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**

**THỐNG KÊ LƯƠNG KHỞI ĐIỂM SINH VIÊN MỚI RA TRƯỜNG**

*Người hướng dẫn*: **GV NGUYỄN THỊ HUỲNH TRÂM**

*Người thực hiện*: **NGUYỄN MAI DUY TRƯỜNG - 52200076**

Lớp **: 22050201**

Khoá  **: 26**

**THÀNH PHỐ HỒ CHÍ MINH, 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**

**THỐNG KÊ LƯƠNG KHỞI ĐIỂM SINH VIÊN MỚI RA TRƯỜNG**

Người hướng dẫn: **GV NGUYỄN THỊ HUỲNH TRÂM**

Người thực hiện: **NGUYỄN MAI DUY TRƯỜNG**

Lớp **: 22050201**

Khoá  **: 26**

**THÀNH PHỐ HỒ CHÍ MINH, NĂM 2023**

LỜI CẢM ƠN

Em xin chân thành cảm ơn cô Nguyễn Thị Huỳnh Trâm là giảng viên dạy lý thuyết và thầy Nguyễn Lâm là giảng viên thực hành môn xác suất thống kê đã giúp đỡ và hỗ trợ em về kiến thức chuyên môn trong suốt thời gian 2 tuần giúp em có thể hoàn thành bài tập này một cách tốt nhất.

**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 GV 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 tháng năm*

*Tác giả*

*(ký tên và ghi rõ họ tên)*

*Nguyễn Mai Duy Trường*

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

Vấn đề:bài báo cáo thống kê mức lương khởi điểm của sinh viên mới ra trường bằng cách xem tập dữ liệu 65 sinh viên như một tổng thể và tiến hành thực hiện các thao tác thống kê như tìm trung bình,phương sai, độ lệch chuẩn, khoảng tin cậy, kiểm định giả thuyết và một số thao tác tìm tỉ lệ và xác suất.

Cách giải quyết vấn đề: sử dụng các kiến thức đã được học trên lớp lý thuyết và kết hợp sử dụng ngôn ngữ lập trình python và các tiện ích đã học trên lớp thực hành.

Kết quả đạt được: hiểu sâu hơn và sử dụng thành thạo hơn các thao tác thống kê thông qua việc tự mình thực hiện bài báo cáo và áp dụng các kiến thức đã học.

MỤC LỤC

[LỜI CẢM ƠN i](#_Toc154707365)

[TÓM TẮT iv](#_Toc154707366)

[MỤC LỤC 1](#_Toc154707367)

[DANH MỤC KÍ HIỆU VÀ CHỮ VIẾT TẮT 3](#_Toc154707368)

[DANH MỤC CÁC BẢNG BIỂU, HÌNH VẼ, ĐỒ THỊ 4](#_Toc154707369)

[CHƯƠNG 1 – THỐNG KÊ VÀ TÍNH XÁC SUẤT CỦA BỘ DỮ LIỆU 5](#_Toc154707370)

[1.1 Giá trị trung bình và độ lệch chuẩn của tổng thể 5](#_Toc154707371)

[1.2 Tỉ lệ nữ có mức lương cao hơn mức lương trung bình 5](#_Toc154707372)

[1.3 Tỉ lệ nam có mức lương cao hơn mức lương trung bình 7](#_Toc154707373)

[1.4 Tính xác suất 8](#_Toc154707374)

[1.4.1 Xác suất chọn được người có mức lương thấp hơn mức lương trung bình 8](#_Toc154707375)

[1.4.2 Xác suất chọn được người có mức lương cao hơn mức lương trung bình biết rằng người đó là nam 10](#_Toc154707376)

[1.5 Số cách chọn tập mẫu 30 phần tử và 15 phần tử từ tống thể 11](#_Toc154707377)

[1.6 Phương pháp lấy mẫu 30 phần tử và 15 phần tử 11](#_Toc154707378)

[1.7 Giá trị trung bình và độ lệch chuẩn của tập mẫu 30 phần tử và 15 phần tử 13](#_Toc154707379)

[1.8 Khoảng tin cậy của tập mẫu 30 phần tử và 15 phần tử 15](#_Toc154707380)

[1.9 Kiểm định giả thuyết 17](#_Toc154707381)

[CHƯƠNG 2 – GIẢ SỬ TẬP DỮ LIỆU 65 PHẦN TỬ LÀ TẬP MẪU 18](#_Toc154707382)

[2.1 Tính trung bình và độ lệch chuẩn 18](#_Toc154707383)

[2.2 Tìm khoảng tin cậy 19](#_Toc154707384)

[2.3 Kiểm định giả thuyết 21](#_Toc154707385)

[CHƯƠNG 3 – MỨC LƯƠNG SAU 10 NĂM 22](#_Toc154707386)

[3.1 Giá trị rung bình 22](#_Toc154707387)

[3.2 Phương sai 23](#_Toc154707388)

[3.3 Độ lệch chuẩn 24](#_Toc154707389)

[TÀI LIỆU THAM KHẢO 27](#_Toc154707390)

[PHỤ LỤC 28](#_Toc154707391)

