TỔNG LIÊN ĐOÀN LAO ĐỘNG VIỆT NAM

**TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

**KHOA CÔNG NGHỆ THÔNG TIN**



**ĐẶNG THÀNH NHÂN – 522H0006**

**BÁO CÁO CUỐI KỲ**

**XÁC SUẤT VÀ THỐNG KÊ**

**ỨNG DỤNG CHO CÔNG NGHỆ THÔNG TIN**

**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**



**ĐẶNG THÀNH NHÂN – 522H0006**

**BÁO CÁO CUỐI KỲ**

**XÁC SUẤT VÀ THỐNG KÊ**

**ỨNG DỤNG CHO CÔNG NGHỆ THÔNG TIN**

Người hướng dẫn

**TRẦN LƯƠNG QUỐC ĐẠI**

**LÊ TRỌNG DIỆU HIỀN**

**THÀNH PHỐ HỒ CHÍ MINH, NĂM 2023**

**LỜI CẢM ƠN**

Lời đầu tiên em xin gửi lời cảm ơn sâu sắc tới thầy Trần Lương Quốc Đại và cô Lê Trọng Diệu Hiền đã dành thời gian để hỗ trợ và giúp đỡ em trong quá trình học tập và hoàn thành bài tiểu luận. Những lời khuyên và nhận xét giúp em hiểu rõ hơn và hoàn thành bài tiểu luận này. Do chưa có nhiều kinh nghiệm làm tiểu luận cũng như những hạn chế về kiến thức nên trong bài tiểu luận chắc chắn sẽ có những sai sót. Rất mong nhận được những nhận xét và ý kiến đóng góp từ thầy để em có thể hoàn thiện hơn. Lời cuối cùng em chân thành cảm ơn thầy đã đọc và cho đánh giá. Chúc thầy có nhiều sức khỏe, thành công và hạnh phúc.

*TP. Hồ Chí Minh, ngày ... tháng … năm 20..*

*Tác giả*

*(Ký tên và ghi rõ họ tên)*



*Đặng Thành Nhân*

**CÔNG TRÌNH ĐƯỢC HOÀN THÀNH**

**TẠI TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

Tôi xin cam đoan đây là công trình nghiên cứu của riêng tôi và được sự hướng dẫn khoa học của thầy Trần Lương Quốc Đại và cô Lệ Trọng Diệu Hiền. 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 Dự 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 Dự 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 20..*

*Tác giả*

*(Ký tên và ghi rõ họ tên)*



*Đặng Thành Nhân*

**MỤC LỤC**

[CHƯƠNG 1. PHẦN I 1](#_Toc154690608)

[1.1 Câu 1 1](#_Toc154690609)

[1.2 Câu 2 3](#_Toc154690610)

[1.3 Câu 3 3](#_Toc154690611)

[1.4 Câu 4 4](#_Toc154690612)

[1.5 Câu 5 6](#_Toc154690613)

[1.6 Câu 6 7](#_Toc154690614)

[1.7 Câu 7 9](#_Toc154690615)

[1.8 Câu 8 10](#_Toc154690616)

[1.9 Câu 9 13](#_Toc154690617)

[CHƯƠNG 2. PHẦN II 15](#_Toc154690618)

[2.1 Câu 10 15](#_Toc154690619)

[CHƯƠNG 3. PHẦN III 17](#_Toc154690620)

[3.1 Câu 11 17](#_Toc154690621)

[TÀI LIỆU THAM KHẢO 18](#_Toc154690622)

# PHẦN I

## Câu 1

* Yêu cầu: Tính giá trị trung bình và độ lệch chuẩn của mức lương khởi điểm
* Tính giá trị trung bình:

***Công thức tính giá trị trung bình:***

***Trong đó:***

* : là giá trị trung bình của mức lương khởi điểm của các sinh viên.
* **n**: là tổng số lượng sinh viên.
* : là mức lương khởi điểm tương ứng với sinh viên thứ n trong tập.
* Dựa trên tập dữ liệu mức lương khởi điểm của 65 sinh viên mới ra trường. Ta có:
  + Vậy giá trị trung bình của mức lương khởi điểm của các sinh viên mới ra trường là:
* Tính phương sai:

***Công thức tính phương sai:***

***Trong đó:***

* : là giá trị phương sai của mức lương khởi điểm của các sinh viên.
* : là giá trị trung bình của mức lương khởi điểm của các sinh viên.
* **n**: là tổng số lượng sinh viên.
* : là mức lương khởi điểm tương ứng với sinh viên thứ **n** trong tập.
* Dựa trên tập dữ liệu mức lương khởi điểm của 65 sinh viên mới ra trường. Ta có:
* Tính độ lệch chuẩn:

***Công thức tính độ lệch chuẩn:***

***Trong đó:***

* là giá trị độ lệch chuẩn của mức lương khởi điểm của các sinh viên.
* là giá trị phương sai của mức lương khởi điểm của các sinh viên.
* Dựa trên tập dữ liệu mức lương khởi điểm của 65 sinh viên mới ra trường. Ta có:
  + Vậy độ lệch chuẩn mức lương khởi điểm của các sinh viên là

## Câu 2

* Yêu cầu: Tính tỉ lệ nữ có mức lương cao hơn mức lương trung bình
* Gọi tỉ lệ sinh viên nữ có mức lương khởi điểm cao hơn mức trung bình là.
* là số sinh viên nữ có mức lương trên trung bình.
* là tổng số sinh viên nữ.

***Ta có:***

* Số lượng sinh viên nữ có mức lương trên mức trung bình là 19 trong tổng 28 sinh viên nữ.

***Tỉ lệ sinh viên nữ có mức lương trên mức trung bình là:***

* + Vậy tỉ lệ nữ có mức lương trên mức trung bình là

## Câu 3

* Yêu cầu: Tính tỉ lệ nam có mức lương cao hơn mức lương trung bình
* Gọi tỉ lệ một sinh viên là sinh viên nam là.
* Gọi là số sinh viên nam có mức lương trên mức trung bình.
* là tổng số sinh viên nam.

***Ta có:***

* Số lượng sinh viên nam có mức lương trên mức trung bình là 13 trong tổng số 37 sinh viên nam

***Tỉ lệ sinh viên nam có mức lương trên mức trung bình là:***

* + Vậy tỉ lệ nữ có mức lương trên mức trung bình là

## Câu 4

* Yêu cầu 1: Tính xác suất chọn được người có mức lương thấp hơn mức lương trung bình
* Gọi A là biến cố chọn được một sinh viên có mức lương dưới mức trung bình từ nhóm nữ.
* Gọi B là biến cố chọn được một sinh viên có mức lương dưới mức trung bình từ nhóm nam.
* C là biến cố chọn được một sinh viên có mức lương dưới mức trung bình từ nhóm nam và nữ.

***Ta có:***

* Xác suất chọn ra một sinh viên từ một nhóm nam và nữ là
* Số lượng sinh viên nữ có mức lương dưới mức trung bình là 9 trong tổng số 28 sinh viên nữ.
* Số lượng sinh viên nam có mức lương dưới mức trung bình là 24 trong tổng số 37 sinh viên nam.

***Xác xuất để chọn được một sinh viên nữ có mức lương thấp hơn mức trung bình là:***

***Xác xuất để chọn được một sinh viên nam có mức lương thấp hơn mức trung bình là:***

***Xác xuất để chọn được một sinh viên có mức lương thấp hơn mức trung bình là:***

* + Vậy xác suất để chọn ra một sinh viên có mức lương thấp hơn mức trung bình từ 2 nhóm nam và nữ là: 0.485
* Yêu cầu 2: Tính xác suất chọn được một người có mức lương cao hơn mức lương trung bình biết rằng người đó là nam.
* Gọi D là biến cố chọn được một sinh viên có mức lương cao trên mức trung bình.
* Gọi E là biến cố chọn được một sinh viên nam có mức lương cao trên mức trung bình từ 2 nhóm nam và nữ.

***Ta có:***

* Xác suất chọn ra một sinh viên từ một nhóm nam và nữ là
* Số lượng sinh viên nam có mức lương trên mức trung bình là 13 trong tổng số 37 sinh viên nam

***Xác xuất để chọn được một sinh viên có mức lương cao hơn mức trung bình là:***

***Xác xuất để chọn được một sinh viên nam có mức lương cao hơn mức trung bình là:***

* + Vậy xác suất xác xuất để chọn được một sinh viên nam có mức lương cao hơn mức trung bình là 0.3411

## Câu 5

* 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?

