Trả lời 15](#_Toc154664190)

[3.5 Câu 5 16](#_Toc154664191)

[3.5.1 Yêu cầu 16](#_Toc154664192)

[3.5.2 Trả lời 17](#_Toc154664193)

[3.6 Câu 6 17](#_Toc154664194)

[3.6.1 Yêu cầu 17](#_Toc154664195)

[3.6.2 Trả lời 18](#_Toc154664196)

[3.7 Câu 7 19](#_Toc154664197)

[3.7.1 Yêu cầu 19](#_Toc154664198)

[3.7.2 Trả lời 19](#_Toc154664199)

[3.8 Câu 8 20](#_Toc154664200)

[3.8.1 Yêu cầu 20](#_Toc154664201)

[3.8.2 Trả lời 20](#_Toc154664202)

[3.9 Câu 9 24](#_Toc154664203)

[3.9.1 Yêu cầu 24](#_Toc154664204)

[3.9.2 Trả lời 24](#_Toc154664205)

[CHƯƠNG 4 – PHẦN 2: 26](#_Toc154664206)

[4.1 Câu 10 26](#_Toc154664207)

[4.1.1 Yêu cầu: 26](#_Toc154664208)

[4.1.2 Trả lời: 26](#_Toc154664209)

[CHƯƠNG 5 – PHẦN 3 29](#_Toc154664210)

[5.1 Câu 11: 29](#_Toc154664211)

[5.1.1 Yêu cầu: 29](#_Toc154664212)

[5.1.2 Trả lời: 29](#_Toc154664213)

# **DANH MỤC CÁC HÌNH ẢNH**

Hình 1: Code minh họa tạo tự động tập STT S1 và S2

Hình 2: Kết quả khi chạy code python dùng hàm random tạo ra tập STT S1, S2

Hình 3: Tập mẫu S1, S2

# **LỜI MỞ ĐẦU**

Trong thời đại công nghệ số phát triển mạnh mẽ, nhu cầu và đời sống của con người ngày càng được cải thiện. Con người có xu hướng trải nghiệm, tìm kiếm những điều mới mẻ, tận hưởng cuộc sống hiện đại, ứng dụng những công nghệ hiện đại như robot,... Điều này dẫn đến việc ứng dụng, sử dụng ngôn ngữ lập trình để tạo ra các trang web, trò chơi, ứng dụng là điều không thể tránh khỏi. Thậm chí là các loại AI chatbot để tư vấn những thắc mắc của con người chúng ta.

Môn học "Xác suất thông kê ứng dụng cho Công nghệ Thông tin" không chỉ là nền tảng lý thuyết mà còn mang đến cái nhìn sâu sắc về cách chúng ta có thể áp dụng những nguyên lý này vào thực tế.

Môn học mang đến những khám phá mới mẻ từ nhiều khía cạnh quan trọng của xác suất thông kê và cách nó có thể được áp dụng hiệu quả trong lĩnh vực công nghệ thông tin. Mục tiêu là không chỉ là hiểu biết về các khái niệm lý thuyết, mà còn là khả năng áp dụng chúng để giải quyết những thách thức cụ thể trong lĩnh vực Công nghệ Thông

# **CHƯƠNG 1 - MỞ ĐẦU**

Báo cáo này mục tiêu chi tiết hóa về mức lương khởi điểm của sinh viên mới ra trường thông qua việc sử dụng một tập dữ liệu cụ thể. Chúng ta sẽ áp dụng các phương pháp thống kê không chỉ để trình bày con số mà còn để giải thích rõ ý nghĩa của chúng. Qua bảng số liệu và các kết quả thống kê, báo cáo sẽ tập trung khám phá đặc điểm chung của mức lương, những sự khác biệt giữa nam và nữ, cũng như cách mẫu dữ liệu được tập trung để đưa ra nhận định có ý nghĩa về tổng thể.

Nội dung của báo cáo được tổ chức thành bốn phần chính như sau:

1. **Phân Tích Mức Lương Trung Bình và Độ Lệch Chuẩn:**

Trong phần này, chúng ta sẽ đàm phán về mức lương trung bình và độ lệch chuẩn để hiểu rõ sự biến động và tính đồng đều của dữ liệu. Các phương tiện thống kê này sẽ giúp ta có cái nhìn tổng quan về phân phối mức lương.

1. **Phân Tích Giới Tính và Mức Lương:**

Báo cáo sẽ tập trung vào việc phân tích sự chênh lệch về mức lương giữa nam và nữ. Các thông số thống kê sẽ được áp dụng để phân loại và diễn giải sự khác biệt này, đồng thời đưa ra những nhận định sâu sắc về bất bình đẳng giới tính trong lĩnh vực nghề nghiệp.

1. **Lấy Mẫu và Diễn Giải:**

Phần này sẽ giải thích quy trình lấy mẫu dữ liệu, chọn mẫu một cách đại diện và giải thích cách mẫu được tập trung để đảm bảo sự độc lập và tính khảo sát của nghiên cứu.

