**ĐẠI HỌC QUỐC GIA TPHCM**

**TRƯỜNG ĐẠI HỌC BÁCH KHOA**

🙦🟔🙤

**BÁO CÁO BÀI TẬP LỚN**

**MÔN XÁC SUẤT THỐNG KÊ**

**HỌC KỲ 201**

**GVHD : Nguyễn Kiều Dung**

**Lớp : L14 Nhóm : 8**

**Danh sách sinh viên thực hiện:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **STT** | **Lớp/ Tổ** | **Họ tên sinh viên** | **MSSV** | **Ngành học** | **Ký tên tham dự** |
| 1 | L14\_A | Nguyễn Thái An | 1910732 | Khoa Xây dựng |  |
| 2 | L14\_A | Đặng Quốc Đạt | 1911001 | Khoa Xây dựng |  |
| 3 | L14\_B | Nguyễn Như Đạt | 1913050 | Khoa Xây dựng |  |
| 4 | L14\_B | Nguyễn Hoàng Hữu Hùng | 1913607 | Khoa Xây dựng |  |
| 5 | L14\_B | Trần Mạnh Khải | 1913785 | Khoa Xây dựng |  |
| 6 | L14\_B | **Trần Tuấn Kiệt** | 1913886 | Khoa Xây dựng |  |
| 7 | L14\_B | Lê Minh Mẫn | 1914129 | Khoa Xây dựng |  |
| 8 | L14\_C | Phạm Hữu Thắng | 1915246 | Khoa Xây dựng |  |
| 9 | L14-C | Trần Trung | 1915702 | Khoa Xây Dựng |  |

TP Hồ Chí Minh, 14 tháng 12 năm 2020

**MỤC LỤC**

LỜI MỞ ĐẦU

PHẦN BÀI TẬP

[Câu 1](#Sol_1): Chọn 1 biến định lượng nào đó và thực hiện:

- Tìm các đặc trưng từ mẫu dữ liệu.

- Tìm các khoảng tin cậy cho giá trị trung bình và phương sai của tổng thể.

* LẬP BÀI TOÁN
* CƠ SỞ LÝ THUYẾT
* LỜI GIẢI TÍNH TAY
* GIẢI QUYẾT BÀI TOÁN TRÊN EXCEL
* BIỆN LUẬN

[Câu 2](#Sol_2): Chọn dữ liệu 2 biến định lượng (hoặc xử lý số liệu theo nhóm cho phù hợp) để lập bài toán kiểm định so sánh 2 trung bình tổng thể.

1. **LẬP BÀI TOÁN**

* CƠ SỞ LÝ THUYẾT
* LỜI GIẢI TÍNH TAY
* GIẢI QUYẾT BÀI TOÁN TRÊN EXCEl
* BIỆN LUẬN

[Câu 3](#Sol_3): Chọn dữ liệu cho k biến (k ≥3) (hoặc xử lý số liệu theo nhóm cho phù hợp) để lập bài toán so sánh về trung bình k tổng thể.

* LẬP BÀI TOÁN
* CƠ SỞ LÝ THUYẾT
* LỜI GIẢI TÍNH TAY
* GIẢI QUYẾT BÀI TOÁN TRÊN EXCEl
* BIỆN LUẬN

[Câu 4](#Sol_4): Chọn dữ liệu cho 2 biến (hoặc xử lý số liệu theo nhóm cho phù hợp) để lập bài toán kiểm định so sánh về sự phân tán. Trình bày các bước thực hiện và nhận xét kết quả.

1. LẬP BÀI TOÁN
2. CƠ SỞ LÝ THUYẾT
3. LỜI GIẢI TÍNH TAY
4. GIẢI QUYẾT BÀI TOÁN TRÊN EXCEL
5. BIỆN LUẬN

[Câu 5](#Sol_5): Chọn dữ liệu phù hợp để lập bài toán so sánh các tỉ lệ ( hay là bài toán kiểm định tính độc lập). Trình bày các bước thực hiện và nhận xét kết quả.

1. LẬP BÀI TOÁN
2. CƠ SỞ LÝ THUYẾT
3. LỜI GIẢI TÍNH TAY
4. GIẢI QUYẾT BÀI TOÁN TRÊN EXCEL
5. BIỆN LUẬN

[Câu 6](#Sol_6): Khảo sát hệ số tương quan giữa 2 biến định lượng cụ thể, dự đoán phương trình đường hồi quy tuyến tính giữa chúng ( có hình vẽ minh họa) và nhận xét về mối tương quan tuyến tính giữa 2 biến.

1. LẬP BÀI TOÁN
2. CƠ SỞ LÝ THUYẾT
3. LỜI GIẢI TÍNH TAY
4. GIẢI QUYẾT BÀI TOÁN TRÊN EXCEL
5. NHẬN XÉT

---------------------------------------------------

**TÀI LIỆU THAM KHẢO**

**ĐỀ TÀI**

**KHẢO SÁT Ý KIẾN SINH VIÊN VỀ SỰ THAY ĐỔI MÔN HỌC XÁC SUẤT THỐNG KÊ TRONG HỌC KỲ 201**

Để đánh giá sự hiệu quả của môn học, đề tài này được thực hiện nhằm mục đích khảo sát ý kiến của sinh viên về sự thay đổi nội dung môn học. Bài khảo sát dưới đây gồm 2 phần:

+ Phần 1: Ý kiến của sinh viên về sự thay đổi nội dung thi giữa kỳ

+ Phần 2: Sự chuẩn bị của sinh viên trước những thay đổi của bài tập lớn

**BÀI TẬP LỚN SỐ 01**

**[Câu 1](#De_1): Chọn một biến định lượng nào đó và thực hiện:**

**Tìm các đặc trưng từ mẫu dữ liệu**

**Tìm các khoảng tin cậy cho giá trị trung bình và phương sai của tổng thể**

**PHẦN 1: LẬP BÀI TOÁN**

Các bảng số liệu bên dưới cho ta thông tin về điểm đánh giá đối với đề thi giữa kỳ (đơn vị: điểm) của các sinh viên Bách Khoa thông qua các tiêu chí ( tính ứng dụng, tính sáng tạo, độ khó)