***Công thức lấy một tập mẫu từ một tập hợp:***

***Trong đó:***

* là số lượng phần tử trong dữ liệu (n > 0)
* là số lượng phần tử của tập hợp con trong dữ liệu (k <= n)
* Ta có:

S1 = = = 3009106305270645248 (Cách)

S2 = = = 207374699821536 (Cách)

* + Vậy có 3009106305270645248 cách để lấy một tập mẫu S1 và có 207374699821536 cách để lấy một tập mẫu S2

## Câu 6

* Yêu cầu: Diễn giải chi tiết phương pháp lấy ngẫu nhiên tập S1, S2
* Cách tìm tập và S2: Sử dụng hàm trong thư viện ngẫu nhiên (random) của Python được dùng để lấy mẫu ngẫu nhiên từ các sinh viên. Cú pháp sử dụng hàm:

***Trong đó:***

* : là tập hợp các phần tử ban đầu có thể là tuple, string, list, … muốn lấy mẫu.
* : là số lượng phần tử ngẫu nhiên muốn lấy ra.
* Tìm tập S1: Sau khi chạy source code dưới đây sẽ tìm ra tập S1 gồm 30 sinh viên từ tập các sinh viên ban đầu.
* A screenshot of a computer code

  Description automatically generatedSource Code tìm :

Hình 1.1: Source Code Tìm S1

* Kết quả tập S1 thu được:

A table of numbers and letters

Description automatically generated

Hình 1.2: Kết quả tập S1

* A screenshot of a computer program

  Description automatically generatedSource Code tìm :

Hình 2.1: Source Code Tìm S2

* Kết quả tập thu được:

A table of numbers and letters

Description automatically generated

Hình 2.2 Kết quả tập S2

## Câu 7

* Yêu cầu: Tính giá trị trung bình mức lương khởi điểm của tập S1, S2. Nêu kết luận về các giá trị này so với giá trị trung bình của tổng thể.
* Tính giá trị trung bình của tập S1:
* Dựa vào tập S1 ở Câu 6, Ta có:

* Tính giá trị trung bình của tập S2:
* Dựa vào tập S2 ở Câu 6, Ta có:

* + Dựa vào gái trị trung bình của tập S1 và S2 vừa tìm được ta có thể thấy giá trị trung bình tổng thể (85.0308) nằm ở giữa 2 giá trị trung bình của tập S1(77.8) và S2(91.93).

## Câu 8

* 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.

***Công thức tính khoảng tin cậy:***

***Trong đó:***

* : là giá trị trung bình của tập dữ liệu.
* : là độ lệch chuẩn của tập dữ liệu.
* : là kích thước của tập dữ liệu.
* : là là giá trị lấy từ bảng T-Student tương ứng với hệ số tin cậy.
* *:* là giá trị .

***Xét tập S1:***

* Ta có:
* Giá trị phương sai của tập S1:
* Giá trị độ lệch chuẩn của tập S1:
* Khoảng tin cậy của tập S1 với độ tin cậy 80% là:

Độ tin cậy là 80% =>

* + Với độ tin cậy 80% thì khoảng tin cậy đối với tập S1 nằm trong khoảng
* Khoảng tin cậy của tập S1 với độ tin cậy 95% là:

Độ tin cậy là 95% =>

* + Với độ tin cậy 95% thì khoảng tin cậy đối với tập S1 nằm trong khoảng

***Xét tập S2:***

* Ta có:
* Giá trị phương sai của tập S2:
* Giá trị độ lệch chuẩn của tập S2:
* Khoảng tin cậy của tập S2 với độ tin cậy 80% là:

Độ tin cậy là 80% =>

* + Với độ tin cậy 95% thì khoảng tin cậy đối với tập S2 nằm trong khoảng
* Khoảng tin cậy của tập S2 với độ tin cậy 95% là:

Độ tin cậy là 95% =>

* + Với độ tin cậy 95% thì khoảng tin cậy đối với tập S2 nằm trong khoảng (

## Câu 9

* 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?

***Công thức tính mức kiểm định:***

***Trong đó:***

* ***:*** là giá trị trung bình của tập mẫu.
* ***:*** là độ lệch chuẩn của tập mẫu.
* ***:*** là giá trị trung bình của tập mẫu.
* ***:*** là giá trị trung bình của tổng thể.
* ***:*** là kích thước của tập mẫu.

***Cặp giả thuyết và miền bác bỏ:***

***Trong đó:***

* **:** là giả thuyết gốc
* là giả thuyết đối
* là miền bác bỏ
* : là là giá trị lấy từ bảng T-Student tương ứng với hệ số tin cậy.