1. **Dự Đoán Tương Lai:**

Cuối cùng, báo cáo sẽ đưa ra những dự đoán và phân tích về xu hướng tương lai của mức lương dựa trên các mô hình thống kê. Những dự báo này sẽ giúp hiểu rõ hơn về tiến triển của mức lương trong ngành công nghiệp và có thể hỗ trợ quyết định chiến lược trong tương lai.

# **CHƯƠNG 2 - TỔNG QUAN**

Trong một khảo sát mức lương khởi điểm của sinh viên vừa ra trường của một lớp học gồm 65 sinh viên, kết quả được đánh số thứ tự như bảng sau.

Đơn vị của mức lương là X00,000 đồng (Ví dụ 27 tức là 2,700,000 VNĐ)

Giới tính: W là nữ, M là nam.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| STT | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Giới tính | W | M | M | W | M | M | W | M | M | W |
| Tiền lương | 27 | 61 | 52 | 69 | 88 | 85 | 79 | 99 | 77 | 165 |
| STT | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| Giới tính | M | W | M | W | M | M | M | W | M | M |
| Tiền lương | 41 | 83 | 144 | 74 | 143 | 131 | 34 | 59 | 46 | 105 |
| STT | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 |
| Giới tính | M | W | W | W | M | M | W | M | M | W |
| Tiền lương | 61 | 118 | 114 | 138 | 24 | 67 | 130 | 56 | 99 | 125 |
| STT | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 |
| Giới tính | M | M | W | W | M | M | W | M | M | M |
| Tiền lương | 87 | 30 | 119 | 40 | 25 | 44 | 123 | 45 | 25 | 94 |
| STT | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 |
| Giới tính | W | W | W | W | M | M | W | M | M | M |
| Tiền lương | 86 | 128 | 69 | 102 | 91 | 106 | 119 | 139 | 67 | 47 |
| STT | 51 | 52 | 53 | 54 | 55 | 56 | 57 | 58 | 59 | 60 |
| Giới tính | W | W | W | M | M | M | W | W | W | M |
| Tiền lương | 62 | 92 | 124 | 31 | 49 | 68 | 109 | 138 | 105 | 84 |
| STT | 61 | 62 | 63 | 64 | 65 |  |  |  |  |  |
| Giới tính | M | M | W | W | M |  |  |  |  |  |
| Tiền lương | 86 | 66 | 128 | 146 | 59 |  |  |  |  |  |

Bảng 1: Mức lương khởi điểm của sinh viên vừa ra trường của một lớp học gồm 65 sinh viên

# **CHƯƠNG 3 - PHẦN 1**

Giả sử tập dữ liệu trên là tập dữ liệu của một tổng thể

## **3.1 Câu 1**

### **3.1.1 Yêu cầu**

Tính giá trị trung bình và độ lệch chuẩn của mức lương khởi điểm (0.5 điểm)

### **3.1.2 Trả lời**

1. **Tính giá trị trung bình(Mean)**

Gọi µ là giá trị trung bình của tổng thể, N = 65 là tổng số lượng trong tổng thể.

Công thức tính trung bình của tổng thể:

µ

Với:

Tính tổng số lương ( =

Từ đó ta tính giá trị trung bình:

µ = ≈ 85.03076923

Vậy mức lương trung bình khởi điểm của sinh viên vừa ra trường của một lớp học gồm 65 sinh viên là 85.03076923 tức là khoảng 8,503,769 VNĐ.

1. **Tính độ lệch chuẩn (Standard Deviation):**

Gọi σ là độ lệch chuẩn của tổng thể, N = 65 là tổng số lượng trong tổng thể.

Bước 1: Tính phương sai (Variance):

(Trong đó µ là giá trị trung bình)

**Vậy:**

Vậy độ lệch chuẩn của bộ dữ liệu sẽ là : 36.40666413 tức là khoảng 3,640,666 VNĐ

**Kết luận:** Bộ dữ liệu khảo sát mức lương khởi điểm của sinh viên vừa ra trường của một lớp học gồm 65 sinh viên có:

* Giá trị trung bình: µ = 85.0307692308
* Độ lệch chuẩn: = 36.40666413

## **3.2 Câu 2**

### **3.2.1 Yêu cầu**

Tính tỉ lệ nữ có mức lương cao hơn mức lương trung bình (0.25 điểm)

### **3.2.2 Trả lời**

Để tính tỉ lệ nữ có mức lương cao hơn mức lương trung bình thì chúng ta

sử dụng các biến sau:

W là tỉ lệ nữ có mức lương cao hơn mức lương trung bình.

C là số lượng nữ có mức lương cao hơn mức trung bình.

S là tổng số nữ trong tập dữ liệu.

Theo kết quả tính được của câu 1, ta có mức lương trung bình khởi điểm của sinh viên vừa ra trường là 85.03076923 hay nói cách khác là 8,503,796 VNĐ.

Theo tập dữ liệu số sinh viên nữ có mức lương cao hơn mức trung bình sẽ là C = 19 người và tổng số sinh viên nữ có trong tập dữ liệu là S = 28 người.