Bảng 1: Số liệu khảo sát điểm đánh giá về độ ứng dụng của đề thi giữa kỳ

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 8 | 9 | 8 | 6 | 5 | 8 | 7 | 10 | 4 | 8 | 8 | 5 |
| 7 | 6 | 6 | 7 | 7 | 7 | 8 | 7 | 8 | 9 | 7 | 4 |
| 7 | 4 | 9 | 7 | 5 | 6 | 5 | 6 | 8 | 8 | 7 | 7 |
| 7 | 4 | 7 | 10 | 7 | 1 | 5 | 7 | 9 | 8 | 8 | 2 |
| 9 | 7 | 9 | 9 | 9 | 4 | 8 | 6 | 8 | 7 | 10 | 1 |
| 7 | 8 | 5 | 8 | 5 | 7 | 9 | 8 | 10 | 7 | 3 | 8 |
| 8 | 9 | 6 | 9 | 7 | 7 | 7 | 7 | 10 | 9 | 5 | 4 |
| 3 | 4 | 7 | 7 | 8 | 10 | 10 | 7 | 7 | 9 | 5 | 8 |
| 7 | 5 | 9 | 8 | 5 | 8 | 7 | 7 | 4 | 7 | 8 | 5 |
| 5 | 7 | 5 | 7 | 7 | 5 | 4 | 7 | 8 | 7 | 4 | 10 |
| 7 | 9 | 8 | 5 | 10 | 10 | 8 | 6 | 8 | 9 | 8 | 9 |
| 6 | 7 | 5 | 7 | 8 | 5 | 8 | 4 | 9 | 5 | 7 | 7 |
| 10 | 6 | 9 | 6 | 9 | 5 | 10 | 7 | 10 | 7 | 7 | 9 |
| 8 | 8 | 7 | 7 | 6 | 6 | 8 | 8 | 8 | 6 | 1 | 5 |
| 6 | 7 | 6 | 9 |  |  |  |  |  |  |  |  |

Bảng 2: Số liệu khảo sát điểm đánh giá về độ sáng tạo của đề thi giữa kỳ

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 7 | 9 | 7 | 7 | 5 | 7 | 7 | 10 | 4 | 1 | 6 | 5 |
| 5 | 8 | 7 | 7 | 7 | 7 | 8 | 7 | 9 | 8 | 6 | 5 |
| 8 | 5 | 9 | 7 | 5 | 7 | 5 | 4 | 7 | 6 | 7 | 5 |
| 7 | 7 | 7 | 7 | 6 | 10 | 6 | 7 | 5 | 8 | 9 | 2 |
| 6 | 7 | 8 | 7 | 8 | 4 | 8 | 8 | 6 | 9 | 10 | 1 |
| 9 | 8 | 7 | 7 | 5 | 5 | 8 | 7 | 9 | 8 | 3 | 7 |
| 7 | 9 | 7 | 7 | 7 | 7 | 5 | 8 | 10 | 10 | 5 | 6 |
| 3 | 4 | 8 | 7 | 8 | 10 | 1 | 7 | 9 | 8 | 5 | 7 |
| 6 | 5 | 9 | 7 | 5 | 7 | 7 | 6 | 3 | 9 | 7 | 9 |
| 4 | 6 | 6 | 7 | 8 | 5 | 3 | 9 | 10 | 7 | 4 | 9 |
| 7 | 9 | 9 | 5 | 5 | 4 | 8 | 8 | 7 | 9 | 9 | 9 |
| 8 | 7 | 5 | 7 | 4 | 5 | 7 | 4 | 9 | 5 | 6 | 8 |
| 10 | 5 | 10 | 6 | 9 | 7 | 10 | 7 | 10 | 7 | 8 | 8 |
| 8 | 7 | 6 | 8 | 6 | 7 | 7 | 8 | 7 | 6 | 1 | 5 |
| 6 | 7 | 6 | 9 |  |  |  |  |  |  |  |  |

Bảng 3: Số liệu khảo sát điểm đánh giá về độ khó của đề thi giữa kỳ

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 9 | 8 | 8 | 8 | 5 | 7 | 7 | 10 | 10 | 9 | 7 | 5 |
| 5 | 7 | 6 | 5 | 8 | 10 | 8 | 10 | 8 | 7 | 7 | 10 |
| 10 | 7 | 8 | 8 | 5 | 7 | 10 | 9 | 9 | 9 | 8 | 8 |
| 8 | 8 | 7 | 8 | 10 | 8 | 6 | 6 | 8 | 8 | 7 | 3 |
| 5 | 9 | 8 | 8 | 8 | 10 | 8 | 10 | 6 | 1 | 10 | 10 |
| 9 | 6 | 8 | 1 | 6 | 9 | 8 | 8 | 1 | 8 | 4 | 9 |
| 9 | 4 | 8 | 8 | 6 | 7 | 8 | 7 | 10 | 9 | 5 | 8 |
| 8 | 10 | 5 | 6 | 8 | 10 | 10 | 9 | 10 | 7 | 5 | 6 |
| 10 | 6 | 8 | 10 | 5 | 8 | 7 | 9 | 9 | 9 | 6 | 9 |
| 6 | 7 | 7 | 8 | 7 | 10 | 5 | 7 | 9 | 10 | 5 | 8 |
| 10 | 10 | 8 | 5 | 10 | 4 | 6 | 9 | 8 | 9 | 10 | 10 |
| 7 | 8 | 5 | 8 | 10 | 5 | 7 | 10 | 10 | 6 | 8 | 9 |
| 8 | 10 | 10 | 4 | 9 | 8 | 10 | 7 | 10 | 8 | 8 | 8 |
| 9 | 9 | 8 | 10 | 7 | 7 | 4 | 9 | 7 | 9 | 1 | 7 |
| 5 | 9 | 8 | 6 |  |  |  |  |  |  |  |  |

Từ các bảng số liệu trên, hãy thực hiện các yêu cầu sau:

* Tìm các đặc trưng mẫu.
* Tìm các khoảng tin cậy cho giá trị trung bình và phương sai của tổng thể.