***Ta có:***

Mức ý nghĩa

Mức kiểm định:

***Xét giả thuyết trái :***

Với miền bác bỏ

(Sai)

* + Với mức ý nghĩa 0.05, không thể cho rằng giá trị trung bình của tập mẫu S1 đúng với thực tế.

***Xét giả thuyết phải :***

Với miền bác bỏ

(Sai)

* + Với mức ý nghĩa 0.05, không thể cho rằng giá trị trung bình của tập mẫu S1 đúng với thực tế.

***Xét giả thuyết 2 bên của giá trị trung bình :***

Với miền bác bỏ

(Sai)

* + Với mức ý nghĩa 0.05, không thể cho rằng giá trị trung bình của tập mẫu S1 đúng với thực tế.

# PHẦN II

## Câu 10

* Yêu cầu: 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. Hãy thực hiện lại các bước tính toán theo yêu cầu trong câu 8, 9.

***Xét tập X gồm 65 sinh viên trên:***

***Ta có:***

* Khoảng tin cậy của tập S1 với độ tin cậy 80% là:

Độ tin cậy là 80% =>

* + Với độ tin cậy 80% thì khoảng tin cậy đối với tập S1 nằm trong khoảng
* Khoảng tin cậy của tập S1 với độ tin cậy 95% là:

Độ tin cậy là 95% =>

* + Với độ tin cậy 95% thì khoảng tin cậy đối với tập S1 nằm trong khoảng

***Ta có:***

Mức ý nghĩa

Mức kiểm định:

***Xét giả thuyết trái :***

Với miền bác bỏ

(Sai)

* + Với mức ý nghĩa 0.05, không thể cho rằng giá trị trung bình của tập mẫu X đúng với thực tế.

***Xét giả thuyết phải :***

Với miền bác bỏ

(Sai)

* + Với mức ý nghĩa 0.05, không thể cho rằng giá trị trung bình của tập mẫu X đúng với thực tế.

***Xét giả thuyết 2 bên của giá trị trung bình :***

Với miền bác bỏ

(Sai)

* + Với mức ý nghĩa 0.05, không thể cho rằng giá trị trung bình của tập mẫu X đúng với thực tế.

# PHẦN III

## Câu 11

* 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. (2 điểm)

**Xét giá trị trung bình của tiền lương**

* Gọi μ0 là giá trị trung bình của tiền lương lúc mới ra trường.
* 
* Gọi μ1 là giá trị trung bình của tiền lương sau 10 năm làm việc.
* 
  + Vậy nên giá trị phương sai của tiền lương sau 10 năm làm việc sẽ lớn gấp khoảng lần so với trung bình phương sai lúc mới ra trường.

**Xét giá trị phương sai của tiền lương:**

* Gọi σ02là giá trị phương sai của tiền lương lúc mới ra trường, ta có:



* Gọi σ12là giá trị phương sai của tiền lương sau 10 năm làm việc, với X là mức lương sau 10 năm, ta có



* Vậy nên giá trị phương sai của tiền lương sau 10 năm làm việc sẽ lớn gấp khoảng lần so với trung bình phương sai lúc mới ra trường.

**Xét giá trị độ lệch chuẩn của tiền lương:**

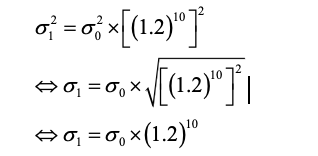
* Gọi σ0 là giá trị độ lệch chuẩn của tiền lương lúc mới ra trường.



* Gọi σ1là giá trị độ lệch chuẩn của tiền lương sau 10 năm làm việc.



***Ta có:***



* + Vậy nên giá trị độ lệch chuẩn của tiền lương sau 10 năm làm việc sẽ lớn gấp khoảng lần so với giá trị độ lệch chuẩn của tiền lương lúc mới ra trường

# TÀI LIỆU THAM KHẢO

Tiếng Anh

<https://stats.libretexts.org/Bookshelves/Probability_Theory/Probability_Mathematical_Statistics_and_Stochastic_Processes_(Siegrist)/08%3A_Set_Estimation/8.02%3A_Estimation_the_Normal_Model>

<https://www.scribbr.com/statistics/confidence-interval/>

<https://www.statlect.com/glossary/test-statistic>