DANH MỤC KÍ HIỆU VÀ CHỮ VIẾT TẮT

**CÁC KÝ HIỆU**

**CÁC CHỮ VIẾT TẮT**

DANH MỤC CÁC BẢNG BIỂU, HÌNH VẼ, ĐỒ THỊ

**DANH MỤC HÌNH**

[Hình 1: Sơ đồ cây chọn một nhóm nam nữ sau đó chọn tiếp một người 8](#_Toc154707050)

[Hình 2: Mã nguồn python chọn 2 tập mẫu từ tổng thể 11](#_Toc154707051)

[Hình 3: Kết quả 2 tập mẫu S1 và S2 12](#_Toc154707052)

[Hình 4: Khoảng tin cậy của 2 tập mẫu 16](#_Toc154707053)

[Hình 5: Khoảng tin cậy của tập mẫu 65 phần tử 20](#_Toc154707054)

**DANH MỤC BẢNG**

[Bảng 1: Mức lương của nữ 5](#_Toc154707085)

[Bảng 2: Mức lương của nam 6](#_Toc154707086)

CHƯƠNG 1 – THỐNG KÊ VÀ TÍNH XÁC SUẤT CỦA BỘ DỮ LIỆU

1.1 Giá trị trung bình và độ lệch chuẩn của tổng thể

-Gọi là giá trị trung bình của tổng thế và N là số quan sát trong tổng thể. Giá trị trung bình của mức lương khởi điểm là:

=

=

Vậy mức lương khởi điểm có giá trị trung bình = 8,503,000 đồng .

-Gọi là độ lệch chuẩn, N là số quan sát trong tổng thể, là giá trị trung bình của tổng thể. Độ lệch chuẩn của mức lương khởi điểm là:

=

=36.41

Vậy mức lương khởi điểm có độ lệch chuẩn = 3,641,000 đồng.

1.2 Tỉ lệ nữ có mức lương cao hơn mức lương trung bình

-Thống kê lại các quan sát là có giới tính nữ và sắp xếp theo thứ tự tăng dần ta được bảng dữ liệu:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| STT | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Giới tính | W | W | W | W | W | W | W | W | W | W |
| Tiền lương | 27 | 40 | 59 | 62 | 69 | 69 | 74 | 79 | 83 | 86 |
| STT | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| Giới tính | W | W | W | W | W | W | W | W | W | W |
| Tiền lương | 92 | 102 | 105 | 109 | 114 | 118 | 119 | 119 | 123 | 124 |
| STT | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 |
| Giới tính | W | W | W | W | W | W | W | W |
| Tiền lương | 125 | 128 | 128 | 130 | 138 | 138 | 146 | 165 |

Bảng 1: Mức lương của nữ

-Với mức lương trung bình khởi điểm = 85.03 ta thấy có 19 quan sát có mức lương cao hơn mức lương trung bình trong tổng số 28 nữ.

-Gọi nfemale là số lương quan sát có mức lương cao hơn mức lương trung bình, N số quan sát của tổng thể, Tfemale là tỉ lệ quan sát nữ có mức lương cao hơn mức lương trung bình so với số lượng của tổng thể. Ta có:

=

= 29.23%

Vậy tỉ lệ nữ có mức lương cao hơn mức lương trung bình so với tổng thể là 29.23%.

-Gọi nfemale là số lương quan sát có mức lương cao hơn mức lương trung bình, Nfemale tổng số nữ, gọi tfemale là tỉ lệ nữ có mức lương trung bình cao hơn mức lương trung bình so với tổng số lượng nữ. Ta có:

=

= 67.86%

Vậy tỉ lệ nữ có mức lương cao hơn mức lương trung bình so với tổng số nữ là 67.86%.

1.3 Tỉ lệ nam có mức lương cao hơn mức lương trung bình

-Thống kê lại các quan sát là có giới tính nam và sắp xếp theo thứ tự tăng dần ta được bảng dữ liệu:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| STT | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Giới tính | M | M | M | M | M | M | M | M | M | M |
| Tiền lương | 24 | 25 | 25 | 30 | 31 | 34 | 41 | 44 | 45 | 46 |
| STT | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| Giới tính | M | M | M | M | M | M | M | M | M | M |
| Tiền lương | 47 | 49 | 52 | 56 | 59 | 61 | 61 | 66 | 67 | 67 |
| STT | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 |
| Giới tính | M | M | M | M | M | M | M | M | M | M |
| Tiền lương | 68 | 77 | 84 | 85 | 86 | 87 | 88 | 91 | 94 | 99 |
| STT | 31 | 32 | 33 | 34 | 35 | 36 | 37 |
| Giới tính | M | M | M | M | M | M | M |
| Tiền lương | 99 | 105 | 106 | 131 | 139 | 143 | 144 |

Bảng 2: Mức lương của nam

-Với mức lương trung bình khởi điểm = 85.03 ta thấy có 13 quan sát có mức lương cao hơn mức lương trung bình trong tổng số 37 nam.