**PHẦN 2: CƠ SỞ LÝ THUYẾT**

**Nội dung 1:** Các đặc trưng mẫu

1. **Tỷ lệ mẫu:**

Cho mẫu định tính kích thước n, trong đó có m phần tử có tính chất A.

Khi đó, ta gọi f = m/n là tỷ lệ mẫu.

Trường hợp này mẫu được cho dưới dạng:

+ Kích thước mẫu: n.

+ Số phần tử có tính chất A: m.

1. **Trung bình mẫu:**

Cho X là một biến ngẫu nhiên.

Giả sử X có n quan sát . Trung bình mẫu được định nghĩa bởi:

Trung bình mẫu được dùng để xấp xỉ trung bình tổng thể. Trong đó: trung bình của tổng thể, được gọi tắt là kỳ vọng và được ký hiệu là hay , là trung bình của tất cả giá trị có thể có của X trong quần thể.

1. **Phương sai mẫu – độ lệch mẫu:**

Cho X là một biến ngẫu nhiên.

Giả sử X có n quan sát . Phương sai mẫu là trung bình của bình phương độ lệch mẫu so với kỳ vọng:

Phương sai mẫu hiệu chỉnh\* s2 được định nghĩa theo công thức:

Nghĩa là:

Khi đó: và s lần lượt được gọi là độ lệch mẫu và độ lệch mẫu hiệu chỉnh.

1. **Các đặc trưng khác:**

Ngoài các đặc trưng chính, ta còn có một số đặc trưng khác như:

+ Yếu vị (mode): là giá trị của phần tử có số lần xuất hiện lớn nhất trong mẫu.

+ Hệ số biến thiên (Coefficient of variation – CV): Hệ số biến thiên đo lường mức độ biến động tương đối của mẫu dữ liệu, được dùng khi người ta muốn so sánh mức độ biến động của các mẫu không cùng đơn vị đo.

+ Sai số chuẩn: là giá trị đại diện cho độ lệch chuẩn của giá trị trung bình trong tập dữ liệu. Nó phục vụ như một thước đo biến động cho các biến ngẫu nhiên hay đo lường độ phân tán. Độ phân tán càng nhỏ, dữ liệu càng chính xác.

+ Trung vị mẫu:

Giả sử X có n quan sát . Xếp các quan sát này theo thứ tự tăng dần.

Trung vị mẫu (Median) là giá trị nằm chính giữa dãy số này và chia nó thành 2 phần bằng nhau. Cụ thể, nếu mẫu có dạng x1<…<x2k+1 thì trung vị mẫu là xk+1; nếu mẫu có dạng x1<…<x2k thì trung vị mẫu là trung bình cộng (xk + xk+1)/2

….

**Nội dung 2:** **Khoảng tin cậy**

Cho tham số của tổng thể và là các quan sát ngẫu nhiên. Ta gọi khoảng (c,d) là khoảng ước lượng (hay khoảng tin cậy) của tham số với độ tin cậy nếu

P

Có thể nói, độ tin cậy cho khoảng ước lượng của tham số chính là xác suất để ta đúng khi ước lượng tham số bằng khoảng (c,d). Ngược lại, xác suất mà ta cho phép sai khi ước lượng được gọi là mức ý nghĩa. Kí hiệu là . Ta có

Nếu là một ước lượng không chệch của thì khoảng ước lượng củacó dạng () được gọi là khoảng ước lượng đối xứng, được gọi là độ chính xác của ước lượng.

Giả sử ( là khoảng ước lượng đối xứng của với độ tin cậy thì

P

**BẢNG TÓM TẮT BÀI TOÁN ƯỚC LƯỢNG**

|  |  |  |
| --- | --- | --- |
| **Dạng ước lượng** | **Độ chính xác** | **Khoảng ước lượng:** |
| 1. Tỷ lệ: |  |  |
| 1. Trung bình: | | |
| * Mẫu lớn (: |  |  |
| * Tuân theo đã biết : |  |  |
| * Tuân theo chưa biết : |  |  |
| 1. Phương sai: |  |  |

**PHẦN 3: LỜI GIẢI TÍNH TAY**

**Bảng 1:**

**Nội dung 1:** Các đặc trưng mẫu:

|  |  |
| --- | --- |
| Kích thước mẫu | n = 172 |
| Trung bình mẫu |  |
| Phương sai mẫu |  |
| Độ lệch mẫu | 1,8843 |
| Phương sai mẫu hiệu chỉnh |  |
| Độ lệch mẫu hiệu chỉnh |  |
| Sai số chuẩn |  |
| Trung vị | med = 7 |
| Yếu vị | mod = 7 |
| Giá trị lớn nhất | max = 10 |
| Giá trị nhỏ nhất | min = 1 |
| Khoảng biến thiên | max – min = 9 |
| Tổng | sum = 1160 |

**Nội dung 2:** Khoảng tin cậy cho giá trị trung bình của tổng thể:

Ta có: n = 172; s = 1,8898; = 6,7441

Độ tin cậy: 1 - α = 0,95 Zα = 1,9599

Độ chính xác của ước lượng:

Khoảng tin cậy cho điểm đánh giá trung bình về độ ứng dụng của đề thi giữa kỳ:

**Nội dung 3:** Khoảng tin cậy cho phương sai của tổng thể:

Ta có: n =172; s2 = 3,5716

016

37

Khoảng tin cậy cho phương sai điểm đánh giá trung bình vể độ ứng dụng của đề thi giữa kỳ:

**Bảng 2:**

**Nội dung 1:** Các đặc trưng mẫu:

|  |  |
| --- | --- |
| Kích thước mẫu | n = 172 |
| Trung bình mẫu | 7,6395 |
| Phương sai mẫu | 3,8933 |
| Độ lệch mẫu | 1,9731 |
| Phương sai mẫu hiệu chỉnh | 3,9160 |
| Độ lệch mẫu hiệu chỉnh | 1,9789 |
| Sai số chuẩn | 1509 |
| Trung vị | med = 8 |
| Yếu vị | mod = 8 |
| Giá trị lớn nhất | max = 10 |
| Giá trị nhỏ nhất | min = 1 |
| Khoảng biến thiên | max – min =9 |
| Tổng | sum = 1314 |

**Nội dung 2:** Khoảng tin cậy cho giá trị trung bình của tổng thể:

Ta có: n =172 ; s =1,979 ; = 7,6395

Độ tin cậy: 1 - α = 0,95 Zα = 1,9599

Độ chính xác của ước lượng:

Khoảng tin cậy cho điểm đánh giá trung bình về độ sáng tạo của đề thi giữa kỳ:

**Nội dung 3:** Khoảng tin cậy cho phương sai của tổng thể:

Ta có: n = 172; s2 = 3,9160

016

37

Khoảng tin cậy cho phương sai điểm đánh giá trung bình về độ sáng tạo của đề thi giữa kỳ:

**Bảng 3:**

**Nội dung 1:** Các đặc trưng mẫu:

|  |  |
| --- | --- |
| Kích thước mẫu | n = 172 |
| Trung bình mẫu |  |
| Phương sai mẫu |  |
| Độ lệch mẫu |  |
| Phương sai mẫu hiệu chỉnh |  |
| Độ lệch mẫu hiệu chỉnh |  |
| Sai số chuẩn |  |
| Trung vị | med = 7 |
| Yếu vị | mod = 7 |
| Giá trị lớn nhất | max = 10 |
| Giá trị nhỏ nhất | min = 1 |
| Khoảng biến thiên | max – min = 9 |
| Tổng | sum = 1168 |

**Nội dung 2:** Khoảng tin cậy cho giá trị trung bình của tổng thể:

Ta có: n = 172; s = 1,9262 ; = 6,7906

Độ tin cậy: 1 - α = 0,95 Zα = 1,9599

Độ chính xác của ước lượng:

Khoảng tin cậy cho điểm đánh giá trung bình về độ khó của đề thi giữa kỳ:

**Nội dung 3:** Khoảng tin cậy cho phương sai của tổng thể:

Ta có: n = 172; s2 = 3,7103

016

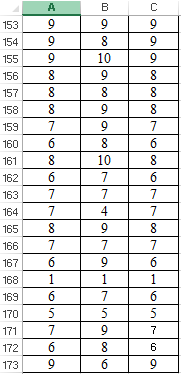
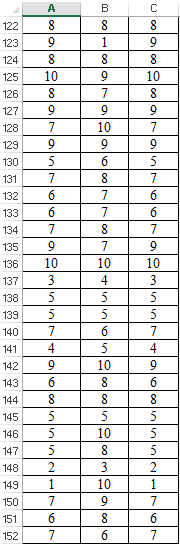
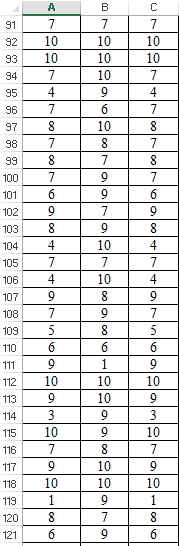
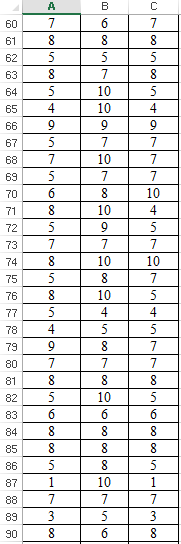
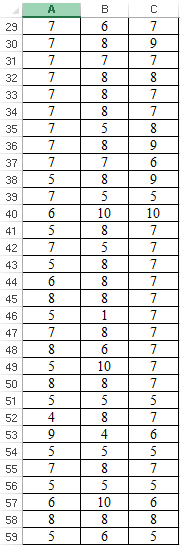
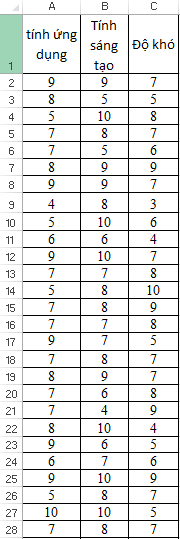
37

Khoảng tin cậy cho phương sai điểm đánh giá trung bình về độ khó của đề thi giữa kỳ

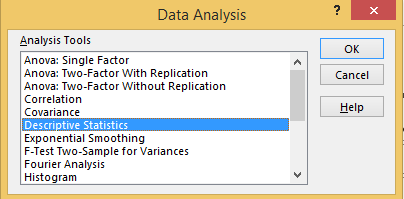
**PHẦN 4: GIẢI QUYẾT BÀI TOÁN TRÊN EXCEL**

**Nội dung 1:** Các đặc trưng mẫu:

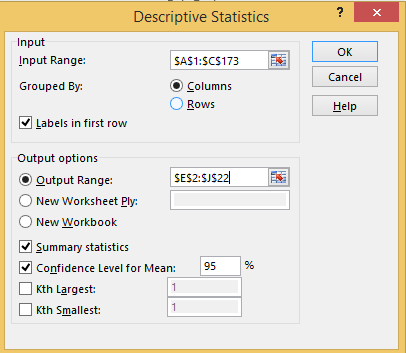
**Bước 1:** Nhập dữ liệu: ta nhập các cột dữ liệu theo hàng dọc.



**Bước 2:** Sử dụng công cụ **Data Analysis/Descriptive Statistics**



**Bước 3:** Chọn các mục như hình:



Input: quét địa chỉ chứa dữ liệu.