**Bước 1: Tính tỉ lệ nữ có mức lương cao hơn mức lương trung bình là:**

tức là khoảng 67,857%

**Bước 2: Kết luận**:

Dựa vào tính toán ở trên, chúng ta thu được tỉ lệ nữ có mức lương cao hơn mức lương trung bình trong bộ dữ liệu là W ≈ 0.6785714286 (67,857%). Tỉ lệ này là một chỉ số quan trọng, cho thấy phần lớn phụ nữ trong tập dữ liệu có mức lương cao hơn mức trung bình.

Tỷ lệ W ≈ 0.6785714286 (67,857%) có thể hiểu là nữ có mức lương cao hơn mức lương trung bình lớn hơn 50%, điều này chứng tỏ rằng trong tập dữ liệu thì phần lớn nữ có mức lương cao hơn mức lương trung bình. Điều này mang lại một cái nhìn tích cực về phân phối thu nhập giữ nam và nữ trong lớp học.

## **3.3 Câu 3**

### **3.3.1 Yêu cầu**

Tính tỉ lệ nam có mức lương cao hơn mức lương trung bình (0.25 điểm)

### **3.3.2 Trả lời**

Để tính tỉ lệ nam có mức lương cao hơn mức lương trung bình, chúng ta sử dụng các biến sau:

M là tỉ lệ nam có mức lương cao hơn mức lương trung bình.

C là số lượng nam có mức lương cao hơn mức trung bình.

S là tổng số nam trong tập dữ liệu.

Theo câu 1 ta có mức lương trung bình khởi điểm của sinh viên vừa ra trường là 85.03076923 tức là 8,503,796 VNĐ.

Theo tập dữ liệu số sinh viên nam có mức lương cao hơn mức trung bình sẽ là C = 13 người và tổng số sinh viên nam có trong tập dữ liệu là S = 37 người.

**Bước 1: Tính tỉ lệ nam có mức lương cao hơn mức lương trung bình là:**

tức là khoảng 35,135%

**Bước 2: Kết luận**:

Dựa vào tính toán, chúng ta thu được tỉ lệ nam có mức lương cao hơn mức lương trung bình là M (35,135%). Tỉ lệ này cho thấy rằng trong tập dữ liệu, phần lớn nam có mức lương thấp hơn mức lương trung bình. Cụ thể, M (35,135%) có nghĩa là tỉ lệ nam có mức lương cao hơn mức lương trung bình thấp hơn 50%, hay có ý rằng phần lớn nam có mức lương thấp hơn mức lương trung bình.

Trong tập dữ liệu này, có thể nói rằng phần lớn sinh viên nam vừa ra trường có mức lương khởi điểm thấp hơn so với mức lương trung bình của toàn bộ lớp học.

## **3.4 Câu 4**

### **3.4.1 Yêu cầu**

Bảng dữ liệu trên bao gồm một nhóm nam và một nhóm nữ. Chọn ngẫu nhiên một nhóm biết xác suất chọn là như nhau. Từ nhóm đã chọn, chọn ngẫu nhiên 1 người để phỏng vấn.

a. Tính xác suất chọn được người có mức lương thấp hơn mức lương trung bình. (1 điểm)

b. Tính xác suất chọn được một người có mức lương cao hơn mức lương trung bình biết rằng người đó là nam. (1 điểm)

### **3.4.2 Trả lời**

1. **Xác xuất để chọn được người có mức lương thấp hơn mức lương trung bình là:**

**Gọi các biến cố:**

* + *M*: Biến cố chọn nhóm nam.
  + *X*: Biến cố chọn một người có mức lương thấp hơn trung bình.
  + *W*: Biến cố chọn nhóm nữ.

**Xác suất chọn một nhóm nam hoặc nữ (P(M)) là:**

​

(Vì xác suất chọn một nhóm nam hay nữ là như nhau)

**Xác suất chọn một người có mức lương thấp hơn mức lương trung bình từ nhóm nam (P(X∣M)):**

P(X∣M): Xác suất chọn một người có mức lương thấp hơn trung bình biết rằng họ thuộc nhóm nam.

(Với TM = 24 là số sinh viên nam có mức lương thấp hơn trung bình và nM = 37 là số sinh viên nam.)

**Xác suất chọn một người có mức lương thấp hơn mức lương trung bình từ nhóm nữ (P(X∣W)):**

P(X∣W): Xác suất chọn một người có mức lương thấp hơn trung bình biết rằng họ thuộc nhóm nữ.

(Với TW = 9 là số sinh viên nữ có mức lương thấp hơn trung bình trong bộ dữ liệu và = 28 là số sinh viên nữ.)

**Từ đó, Xác suất để chọn một người có mức lương thấp hơn mức lương trung bình (P(X)) là:**

(Gọi P(X) là xác suất chọn một người có mức lương thấp hơn trung bình)

**b) Xác suất chọn được người có mức lương thấp hơn trung bình biết rằng người đó là nam (P(M∣X)):**

Với các biến cố này, ta có thể sử dụng công thức Bayes để tính *P*(*M*∣*X*) như sau:

Trong đó:

* *P*(*X*∣*M*): Xác suất chọn một người có mức lương thấp hơn trung bình biết rằng họ thuộc nhóm nam.
* *P*(*M*): Xác suất chọn nhóm nam.
* *P*(*X*): Xác suất chọn một người có mức lương thấp hơn trung bình.