-Gọi nmale là số lương quan sát nam có mức lương cao hơn mức lương trung bình, N số quan sát của tổng thể, Tmale là tỉ lệ quan sát nam có mức lương cao hơn mức lương trung bình so với số lượng của tổng thể. Ta có:

=

= 20%

Vậy tỉ lệ nam có mức lương cao hơn mức lương trung bình so với tổng thể là 20%.

-Gọi nmale là số lương quan sát nam có mức lương cao hơn mức lương trung bình, Nmale tổng số nam, gọi tmale là tỉ lệ nam có mức lương trung bình cao hơn mức lương trung bình so với tổng số lượng nam. Ta có:

=

= 35.13%

Vậy tỉ lệ nam có mức lương cao hơn mức lương trung bình so với tổng số nam là 35.13%.

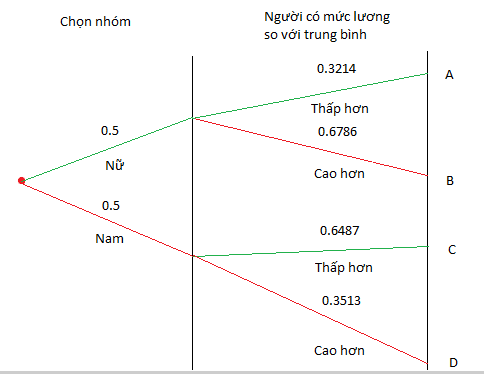
1.4 Tính xác suất

1.4.1 Xác suất chọn được người có mức lương thấp hơn mức lương trung bình

-Đặt biến cố:

* Gọi A là biến cố chọn được người có mức lương thấp hơn mức lương trung bình từ nhóm nữ
* Gọi B là biến cố chọn được người có mức lương cao hơn mức lương trung bình từ nhóm nữ
* Gọi C là biến cố chọn được người có mức lương thấp hơn mức lương trung bình từ nhóm nam.
* Gọi D là biến cố chọn được người có mức lương cao hơn mức lương trung bình từ nhóm nam
* Gọi E là biến cố chọn được người có mức lương thấp hơn mức lương trung bình từ tổng thể

-Vẽ sơ đồ cây:



Hình 1: Sơ đồ cây chọn một nhóm nam nữ sau đó chọn tiếp một người

-Tính xác suất biến cố:

* Xác suất chọn được người có mức lương thấp hơn lương trung bình từ nhóm nữ là :

P(A) = 0.50.3214 = 0.1607

Vậy xác suất chọn được người có mức lương thấp hơn lương trung bình từ nhóm nữ là 0.1607

* Xác suất chọn được người có mức lương cao hơn lương trung bình từ nhóm nữ là :

P(B) = 0.50.6786 = 0.3393

Vậy xác suất chọn được người có mức lương cao hơn lương trung bình từ nhóm nữ là 0.3393

* Xác suất chọn được người có mức lương thấp hơn lương trung bình từ nhóm nam là :

P(C) = 0.50.6487 = 0.32435

Vậy xác suất chọn được người có mức lương thấp hơn lương trung bình từ nhóm nam là 0.32435

* Xác suất chọn được người có mức lương cao hơn lương trung bình từ nhóm nam là :

P(D) = 0.50.3513 = 0.17565

Vậy xác suất chọn được người có mức lương cao hơn lương trung bình từ nhóm nữ là 0.17565

* Xác suất chọn được người có mức lương thấp hơn lương trung bình từ tổng thể là :

P(E) = P(A) + P(C) = 0.1607 + 0.32435 = 0.48505

Vậy xác suất chọn được người có mức lương thấp hơn lương trung bình từ tổng thể là 0.48505

1.4.2 Xác suất chọn được người có mức lương cao hơn mức lương trung bình biết rằng người đó là nam

-Đặt biến cố:

* Gọi F là biến cố chọn được nam
* Gọi G là biến cố chọn được người có mức lương cao hơn lương trung bình.

-Tính xác suất:

* Xác suất chọn được nhóm nam theo như đề bài ta có:

P(F) = 0.5

* Xác suất chọn được người có mức lương cao hơn mức lương trung bình biết rằng người đó là nam là:

P(G|F) = =

=

Vậy xác suất chọn được người có mức lương cao hơn mức lương trung bình biết rằng người đó là nam là 0.3513.

1.5 Số cách chọn tập mẫu 30 phần tử và 15 phần tử từ tống thể

* Gọi S1 là tập mẫu gồm 30 sinh viên được lấy từ tổng thể 65 sinh viên. Số cách chọn tập mẫu S1 là:

( cách)