Chọn Group by: Columns

Chọn: Labels in frist row

Output Range: chọn vị trí cần xuất kết quả

Chọn: Summary statistic

Confidence Level for Mean: nhập độ tin cậy 95%

Kết quả thu được:



Dịch sang tiếng việt:



**Nội dung 2:** Khoảng tin cậy cho giá trị trung bình của tổng thể:

**Bảng 1:**

Dựa trên bảng Descriptive, ta xác định được:

n = 172; s = 1,8898; = 6,7441

Ngoài ra, ta có thể xác định được độ chính xác: , tuy nhiên kết quả trong bảng descriptive cho ta độ chính tính theo tiêu chuẩn student với (dành cho bài toán n <30). Do đó ta cần xác định lại độ chính xác theo công thức:

với:



Ta tiếp tục tính độ chính xác:

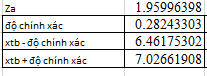


Từ đó, ta xác định được khoảng tin cậy cho điểm đánh giá trung bình về độ ứng dụng của đề thi giữa kỳ:





Kết quả thu được:



Khoảng tin cậy cho điểm đánh giá trung bình về độ ứng dụng của đề thi giữa kỳ:

(− ε; + ε) ↔ (6,7441-

**Bảng 2:**

Dựa trên bảng Descriptive, ta xác định được:

n =172; s = 1,9789; = 7,6395

Ngoài ra, ta có thể xác định được độ chính xác: , tuy nhiên kết quả trong bảng descriptive cho ta độ chính tính theo tiêu chuẩn student với (dành cho bài toán n <30). Do đó ta cần xác định lại độ chính xác theo công thức:

với:



Ta tiếp tục tính độ chính xác:

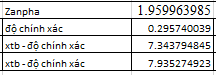


Từ đó, ta xác định được khoảng tin cậy cho điểm đánh giá trung bình về độ sáng tạo của đề thi giữa kỳ:





Kết quả thu được:



Khoảng tin cậy cho điểm đánh giá trung bình về độ sáng tạo của đề thi giữa kỳ:

− ε; + ε) ↔ (7,6395-

**Bảng 3:**

Dựa trên bảng Descriptive, ta xác định được:

n = 172; s = 1,9262; = 6,7906

Ngoài ra, ta có thể xác định được độ chính xác: , tuy nhiên kết quả trong bảng descriptive cho ta độ chính tính theo tiêu chuẩn student với (dành cho bài toán n <30). Do đó ta cần xác định lại độ chính xác theo công thức:

với:



Ta tiếp tục tính độ chính xác:

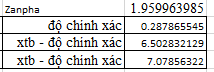


Từ đó, ta xác định được khoảng tin cậy cho điểm đánh giá trung bình về độ khó của đề thi giữa kỳ theo công thức:





Kết quả thu được:



Khoảng tin cậy cho điểm đánh giá trung bình về độ khó của đề thi giữa kỳ:

(− ε; + ε) (6,7906-

**Nội dung 3:** Khoảng tin cậy cho phương sai của tổng thể:

**Bảng 1:**

Dựa trên bảng Descriptive, ta xác định được:

n = 172; s2 =

Ta cần tính:

được tính bằng hàm Chiinv

được tính bằng hàm Chiinv



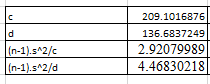


Từ đó, ta xác định được khoảng tin cậy cho phương sai điểm đánh giá trung bình về độ ứng dụng của đề thi giữa kỳ theo công thức:





Kết quả thu được:



Khoảng tin cậy cho phương sai điểm đánh giá trung bình về độ ứng dụng của đề thi giữa kỳ:

2,9207; 4,4683)

**Bảng 2:**

Dựa trên bảng Descriptive, ta xác định được:

n = 172; s2 = 3,916

Ta cần tính:

được tính bằng hàm Chiinv

được tính bằng hàm Chiinv



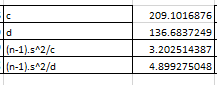


Từ đó, ta xác định được khoảng tin cậy cho phương sai điểm về độ sáng tạo của đề thi giữa kỳ:





Kết quả thu được:



Khoảng tin cậy cho phương sai điểm đánh giá trung bình về độ sáng tạo của đề thi giữa kỳ:

3,2025; 4,8992)

**Bảng 3:**

Dựa trên bảng Descriptive, ta xác định được:

n = 172; s2 = 3,7103

Ta cần tính:

được tính bằng hàm Chiinv

được tính bằng hàm Chiinv



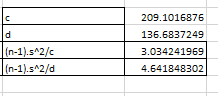


Từ đó, ta xác định được khoảng tin cậy cho phương sai điểm đánh giá trung bình về độ khó của đề thi giữa kỳ:





Kết quả thu được:



Khoảng tin cậy cho phương sai điểm đánh giá trung bình về mức độ khó của đề thi giữa kỳ:

3,0342; 4,6418)

**NHẬN XÉT:**

Với độ tin cậy 95%, ta thu được các khoảng tin cậy:

+Khoảng tin cậy cho tính ứng dụng trung bình của đề thi giữa kì là:

+Khoảng tin cậy cho tính sáng tạo trung bình của đề thi giữa kì là:

+Khoảng tin cậy cho độ khó trung bình của đề thi giữa kì là:)

+ Khoảng tin cậy cho phương sai điểm đánh giá trung bình về mức độ ứng dụngcủa đề thi giữa kỳ:2,9207; 4,4683)

+ Khoảng tin cậy cho phương sai điểm đánh giá trung bình về mức độ sáng tạo của đề thi giữa kỳ: 3,2025; 4,8992)

+ Khoảng tin cậy cho phương sai điểm đánh giá trung bình về mức độ khó của đề thi giữa kỳ: 3,0342; 4,6418)

Ngoài ra, ta còn dễ dàng nhận thấy các đặc trưng khác của mẫu số liệu. Các sinh viên đánh giá đề thi giữa kì của năm nay 6,7 trên 10 điểm về tính ứng dụng của môn học vào thực tế,7,6 điểm trên 10 điểm về tính sáng tạo của đề thi giữa kì, 6,7 trên 10 điểm về độ khó, vậy nên đề thi như vậy là thích hợp, cần được phát huy, và khoảng tin cậy của ba số liệu này chênh lệch không đáng kể, suy ra phần lớn sinh viên có năng lực học khá tốt chiếm tỷ lệ khá cao.