## **3.5 Câu 5**

### **3.5.1 Yêu cầu**

Gọi S1 là mẫu gồm 30 sinh viên và tập S2 là mẫu gồm 15 sinh viên. Có bao nhiêu cách lấy 1 tập mẫu S1? Có bao nhiêu cách lấy 1 tập mẫu S2? (0.5 điểm)

### **3.5.2 Trả lời**

Trong trường hợp lấy mẫu không thay đổi (có thể chọn một sinh viên và không bao giờ chọn lại), chúng ta có thể sử dụng công thức tính tổ hợp để xác định số cách lấy mẫu:

Công thức tổ hợp để chọn *k* phần tử từ *n* phần tử là:

**Trong đó:**

* *n*! là giai thừa của *n*, tức là tích của tất cả các số nguyên từ 1 đến *n*.
* *C*(*n*,*k*) là số cách chọn *k* phần tử từ *n* phần tử.

**Số cách lấy mẫu S1 (30 sinh viên từ tổng số 65 sinh viên):**

**Trong đó:**

* + n = 65 là tổng số sinh viên.
  + k = 30 là số sinh viên trong mẫu S1.
  + C(65,30) là số cách chọn k sinh viên từ n sinh viên

**Số cách lấy mẫu S2 (15 sinh viên từ tổng số 65 sinh viên):**

**Trong đó:**

* + n = 65 là tổng số sinh viên.
  + k = 15 là số sinh viên trong mẫu S2.
  + C(65,15) là số cách chọn k sinh viên từ n sinh viên

≈ 207,374,699,821,536 (cách)

**Kết luận**: Vậy với bộ dữ liệu gồm 65 sinh viên ta có:

* Số cách lấy tập mẫy S1 là:
* Số cách lấy tập mẫu S2 là: = 207,374,699,821,536 cách.

## **3.6 Câu 6**

### **3.6.1 Yêu cầu**

Diễn giải chi tiết phương pháp lấy ngẫu nhiên tập S1, S2 bằng cách chụp màn hình kết quả làm minh chứng. Đồng thời điền kết quả từ câu 6 đến câu 10 vào link sau:

(Lưu ý: Sinh viên điền rõ thông tin họ tên, MSSV, và các kết quả của mình vào link. Chú ý sao cho bộ dữ liệu S1, S2 của mỗi bạn phải khác nhau.) (1 điểm)

Gợi ý: Có thể dùng hàm random để chọn ra từng giá trị trong 1 mẫu dựa vào số thứ tự trong bảng dữ liệu.

### **3.6.2 Trả lời**

Phương pháp lấy ngẫu nhiên tập S1 và S2:

Ta sẽ dùng code của python để lấy ngẫu nhiên tập STT S1, S2 sau đó dựa vào số thứ tự trong bảng dữ liệu ta sẽ được tập S1, S2.

Đầu tiên ta sẽ dùng hàm def Tao\_mau\_STT(size) với tham số size đại diện cho kích thước của tập số thứ tự.

Tiếp tục là sử dụng “random.sample(range(1, 66), size)”: Hàm sample từ thư viện random được sử dụng để lấy ngẫu nhiên của size phần tử từ dãy số từ 1 đến 65 mà không có sự lặp lại. Điều này đảm bảo rằng các phần tử trong tập mẫu S1 và S2 không trùng nhau.

Size là tham số do chúng ta truyền vào hàm để xác định số lượng phần tử muốn lấy mẫu.

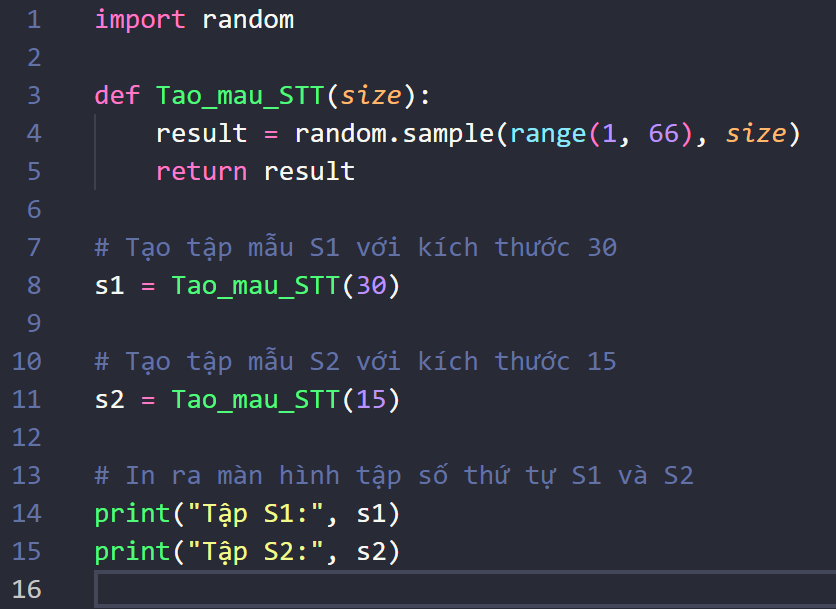
return result: trả về tập mẫu từ hàm trên

“Tao\_mau\_STT(30)”: Gọi hàm Tao\_mau\_STT để tạo tập mẫu S1 với kích thước là 30.