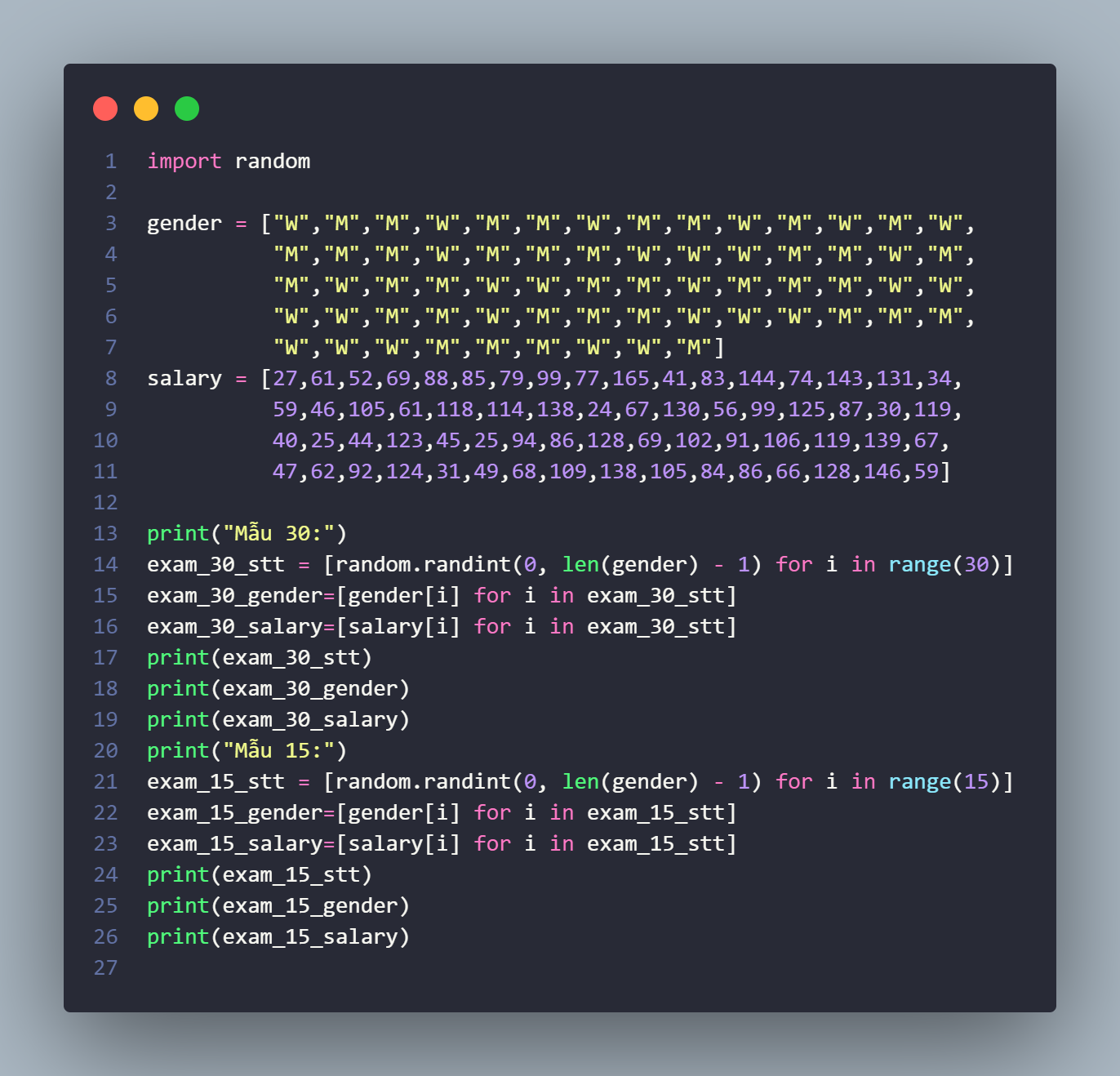
Vậy có 3009106305270645216 cách để chọn tập mẫu S1.

* Gọi S2 là tập mẫu gồm 15 sinh viên được lấy từ tổng thể 65 sinh viên. Số cách chọn tập mẫu S2 là:

( cách)

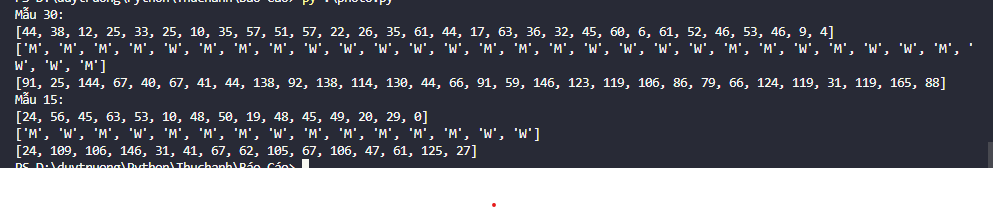
Vậy có 207374699821536 cách để chọn tập mẫu S2.

1.6 Phương pháp lấy mẫu 30 phần tử và 15 phần tử

* Sử dụng sự trợ giúp của ngôn ngữ lập trình Python và thư viện random
* Đầu tiên tạo 2 danh sách lần lượt chứa giới tính và mức lương khởi điểm tương ứng nhau theo như đề bài.
* Tạo hai mảng để lưu trữ các số thứ tự của tập mẫu 30 phần tử và tập mẫu 15 phần tử.
* Sử dụng vòng lặp for kết hợp với hàm randint() của thư viện random để lấy ngẫu nhiên 30 số thứ tự từ 0 đến 64 trong mảng Python tương ứng với số thứ tự từ 1 đến 65 của đề bài và tập mẫu 15 phần tử cũng sẽ lấy 15 số thứ tự ngẫu nhiên theo cách tương tự.
* Tạo 2 mảng chứa giới tính và 2 mảng chứa mức lương của 2 tập mẫu.
* Đối với việc lấy giới tính và mức lương cho tập mẫu thì sử dụng cùng một phương pháp là dùng vòng lặp for duyệt từ mảng lưu số thứ tự đã lấy ngẫu nhiên trước đó và lấy trực tiếp giá trị từ việc gọi index của số thứ tự đó từ 2 mảng giới tính và mức lương khởi điểm đã tạo từ lúc đầu.