|  |
| --- |
| **[Câu 2](#De_2): Chọn dữ liệu 2 biến định lượng (hoặc xử lý số liệu theo nhóm cho phù hợp) để lập bài toán kiểm định so sánh 2 trung bình tổng thể.** |

**PHẦN 1: LẬP BÀI TOÁN**

Để đánh giá sự ảnh hưởng của các khoa lên việc áp dụng kiến thức đã dạy trong lớp vào đề thi giữa kì của sinh viên, ta thực hiện khảo sát về phần trăm kiến thức đã dạy trong lớp áp dụng vào đề thi giữa kì của sinh viên 2 khoa Điện-Điện tử và Xây Dựng :

**Nhóm 1:** Sinh viên khoa Điện-Điện tử

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 60 | 90 | 50 | 50 | 90 | 5 | 70 | 75 | 70 | 40 | 60 | 80 |
| 90 | 60 | 60 | 0 | 70 | 50 | 85 | 50 | 90 | 100 | 50 | 50 |
| 100 | 70 | 5 | 80 | 40 | 4 | 35 | 50 | 70 | 80 | 50 | 80 |
| 80 | 75 | 10 | 80 | 30 | 60 | 90 | 70 | 70 | 40 | 80 | 60 |
| 40 | 10 | 80 | 80 | 30 | 100 | 20 | 80 | 70 | 40 | 60 | 60 |
| 60 | 100 | 50 | 80 | 100 | 70 | 30 | 70 | 60 | 30 | 60 | 1 |
| 90 | 80 |  |  |  |  |  |  |  |  |  |  |

**Nhóm 2:** Sinh viên khoa Xây Dựng

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 90 | 90 | 30 | 70 | 100 | 30 | 0 | 5 | 90 | 90 | 30 | 70 |
| 10 | 60 | 60 | 40 | 50 | 50 | 70 | 80 | 80 | 80 | 80 | 30 |
| 50 | 90 | 60 | 80 | 80 | 50 | 70 | 80 | 80 | 60 | 40 | 65 |
| 60 | 50 | 10 | 70 | 60 | 60 | 80 | 75 | 70 | 70 | 65 | 75 |
| 80 |  |  |  |  |  |  |  |  |  |  |  |

Với mức ý nghĩa 5% hãy so sánh phần trăm kiến thức đã dạy trong lớp áp dụng vào đề thi giữa kì của sinh viên 2 khoa Điện-Điện tử và Xây Dựng. Giả thiết số liệu được khảo sát độc lập và phần trăm kiến thức đã dạy trong lớp áp dụng vào đề thi giữa kì thu được ở 2 nhóm tuân theo quy luật phân phối chuẩn.

**PHẦN 2: CƠ SỞ LÝ THUYẾT**

**Dạng bài:**  Kiểm định trung bình.

**Phương pháp giải**: So sánh giá trị trung bình với phương sai khác nhau.

**Lý thuyết so sánh giá trị trung bình với phương khác nhau**:

Với hai mẫu lớn (N>30) độc lập và có phương sai khác nhau (hai mẫu phân biệt), ta có thể áp dụng trắc nghiệm t dị phương sai (betero – scedastic – test) để so sánh giá trị trung bình của hai mẫu ấy.

* Trắc nghiệm bên phải:

**Giả thiết H0:** 1 = 2

**Giả thiết H1:** 1 > 2

**Miền bác bỏ bên phải:** Wα= (

* Trắc nghiệm bên trái:

**Giả thiết H0:** 1 = 2

**Giả thiết H1:** 1 < 2

**Miền bác bỏ bên trái**: Wα= (- ∞; - )

* Trắc nghiệm hai bên:

**Giả thiết H0:** 1 = 2

**Giả thiết H1:** 1 2

**Miền bác bỏ hai bên:** Wα= ( - ∞;- ) (

**Tiêu chuẩn kiểm định:**

**Biện luận:**

Nếu Z Wα ⇒ Bác bỏ giả thiết H0, chấp nhận giả thiết H1.

Nếu Z Wα ⇒ chưa bác bỏ giả thiết H0.

**PHẦN 3: LỜI GIẢI TÍNH TAY**

Gọi a1, a2 lần lượt là phần trăm trung bình kiến thức đã dạy trên lớp áp dụng vào đề thi giữa kỳ của 2 khoa Điện-Điện tử và Xây Dựng

Ta có: ; n1 = 74; n2 = 49 sx2=688,5474; sy2=596,7049

**Giả thiết H0:** Phần trăm trung bình kiến thức đã dạy trên lớp áp dụng vào đề thi giữa kỳ của 2 khoa Điện-Điện tử và Xây Dựng bằng nhau.

**Giả thiết H1:** Phần trăm trung bình kiến thức đã dạy trên lớp áp dụng vào đề thi giữa kỳ của sinh viên khoa Điện-Điện tử lớn hơn khoa Xây Dựng.

Ta có

**Miền bác bỏ:** Wα = ( Z2α;) = (1.64;)

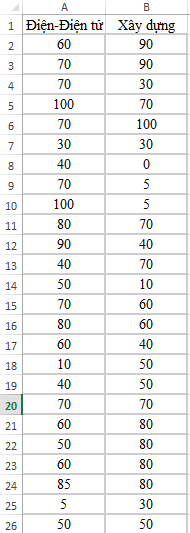
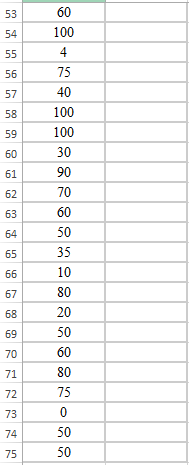
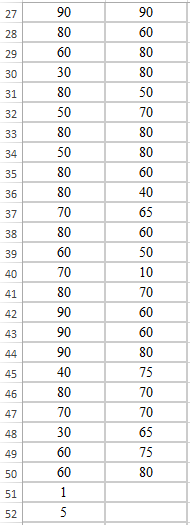
**Tiêu chuẩn kiểm định:**