“Tao\_mau\_STT(15)”: Gọi hàm Tao\_mau\_STT để tạo tập mẫu S2 với kích thước là 15.

Cuối cùng, in ra màn hình tập số thứ tự S1 và S2 để xem kết quả. Kết quả là hai tập số ngẫu nhiên có kích thước là 30 và 15.

Code minh họa:



Hình 1: Code minh họa tạo tự động tập STT S1 và S2

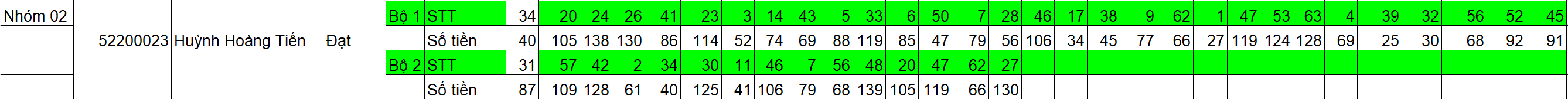
Output:

A computer screen with many small letters

Description automatically generated with medium confidence

Hình 2: Kết quả khi chạy code python dùng hàm random tạo ra tập STT S1, S2

**Kết luận:** Từ tập số thứ tự S1, S2 trên dựa vào số thứ tự trong bảng dữ liệu thì ta thu được kết quả tập mẫu S1, S2 như thế này:



Hình 3: Tập mẫu S1, S2

## **3.7 Câu 7**

### **3.7.1 Yêu cầu**

Tính giá trị trung bình mức lương khởi điểm của tập S1, S2. Bạn có kết luận gì về các giá trị này so với giá trị trung bình của tổng thể. (0.5 điểm)

### **3.7.2 Trả lời**

* **Đối với tập mẫu S1**

Gọi là giá trị trung bình của tập mẫu S1, n1 = 30 là số sinh viên có trong tập mẫu S1 và là mức lương của từng sinh viên theo số thứ tự là i trong tập mẫu S1.

Áp dụng công thức tính trung bình của tổng thể:

**Ta có:**

**Vậy:**

* **Đối với tập mẫu S2**

Gọi là giá trị trung bình của tập mẫu S2, n2 = 15 là số sinh viên có trong tập mẫu S2 và là mức lương của từng sinh viên theo số thứ tự là i trong tập mẫu S2.

Áp dụng công thức tính trung bình của tổng thể:

**Ta có:**

**Vậy:**

Từ các kết quả tính toán mà ta có được:

* Giá trị trung bình của tổng thể (x) là 85.03076923 tức là 8,503,796 VNĐ.
* Giá trị trung bình của tập mẫu S1 ( = (tức là khoảng 7,943,333 VNĐ).

Giá trị trung bình mới của mức lương khởi điểm của sinh viên nam trong tập mẫu S1 là 79.43333333, tương đương với khoảng 7,943,333 VNĐ. Kết quả này thấp hơn giá trị trung bình của toàn bộ quần thể.

* Giá trị trung bình của tập mẫu S2 ( = (tức là khoảng 9,353,333 VNĐ).

Giá trị trung bình mới của mức lương khởi điểm của sinh viên nữ trong tập mẫu S2 là 93.53333333, tương đương với khoảng 9,353,333 VNĐ. Kết quả này cũng thấp hơn giá trị trung bình của toàn bộ quần thể.

Nhìn chung, cả hai tập mẫu S1 và S2 đều có giá trị trung bình mức lương thấp hơn so với giá trị trung bình của toàn bộ quần thể, nhưng với giá trị mới, chênh lệch giữa giá trị trung bình của tập mẫu S1 và giá trị trung bình của toàn bộ quần thể là khoảng 5,560,463 VNĐ, trong khi tập mẫu S2 có sự chênh lệch giảm xuống còn khoảng 516,463 VNĐ. Kết quả này vẫn chứng tỏ rằng cả nam và nữ trong tập mẫu S1 và S2 đều có mức lương khởi điểm thấp hơn so với mức lương trung bình của toàn bộ quần thể.