Hình 2: Mã nguồn python chọn 2 tập mẫu từ tổng thể

* Kết quả ta sẽ có được 2 tập mẫu tương ứng với 4 bảng dữ liệu bao gồm: S1 - Tập mẫu 30 phần tử ( 1 mảng chứa giới tính và một mảng chứa mứclương) và S2 - Tập mẫu 15 phần tử ( 1 mảng chứa giới tính và một mảng chứa mức lương).

Hình 3: Kết quả 2 tập mẫu S1 và S2

1.7 Giá trị trung bình và độ lệch chuẩn của tập mẫu 30 phần tử và 15 phần tử

-Tập mẫu S1:

* Gọi là giá trị trung bình của mức lương của S1 và n là cỡ mẫu. Giá trị trung bình của mức lương của S1 là:

=

=

Vậy mức lương của tập mẫu S1 có giá trị trung bình = 9,207,000 đồng .

* Gọi là độ lệch chuẩn của tập mẫu S1, n là cỡ mẫu, là giá trị trung bình mức lương của S1. Độ lệch chuẩn của mức lương tập mẫu S1 là:

=

= 38.736

Vậy mức lương của S1 có độ lệch chuẩn = 3,873,600 đồng.

-Tập mẫu S2:

* Gọi là giá trị trung bình của mức lương của S2 và n là cỡ mẫu. Giá trị trung bình của mức lương của S2 là:

=

=

Vậy mức lương của tập mẫu S2 có giá trị trung bình = 7,493,000 đồng .

* Gọi là độ lệch chuẩn của tập mẫu S2, n là cỡ mẫu, là giá trị trung bình mức lương của S2. Độ lệch chuẩn của mức lương tập mẫu S2 là:

=

= 38.538

Vậy mức lương của S2 có độ lệch chuẩn = 3,853,800 đồng.

-Kết luận:

* Giá trị trung bình của cả 2 tập mẫu đều khác biệt và chênh lệch đáng kể so với giá trị trung bình của mức lương khởi điểm.
* Độ lệch chuẩn của cả 2 tập mẫu và độ lệch chuẩn của mức lương khởi điểm chệnh lệch khá nhỏ và khá tương đồng với nhau.

1.8 Khoảng tin cậy của tập mẫu 30 phần tử và 15 phần tử

-Độ tin cậy 80%:

+Với là mức ý nghĩa ta có công thức xác định độ tin cậy như sau:

Độ tin cậy = (1 -

⬄ 80% = (1 -

⬄

Vậy mức ý nghĩa của độ tin cậy 80% là 0.2.

+Dựa vào đồ thị xác suất tích lũy cho phân phối chuẩn ta được = 1.28

+Tập mẫu S1:

Với là giá trị trung bình của mẫu S1, là độ lệch chuẩn của tổng thể, n là cỡ mẫu. Khoảng tin cậy của tập mẫu S1 là:

=

= (83.56 , 100.58)

Vậy giá trị trung bình mức lương khởi điểm của sinh viên khoảng từ 8,356,000 đến 10,058,000 đồng với độ tin cậy 80%.

+Tập mẫu S2:

Với là giá trị trung bình của mẫu S2, là độ lệch chuẩn của tổng thể, n là cỡ mẫu. Khoảng tin cậy của tập mẫu S2 là:

=

= (62.9 , 86.96)

Vậy giá trị trung bình mức lương khởi điểm của sinh viên khoảng từ 6,290,000 đến 8,696,000 đồng với độ tin cậy 80%.

-Độ tin cậy 95%:

+Với là mức ý nghĩa ta có công thức xác định độ tin cậy như sau:

Độ tin cậy = (1 -

⬄ 95% = (1 -

⬄

Vậy mức ý nghĩa của độ tin cậy 95% là 0.05

+Dựa vào đồ thị xác suất tích lũy cho phân phối chuẩn ta được = 1.96

+Tập mẫu S1:

Với là giá trị trung bình của mẫu S1, là độ lệch chuẩn của tổng thể, n là cỡ mẫu. Khoảng tin cậy của tập mẫu S1 là:

=

= (79.04 , 105.1)

Vậy giá trị trung bình mức lương khởi điểm của sinh viên khoảng từ 7,904,000 đến 10,510,000 đồng với độ tin cậy 95%.