Vì Zqs ∉ Wα=> chấp nhận Ho

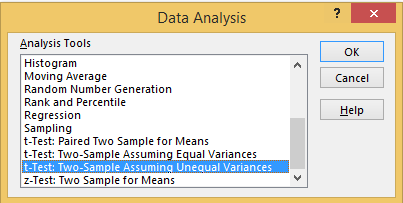
**Kết luận:** Vậy phần trăm kiến thức trung bình đã dạy trên lớp áp dụng vào đề thi giữa kỳ của sinh viên khoa Điện-Điện tử và Xây Dựng như nhau (≈60%). Điều đó cho thấy dù sinh viên học khoa Điện-Điện tử hay Xây Dựng thì phần trăm kiến thức trung bình đã dạy trên lớp áp dụng vào đề thi giữa kỳ như nhau.

**PHẦN 4: GIẢI QUYẾT BÀI TOÁN TRÊN EXCEL**

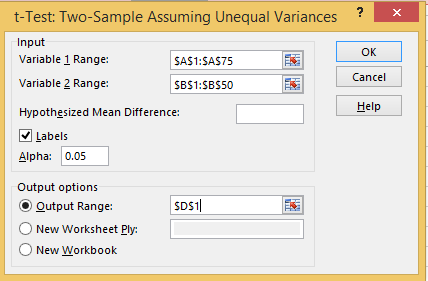
**Bước 1:** Nhập dữ liệu:

**Bước 2:** Sử dụng công cụ Data Analysis/t-test: Two-sample Assuming Unequal Variances

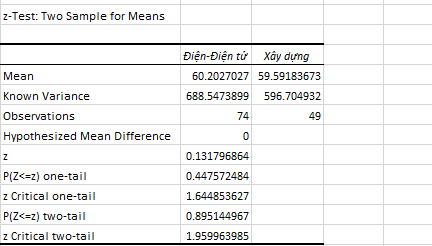


**Bước 3:** Chọn các mục như hình:



* Varible 1 Range: Quét địa chỉ chứa dữ liệu 1.
* Varible 2 Range: Quét địa chỉ chứa dữ liệu 2.
* Chọn Labels
* Alpha: nhập 0.05 (mức ý nghĩa 5%)
* Output Range: chọn Vị trí xuất kết quả.

Kết quả:

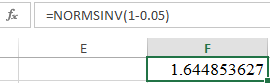


**Bước 4:** Biện luận bài toán

Đặt giả thuyết :

**Giả thiết H0:** Phần trăm kiến thức trung bình đã dạy trên lớp áp dụng vào đề thi giữa kỳ của sinh viên khoa Điện-Điện tử và Xây Dựng bằng nhau

**Giả thiết H1:** Phần trăm kiến thức trung bình đã dạy trên lớp áp dụng vào đề thi giữa kỳ của sinh viên khoa Điện-Điện tử lớn hơn Xây Dựng

Ta có : Tiêu chuẩn kiểm định T (t Stat) = 0,1318 và giá trị ngưỡng T critical one tail = 1,6591. Tuy nhiên giá trị ngưỡng ta tính được từ công cụ t-test chỉ đúng đối với trường hợp (n < 30), kết quả cho ta phân vị của phân phối student. Do đó, ta cần tính lại ngưỡng Z2 ở bài toán kiểm định 1 phía bên phải này. Ta sẽ sử dụng hàm Norm.s.inv để tìm Z2.

Kết quả:

**Miền bác bỏ:** Wα = ( Z2α;) = (1.6449;)

**Tiêu chuẩn kiểm định:** Zqs=0,1318

Vì Zqs=0,1318 ∉ Wα=> Chấp nhận H0

**Kết luận:** Vậy phần trăm kiến thức trung bình đã dạy trên lớp áp dụng vào đề thi giữa kỳ của sinh viên khoa Điện-Điện tử và Xây Dựng như nhau (≈60%). Điều đó cho thấy dù sinh viên học khoa Điện-Điện tử hay Xây Dựng thì phần trăm kiến thức trung bình đã dạy trên lớp áp dụng vào đề thi giữa kỳ như nhau.

**[Câu 3](#De_3): Chọn dữ liệu cho k biến (k ) (hoặc xử lý số liệu theo nhóm cho phù hợp) để lập bài toán so sánh về phân tích phương sai.**

**PHẦN 1: LẬP BÀI TOÁN**

Để đánh giá mức độ ứng phó của y tế các nước với đại dịch Covid-19, ta lấy dữ liệu tỉ lệ ca tử vong (đơn vị: %) trên tổng số ca mắc của các quốc gia theo từng châu lục, số liệu thu được đến ngày 22/5/2021 như sau:

**Nhóm 1:** *Tỉ lệ* ca tử vong trên tổng số ca bệnh do nCoV của các nước châu Phi

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 2.69 | 2.22 | 1.26 | 1.52 | 1.23 | 0.13 | 1.59 | 0.88 | 1.37 | 3.51 |
| 3.77 | 1.31 | 0.63 | 2.52 | 1.32 | 5.81 | 1.34 | 0.36 | 3.62 | 1.51 |
| 0.61 | 2.93 | 0.84 | 0.69 | 1.81 | 1.81 | 3.01 | 3.97 | 1.7 | 1.94 |
| 3.36 | 3.6 | 2.4 | 1.31 | 1.76 | 1.18 | 1.45 | 3.58 | 1.25 | 1.31 |
| 0 | 1.54 | 2.75 | 0.36 | 1.92 | 5.25 | 3.42 | 1.08 | 7.36 | 4.13 |
| 0.94 | 3.65 | 0.8 | 1.36 | 4.1 |  |  |  |  |  |