## **3.8 Câu 8**

### **3.8.1 Yêu cầu**

Uớc lượng khoảng tin cậy đối của giá trị trung bình của tập S1, S2 với độ tin cậy là 80% và 95%. Bạn có kết luận gì về khoảng giá trị này so với giá trị trung bình của tổng thể và so với nhau. (1 điểm)

### **3.8.2 Trả lời**

* Ước lượng khoảng tin cậy với độ tin cậy 80% cho giá trị trung bình của tập S1:

Ta có:

Với α : mức ý nghĩa

🡺α = 0.2 🡺

Dò vào bảng CUMULATIVE PROBABILITIES FOR THE STANDARD NORMAL DISTRIBUTION với mức ý nghĩa ta được

Công thức ước lượng khoảng tin cậy theo công thức:

Trong đó:

* là giá trị trung bình mẫu của tập S1
* là độ lệch chuẩn của tổng thể
* là số lượng sinh viên trong tập mẫu S1

Từ đó suy ra:

L =

U =

Từ đó suy ra độ tin cậy 80% cho giá trị trung bình của tập S1 là:

(L,U) = (70.92527949, 87.94138717)

Ta có:

Ước lượng số bộ dữ liệu chứa giá trị trung bình tổng thể:

**Kết luận:**

Vậy mức lương trung bình của sinh viên vừa mới ra trường của một lớp học 65 sinh viên khoảng 70.92527949 tức là khoảng 7,092,527 VNĐ đến 87.94138717 tức là khoảng 8,794,138 VNĐ với độ tin cậy là 80%.

(Bộ dữ liệu vừa tính của tập mẫu S1 của em có chứa giá trị trung bình tổng thể là 85.03076923. Vậy bộ dữ liệu nảy thuộc bộ dữ liệu có chứa giá trị trung bình của tổng thể vậy dự đoán của em là đúng).

* Ước lượng khoảng tin cậy với độ tin cậy 80% cho giá trị trung bình của tập S2:

Ta có:

Với α : mức ý nghĩa

🡺α = 0.2 🡺

Dò vào bảng CUMULATIVE PROBABILITIES FOR THE STANDARD NORMAL DISTRIBUTION với mức ý nghĩa ta được

Công thức ước lượng khoảng tin cậy theo công thức:

Trong đó:

* là giá trị trung bình mẫu của tập S2
* là độ lệch chuẩn của tổng thể
* là số lượng sinh viên trong tập mẫu S2

Từ đó suy ra:

L =85.02527949

U =

Từ đó suy ra độ tin cậy 80% cho giá trị trung bình của tập S2 là:

(L,U) = (85.02527949, 102.0413872)

Ta có:

Ước lượng số bộ dữ liệu chứa giá trị trung bình tổng thể:

**Kết luận:**

Vậy mức lương trung bình của sinh viên vừa mới ra trường của một lớp học 65 sinh viên khoảng 85.02527949 tức là khoảng 8,025,279 VNĐ đến 102.0413872 tức là khoảng 10,204,138 VNĐ với độ tin cậy là 80%.

(Bộ dữ liệu vừa tính của tập mẫu S1 của em có chứa giá trị trung bình tổng thể là 85.03076923. Vậy bộ dữ liệu nảy thuộc bộ dữ liệu có chứa giá trị trung bình của tổng thể vậy dự đoán của em là đúng).

* Ước lượng khoảng tin cậy với độ tin cậy 95% cho giá trị trung bình của tập S1:

Ta có:

Với α : mức ý nghĩa

🡺α = 0.05 🡺

Dò vào bảng CUMULATIVE PROBABILITIES FOR THE STANDARD NORMAL DISTRIBUTION với mức ý nghĩa ta được

Công thức ước lượng khoảng tin cậy theo công thức:

Trong đó:

* là giá trị trung bình mẫu của tập S1
* là độ lệch chuẩn của tổng thể
* là số lượng sinh viên trong tập mẫu S1

Từ đó suy ra:

L =

U = 92.46129077

Từ đó suy ra độ tin cậy 95% cho giá trị trung bình của tập S1 là:

(L,U) = (66.40537589, 92.46129077)

Ta có:

Ước lượng số bộ dữ liệu chứa giá trị trung bình tổng thể:

**Kết luận:**

Vậy mức lương trung bình của sinh viên vừa mới ra trường của một lớp học 65 sinh viên khoảng 66.40537589 tức là khoảng 6,640,537 VNĐ đến 92.46129077 tức là khoảng 9,246,129 VNĐ với độ tin cậy là 95%.

(Bộ dữ liệu vừa tính của tập mẫu S2 của em có chứa giá trị trung bình tổng thể là 85.03076923. Vậy bộ dữ liệu này thuộc bộ dữ liệu có chứa giá trị trung bình của tổng thể vậy dự đoán của em là đúng).

* Ước lượng khoảng tin cậy với độ tin cậy 95% cho giá trị trung bình của tập S2:

Ta có:

Với α : mức ý nghĩa

🡺α = 0.05 🡺

Dò vào bảng CUMULATIVE PROBABILITIES FOR THE STANDARD NORMAL DISTRIBUTION với mức ý nghĩa ta được

Công thức ước lượng khoảng tin cậy theo công thức:

Trong đó:

* là giá trị trung bình mẫu của tập S2
* là độ lệch chuẩn của tổng thể
* là số lượng sinh viên trong tập mẫu S2

Từ đó suy ra:

L =

U =

Từ đó suy ra độ tin cậy 95% cho giá trị trung bình của tập S2 là:

(L,U) = (80.50537589, 106.5612908)

Ta có:

Ước lượng số bộ dữ liệu chứa giá trị trung bình tổng thể:

**Kết luận:**

Vậy mức lương trung bình của sinh viên vừa mới ra trường của một lớp học 65 sinh viên khoảng 80.50537589 tức là khoảng 8,050,375 VNĐ đến 106.5612908 tức là khoảng 10,656,129 VNĐ với độ tin cậy là 95%.