+Tập mẫu S2:

Với là giá trị trung bình của mẫu S2, là độ lệch chuẩn của tổng thể, n là cỡ mẫu. Khoảng tin cậy của tập mẫu S2 là:

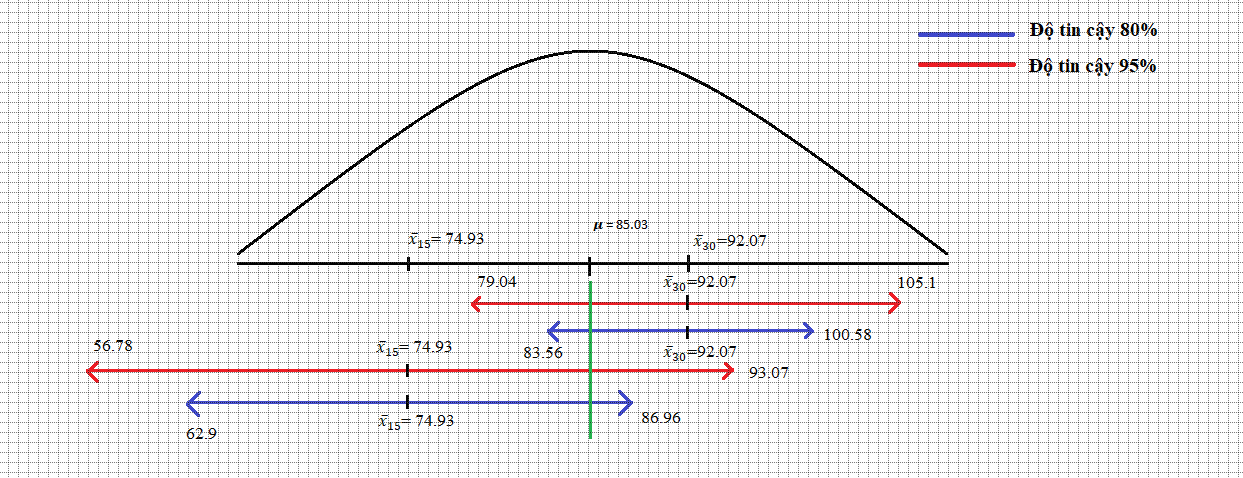
=

= (56.78 , 93.07)

Vậy giá trị trung bình mức lương khởi điểm của sinh viên khoảng từ 5,678,000 đến 9,307,000 đồng với độ tin cậy 95%.

-Kết luận:

* So với tổng thể, các khoảng tin cậy của cả hai tập mẫu đều chứa giá trị trung bình của tổng thể là 85.03
* Giữa 2 tập mẫu thì độ chính xác của mẫu S1 tốt hơn mẫu S2 do độ rộng khoảng tin cậy nhỏ hơn so với mẫu S2.



Hình 4: Khoảng tin cậy của 2 tập mẫu

1.9 Kiểm định giả thuyết

* = 92.07
* Mức ý nghĩa = 0.05
* Khoảng tin cậy: 79.04 đến 105.1
* Chọn = 81
* Dự đoán rằng giá trị trung bình của tổng thể là 81
* Với là giá trị trung bình của mẫu S1, là giá trị trung bình của tổng thể dự đoán, là độ lệch chuẩn của tổng thể, n là cỡ mẫu.Test thống kê:

z =

= = 1.665

* Với mức ý nghĩa = 0.05 và dựa vào đồ thị xác suất tích lũy cho phân phối chuẩn ta được = 1.64
* Giả thuyết: : 81, : <81

-Theo phương pháp giá trị tới hạn thì z = 1.665 > -= -1.64

-Do đó, không thể bác bỏ giải thuyết : 81 vì không đủ bằng chứng thống kê chứng minh rằng : <81.

=>Đúng với thực tế, không mắc sai lầm.

* Giả thuyết: : 81, : >81

-Theo phương pháp giá trị tới hạn thì z = 1.665 > = 1.64

-Do đó, có thể bác bỏ giải thuyết : 81 vì đã có đủ bằng chứng thống kê chứng minh rằng : >81.

=>Đúng với thực tế, không mắc sai lầm.

* Giả thuyết: : = 81, : 81

-Với mức ý nghĩa = 0.05 và dựa vào đồ thị xác suất tích lũy cho phân phối chuẩn ta được = 1.96

-Theo phương pháp giá trị tới hạn thì z = 1.665 > -= -1.96

z =1.665 < = 1.96

-Do đó, không thể bác bỏ giải thuyết : = 81 vì không đủ bằng chứng thống kê chứng minh rằng : 81

=> Không đúng với thực tế, mắc sai lầm loại II

CHƯƠNG 2 – GIẢ SỬ TẬP DỮ LIỆU 65 PHẦN TỬ LÀ TẬP MẪU

2.1 Tính trung bình và độ lệch chuẩn

-Gọi là giá trị trung bình của mức lương của tập mẫu và n là cỡ mẫu. Giá trị trung bình của mức lương trong tập mẫu 65 phần tử là:

=

=

Vậy mức lương của tập mẫu có giá trị trung bình = 8,503,000 đồng .

-Gọi là độ lệch chuẩn của tập mẫu 65 phần tử, n là cỡ mẫu, là giá trị trung bình mức lương của tập mẫu. Độ lệch chuẩn của mức lương tập mẫu là:

=

=36.69

Vậy mức lương của S1 có độ lệch chuẩn = 3,669,000 đồng.