(Bộ dữ liệu vừa tính của tập mẫu S2 của em có chứa giá trị trung bình tổng thể là 85.03076923. Vậy bộ dữ liệu nảy thuộc bộ dữ liệu có chứa giá trị trung bình của tổng thể vậy dự đoán của em là đúng).

## **3.9 Câu 9**

### **3.9.1 Yêu cầu**

Với mức ý nghĩa 0.05, hãy thực kiểm định giả thuyết trái, phải, hai bên của giá trị trung bình của S1 so với giá trị trung bình của tổng thể. Kết luận của kiểm định giả thuyết có đúng với thực tế hay không? (1 điểm)

### **3.9.2 Trả lời**

Với mức ý nghĩa α = 0.05 🡺α = 0.05 🡺

Suy ra ta có độ tin cậy:

🡺

Với mức định nghĩa α = 0.05 thì ta có độ tin cậy là 95%

Gọi = là giá trị trung bình của tập mẫu S1

σ = là độ lệch chuyển của tổng thể

n = 30 là số lượng sinh viên có trong tập mẫu S1

Theo kết quả tính toán từ câu 8, ta có bộ dữ liệu của tập S1 với n = 30 và độ tin cậy 95% thì ta có:

Khoảng độ tin cậy là:

(L,U) = (66.40537589, 92.46129077)

Dự đoán giá trị trung bình của tổng thể là với độ tin cậy là 95%

Giả thuyết trái, phải, hai bên của giá trị trung bình của S1 so với giá trị trung bình của tổng thể.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Lower Tail Test | Upper Tail Test | Tow-Tailed Test |
| Hypotheses |  |  |  |

Thử nghiệm thống kê:

Với độ tin cậy là 95% thì z = 0.05 suy ra ta có:

* Giả thuyết trái của S1 so với giá trị trung bình của tổng thể.

Theo phương pháp giá trị giới hạn thì

Do đó không thể bác bỏ giả thuyết vì không đủ bằng chứng thống kế để cho rằng bằng chứng (

* Giả thuyết phải của S1 so với giá trị trung bình của tổng thể.

Theo pương pháp giá tị giới hạn thì

Do đó không thể bác bỏ giả thuyết vì không đủ bằng chứng thống kế để cho rằng bằng chứng (

* Giả thuyết hai bên của S1 so với giá trị trung bình của tổng thể.

Với z = 0.05 🡺 ta được:

Theo phương pháp giá trị giới hạn thì

Do đó không thể bác bỏ giả thuyết vì không đủ bằng chứng thống kế để cho rằng bằng chứng (.

# **CHƯƠNG 4 – PHẦN 2:**

Giả sử tập dữ liệu gồm 65 sinh viên trên là tập mẫu trong một tổng thể là hữu hạn, tuy nhiên do nhiều điều kiện khác nhau nên không thể tiến hành thu thập dữ liệu của tổng thể.

## **4.1 Câu 10**

### **4.1.1 Yêu cầu:**

Hãy thực hiện lại các bước tính toán theo yêu cầu trong câu 8, 9. (1 điểm)

### **4.1.2 Trả lời:**

**Yêu cầu 8:** Uớc lượng khoảng tin cậy đối của giá trị trung bình của tập mẫu gồm 65 phần tử với độ tin cậy là 80% và 95%. Bạn có kết luận gì về khoảng giá trị này so với giá trị trung bình của tổng thể và so với nhau. (1 điểm)

**Với độ tin cậy là 80%:**

Theo kết quả giá trị trung bình mẫu đã tính thì ta có:

Tiếp theo đó, vì không thu thập dữ liệu của tổng thể nên ta có phương sai của mẫu mới:

Với n là số phần tử của mẫu và là giá trị trung bình mẫu.

Từ đó ta có:

**Vậy:**

Với độ tin cậy là 80% thì ta có khoảng độ tin cậy là:

🡺α = 0.2 🡺

Với mức ý nghĩa là ta có 2 khi tra vào bảng CUMULATIVE PROBABILITIES FOR THE STANDARD NORMAL DISTRIBUTION.

Khoảng tin cậy được ước lượng bởi công thức như trên:

Trong đó thì là giá trị trung bình mẫu.

là độ lệch chuẩn của mẫu.

n là số lượng sinh viên có trong tập mẫu.

Khi thế số vào công thức thì ta có:

Từ đó thì ta có khoảng độ tin cậy 80% là:

(L,U) = (79.1966013,90.86493716)

Vậy mức lương trung bình của sinh viên vừa mới ra trường của một lớp học 65 sinh viên là khoảng 79.1966013 tức là khoảng 7,919,660VNĐ đến 90.86493716 tức là khoảng 9,086,493VNĐ với độ tin cậy là 80%.

**Tiếp tục với độ tin cậy là 95%:**

Với độ ti cậy là 95% nên ta có khoảng độ tin cậy là:

🡺α = 0.05 🡺

Với mức ý nghĩa là ta có khi tra vào bảng CUMULATIVE PROBABILITIES FOR THE STANDARD NORMAL DISTRIBUTION.

Khoảng tin cậy được ước lượng bởi công thức như trên:

Trong đó thì là giá trị trung bình mẫu.

là độ lệch chuẩn của mẫu.

n là số lượng sinh viên có trong tập mẫu.

Tiếp tục thế số vào công thức trên thì ta có:

Từ đó thì ta có khoảng độ tin cậy với 95% là:

(L,U) = (76.11113652,93.95040194)

Vậy mức lương trung bình của sinh viên vừa mới ra trường của một lớp học 65 sinh viên là khoảng 76.11113652 tức là khoảng 7,611,113VNĐ đến 93.95040194 tức là khoảng 9,395.040VNĐ với độ tin cậy là 95%.

**Yêu cầu 9:** Với mức ý nghĩa 0.05, hãy thực kiểm định giả thuyết trái, phải, hai bên của giá trị trung bình của tập dữ liệu gồm 65 phần tử so với giá trị trung bình của tổng thể. Kết luận của kiểm định giả thuyết có đúng với thực tế hay không? (1 điểm)

* + **Trả lời:**

Với mức định nghĩa là 0.05 thì ta có α = 0.05 →

Theo đề bài ta có định nghĩa là 0.05 tức là khoảng độ tin cậy là 95% nên:

Gọi là giá trị trung bình của tập mẫu.

s = là độ lệch chuẩn của tập mẫu.

n = 65 là só lượng sinh viên có trong tập mẫu.

Theo dữ liệu mà ta tính được từ yêu cầu 8 ở trên, tập S với n = 65 và độ tin cậy là 95% thì ta lấy được khoảng độ tin cậy là:

(L,U) = (76.11113652,93.95040194)

Dự đoán giá trị trung bình của tổng thể là với độ tin cậy là 95%

Từ đó, giả thuyết trái, phải, hai bên của giá trị trung bình của S1 so với giá trị trung bình của tổng thể là:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Lower Tail Test | Upper Tail Test | Tow-Tailed Test |
| Hypotheses |  |  |  |

Thử nghiệm thống kê:

Với độ tin cậy là 95% thì z = 0.05 ta được:

* **Kiểm định giả thuyết trái:**

Theo pương pháp giá tị giới hạn thì

Do đó không thể bác bỏ giả thuyết vì không đủ bằng chứng thống kế để cho rằng bằng chứng (

* **Kiểm định giả thuyết phải:**

Theo pương pháp giá tị giới hạn thì

Do đó không thể bác bỏ giả thuyết vì không đủ bằng chứng thống kế để cho rằng bằng chứng (

* **Kiểm định giả thuyết hai bên:**

Với z = 0.05 → ta được:

Theo pương pháp giá tị giới hạn thì

Do đó không thể bác bỏ giả thuyết vì không đủ bằng chứng thống kế để cho rằng bằng chứng (.

# **CHƯƠNG 5 – PHẦN 3**

## **5.1 Câu 11:**

### **5.1.1 Yêu cầu:**

Giả sử rằng sau mỗi 1 năm làm việc, tiền lương của mỗi người sẽ tăng 20%. Không tính trực tiếp lương của mỗi người, có thể kết luận gì về giá trị trung bình, phương sai, độ lệch chuẩn của tiền lương sau 10 năm làm việc so với lúc mới ra trường? Giả sử rằng cả 65 người không ai bỏ việc trong thời gian này.

### **5.1.2 Trả lời:**

Giả sử giá trị trung bình của tiền lương khởi điểm là  **và độ lệch chuẩn là**

**Ta có giả sử rằng sau mỗi 1 năm làm việc thì tiền lương mỗi người sẽ tăng 20% và khong có ai bỏ việc trong thời gian 10 năm. Do đó, sau 10 năm làm việc tiền lương của mỗi người sẽ là so với lúc mới ra trường.**

**Giả sử là giá trị trung bình mức lương của sinh viên i sau n năm làm việc. Theo giả định đó thì ta có:**

**Trong đó ta có: là giá trị trung bình mức lương của sinh viên sau 10 năm làm việc**

là giá trị trung bình mức lương của sinh viên mới ra trường

1. **Giá trị trung bình sau 10 năm:**
2. **Phương sai sau 10 năm:**
3. **Độ lệch chuẩn sau 10 năm:**

**Kết luận:** Với giả định là không có ai bỏ việc trong thời gian này thì từ kết quả mà ta tính được có thể thấy được rằng giá trị trung bình mức lương của sinh viên sau 10 năm tăng gấp 20 lần so với mức lương mới ra trường, phương sai sau 10 năm tăng gấp 400 lần so với ban đầu và độ lệch chuẩn của tiền lương sau 10 năm làm việc sẽ tăng gấp 20 lần so với mức mới ra trường.