2.2 Tìm khoảng tin cậy

-Độ tin cậy 80%:

+Với là mức ý nghĩa ta có công thức xác định độ tin cậy như sau:

Độ tin cậy = (1 -

⬄ 80% = (1 -

⬄

Vậy mức ý nghĩa của độ tin cậy 80% là 0.2.

+Giả sử tập 65 phần tử đã cho là một tập mẫu:

Dựa vào đồ thị xác suất tích lũy cho phân phối chuẩn với cỡ mẫu là 65 và = 0.1 ta được = 1.282

Với là giá trị trung bình của mẫu, là độ lệch chuẩn của tập mẫu, n là cỡ mẫu. Khoảng tin cậy của tập tập mẫu là:

=

= (79.2 , 90.86)

Vậy giá trị trung bình mức lương khởi điểm của sinh viên khoảng từ 8,503,000 đến 9,086,000 đồng với độ tin cậy 80%.

-Độ tin cậy **95**%:

+Với là mức ý nghĩa ta có công thức xác định độ tin cậy như sau:

Độ tin cậy = (1 -

⬄ 95% = (1 -

⬄

Vậy mức ý nghĩa của độ tin cậy 95% là 0.05

+Giả sử tập 65 phần tử đã cho là một tập mẫu:

Dựa vào đồ thị xác suất tích lũy cho phân phối chuẩn với cỡ mẫu là 65 và = 0.025 ta được = 1.96

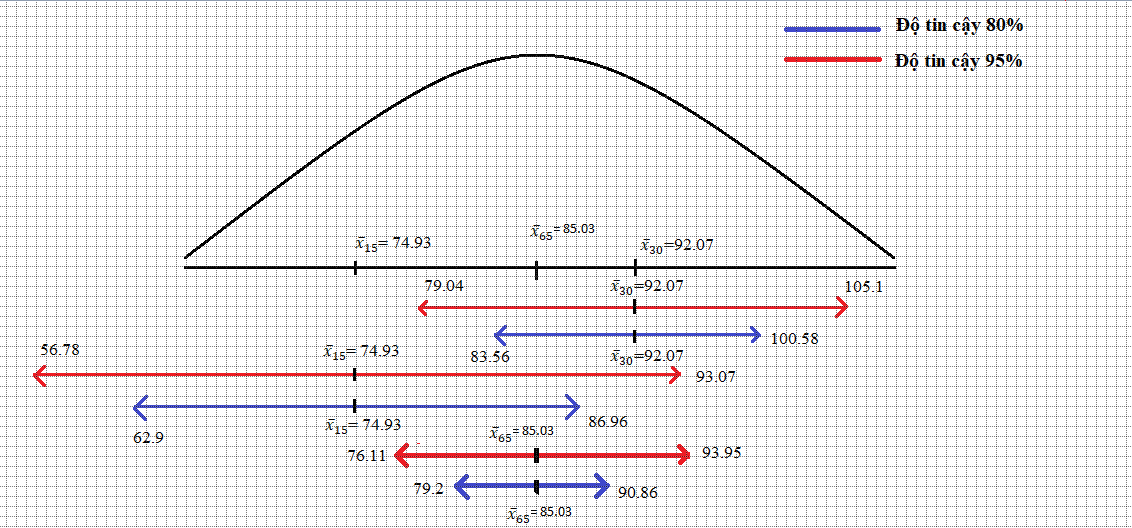
Với là giá trị trung bình của mẫu, là độ lệch chuẩn của tập mẫu, n là cỡ mẫu. Khoảng tin cậy của tập tập mẫu là:

=

= (76.11 , 93.95)

Vậy giá trị trung bình mức lương khởi điểm của sinh viên khoảng từ 7,611,000 đến 9,395,000 đồng với độ tin cậy 95%.

-Kết luận: độ rộng khoảng tin cậy của tập mẫu 65 phần tử này nhỏ hơn so với 2 tập mẫu S1 và S2 trước đó.



Hình 5: Khoảng tin cậy của tập mẫu 65 phần tử

2.3 Kiểm định giả thuyết

* = 85.03
* Mức ý nghĩa = 0.05
* Khoảng tin cậy: 76.11 đến 93.95
* Chọn = 81
* Dự đoán rằng giá trị trung bình của tổng thể là 81
* Với là giá trị trung bình của mẫu 65 phần tử, là giá trị trung bình của tổng thể dự đoán, *s* là độ lệch chuẩn của tập mẫu 65 phần tử, n là cỡ mẫu.Test thống kê:

t =

= = 0.88555

* Với mức ý nghĩa = 0.05 và dựa vào đồ thị xác suất tích lũy cho phân phối chuẩn ta được = 1.645
* Giả thuyết: : 81, : <81

-Theo phương pháp giá trị tới hạn thì t = 0.88555 > -= -1.645

-Do đó, không thể bác bỏ giải thuyết : 81 vì không đủ bằng chứng thống kê chứng minh rằng : <81.

* Giả thuyết: : 81, : >81

-Theo phương pháp giá trị tới hạn thì t = 0.88555 < = 1.645

-Do đó, không thể bác bỏ giải thuyết : 81 vì không đủ bằng chứng thống kê chứng minh rằng : >81.

* Giả thuyết: : = 81, : 81

-Với mức ý nghĩa = 0.05 và dựa vào đồ thị xác suất tích lũy cho phân phối chuẩn ta được = 1.96

-Theo phương pháp giá trị tới hạn thì t = 0.88555 > -= -1.96

t =0.88555 < = 1.96

-Do đó, không thể bác bỏ giải thuyết : = 81 vì không đủ bằng chứng thống kê chứng minh rằng : 81

=> Không thể kết luận việc kiểm tra các giả thuyết ở trên có mắc sai lầm hay không vì ta không thể biết được giá trị trung bình của tổng thể.Tuy nhiên, nếu việc kiểm định giả thuyết ở trên tất cả đều chính xác thì giá trị trung bình của tổng thể ( mức lương khởi điểm trung bình của sinh viên vừa ra trường) là 8,000,000 đồng.

CHƯƠNG 3 – MỨC LƯƠNG SAU 10 NĂM

3.1 Giá trị rung bình

Với là giá trị trung bình của mức lương khởi điểm, là giá trị lương của mỗi người, N số người trong tổng thể và giả sử mỗi năm mức lương mỗi người tăng 20%. Ta có công thức tính giá trị trung bình của mức lương khởi điểm:

Giá trị trung bình của mức lương sau 1 năm :

Giá trị trung bình của mức lương sau 2 năm :

Giá trị trung bình của mức lương sau 3 năm :

=

Suy ra giá trị trung bình của mức lương sau n năm nếu mỗi năm mức lương mỗi người tăng 20% là:

Kết luận: Nếu mức lương mỗi người mỗi năm đều tăng 20% thì sau mỗi năm mức lương trung bình sẽ tăng theo cấp số nhân.

Từ đó suy ra mức lương trung bình sau 10 năm là :

=526.483

Vậy mức lương trung bình sau 10 năm là 52,643,000 đồng sau 10 năm gấp 6.19 lần mức lương trung bình khởi điểm.

3.2 Phương sai

Với là phương sai của mức lương khởi điểm, là giá trị trung bình của mức lương khởi điểm; là giá trị lương của mỗi người, N số người trong tổng thể và giả sử mỗi năm mức lương mỗi người tăng 20%. Ta có công thức tính phương sai của mức lương khởi điểm:

Phương sai của mức lương sau 1 năm :

=

=

Phương sai của mức lương sau 2 năm :

=

=

Phương sai của mức lương sau 3 năm :

=

=

Suy ra phương sai của mức lương sau n năm nếu mỗi năm mức lương mỗi người tăng 20% là:

Kết luận: Nếu mức lương mỗi người mỗi năm đều tăng 20% thì sau mỗi năm phương sai sẽ tăng theo cấp số nhân.

Từ đó suy ra mức lương trung bình sau 10 năm là :

=50823.7

Vậy phương sai sau 10 năm gấp 38.34 lần phương sai khởi điểm.

3.3 Độ lệch chuẩn

Với là độ lệch chuẩn của mức lương khởi điểm, là giá trị trung bình của mức lương khởi điểm; là giá trị lương của mỗi người, N số người trong tổng thể và giả sử mỗi năm mức lương mỗi người tăng 20%. Ta có công thức tính độ lệch chuẩn của mức lương khởi điểm:

Độ lệch chuẩn của mức lương sau 1 năm :

=

=

Độ lệch chuẩn của mức lương sau 2 năm :

=

=

Độ lệch chuẩn của mức lương sau 3 năm :

=

=

Suy ra độ lệch chuẩn của mức lương sau n năm nếu mỗi năm mức lương mỗi người tăng 20% là:

Kết luận: Nếu mức lương mỗi người mỗi năm đều tăng 20% thì sau mỗi năm độ lệch chuẩn sẽ tăng theo cấp số nhân.

Từ đó suy ra độ lệch chuẩn sau 10 năm là :

=225.44

Vậy độ lệch chuẩn sau 10 năm là 22,544,000 đồng gấp 6.19 lần độ lệch chuẩn khởi điểm bằng với tỉ lệ của mức lương trung bình sau 10 năm so với mức lương khởi điểm.

TÀI LIỆU THAM KHẢO

**Tiếng Việt**

1. [Giáo trình môn học xác suất thông kê cung cấp bởi GV Nguyễn Thị Huỳnh Trâm](https://drive.google.com/drive/u/1/folders/10rQCCK1hhMyUEQPDuuQaV4CcM8G6Js1F?fbclid=IwAR1imE5FPttWvhcQVm8cYaaNFw0aehTfSAZwApw-bOkKWcMBCA0Z_5u6QVI)
2. [Xác suất có điều kiện](https://vi.wikipedia.org/wiki/X%C3%A1c_su%E1%BA%A5t_c%C3%B3_%C4%91i%E1%BB%81u_ki%E1%BB%87n)

**Tiếng Anh**

PHỤ LỤC