**Nhóm 2:** *Tỉ lệ* ca tử vong trên tổng số ca bệnh do nCoV các nước châu Á

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 4.26 | 1.98 | 1.46 | 0.38 | 1.57 | 0.07 | 1.28 | 0.68 | 5.09 | 1.37 |
| 1.78 | 1.12 | 2.78 | 2.78 | 1.39 | 0.76 | 1.7 | 1.28 | 0.8 | 0.58 |
| 1.71 | 0.11 | 1.43 | 0 | 0.44 | 0.23 | 0.47 | 2.25 | 1.22 | 0 |
| 1.07 | 2.25 | 1.13 | 1.69 | 0.25 | 1.65 | 0.05 | 1.42 | 0.73 | 7.19 |
| 0.44 | 0.68 | 0.6 | 0.2 | 0.89 | 0 | 0.3 | 0.69 | 0.82 | 19.61 |

**Nhóm 3:** *Tỉ lệ* ca tử vong trên tổng số ca bệnh do nCoV của các nước châu Âu

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1.85 | 0.94 | 1.64 | 0.72 | 2.37 | 4.48 | 4.2 | 2.23 | 0.49 | 1.81 |
| 0.92 | 0.97 | 0 | 1.02 | 1.81 | 2.39 | 0 | 3.02 | 0 | 3.68 |
| 0.44 | 1.94 | 0 | 2.99 | 0 | 2.1 | 1.78 | 1.93 | 1.54 | 1.16 |
| 1.37 | 2.38 | 1.28 | 1.58 | 1.08 | 3.42 | 0.64 | 2.54 | 2.01 | 2.78 |
| 2.35 | 1.77 | 0.95 | 3.16 | 1.73 | 2.19 | 1.36 | 1.57 | 2.29 | 2.86 |
| 0 |  |  |  |  |  |  |  |  |  |

**Nhóm 4:** *Tỉ lệ* ca tử vong trên tổng số ca bệnh do nCoV của các nước bắc Mỹ

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 0 | 3.34 | 0 | 1.95 | 1.18 | 2.53 | 0 | 0 | 1.85 | 0 |
| 1.26 | 0.65 | 0 | 0 | 1.28 | 3.06 | 0 | 0.62 | 3.24 | 2.03 |
| 2.65 | 1.9 | 9.25 | 0 | 2.57 | .69 | 0 | 1.56 | 0.61 | 0 |
| 1.88 | 0 | 1.78 |  |  |  |  |  |  |  |

Với mức ý nghĩa 5% hãy so sánh tỉ lệ tử vong trên tổng số ca bệnh của 4 nhóm trên. Cho nhận xét. Giả sử tỉ lệ này của 3 nhóm tuân theo phân phối chuẩn và có cùng phương sai.

**PHẦN 2: CƠ SỞ LÝ THUYẾT**

**Dạng bài:** Kiểm định trung bình

**Phương pháp giải:** Phân tích phương sai 1 yếu tố

**Lý thuyết phân tích phương sai 1 yếu tố**:

Phép phân tích phương sai là so sánh trung bình của 2 hay nhiều nhóm dựa trên các giá trị trung bình của các mẫu quan sát từ các nhóm này, và thông qua kiểm định giả thiết để kết luận về sự bằng nhau của các trung bình tổng thể này.

**Phương pháp phân tích phương sai 1 yếu tố:**

Phép phân tích phương sai được dùng trong các trắc nghiệm để so sánh các giá trị trung bình của hai hay nhiều mẫu được lấy từ các phân số.

Đây có thể được xem như phần mở rộng các trắc nghiệm t hay z (so sánh hai giá trị trung bình).

Mục đích của sự phân tích phương sai một yếu tố là đánh giá sự ảnh hưởng của 1 yếu tố (nhân tạo hay tự nhiên) nào đó trên các giá trị quan sát, Yi(i=0,1,2,…,k)

**Điều kiện:**

Các giá trị được chọn 1 cách ngẫu nhiên và độc lập.

Mẫu N lớn hơn 30 hoặc tuân theo phân phối chuẩn.

Phương sai của các nhóm bằng nhau.

**Mô hình:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Yếu tố thí nghiệm | | | |  |
| 1 | 2 | ........ | k |
|  | Y11 | Y21 | ........ | Yk1 |  |
| Y12 | Y22 | ........ | Yk2 |
| .......... | ......... | ........ | ........... |
| Y1N | Y2N | ........ | YkN |
| Tổng cộng  trung bình | T1  \_ | T2  \_ | ........  ........ | Tk  \_ | T  \_ |

**Bảng Anova:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Nguồn sai số | Bậc sai số | Tổng số bình phương | Bình phương trung bình | Gía trị thống kê |
| Yếu tố |  |  |  |  |
| Sai số |  |  |  |  |
| Tổng cộng |  |  |  |  |

Đặt giả thiết:

H0: μ1 = μ2 = ...... = μn ⭤ “Các giá trị trung bình bằng nhau”

H1: μi ≠ μj ⭤ “Ít nhất có hai giá trị trung bình khác nhau”

Tiêu chuẩn kiểm định: F = MSF/MSE

Biện luận: Nếu F < → Chấp nhận giả thiết H0.

**PHẦN 3: LỜI GIẢI TÍNH TAY**

**Giả thiết H0:** Tỉ lệ ca tử vong trên tổng số ca bệnh do nCov-19 của 4 châu lục là như nhau.

**Giả thiết H1:** Có ít nhất 2 châu lục mà tỉ lệ ca tử vong trên tổng số ca bệnh là khác nhau.

**Miền bác bỏ:** = ( = (2,65; +)

Trung bình mẫu từng nhóm:

1 = 2,14109

2 = 1,65260

3 = 1,72019

4 = 1,42061

Trung bình chung 4 mẫu:

Tổng chênh lệch bình phương cho từng nhóm:

Tổng chênh lệch bình phương trong các nhóm:

SSW = SS1 + SS2 + SS3 + SS4 = = 696,787

Tổng chênh lệch bình phương giữa các nhóm:

SSG = n1(1 - )2 + n2(2 - )2 +n3(3 - )2 + n4(4 - )2

= 12,41695

Phương sai giữa các nhóm:

MSG = = = 4,1390

MSW = = 3,76642

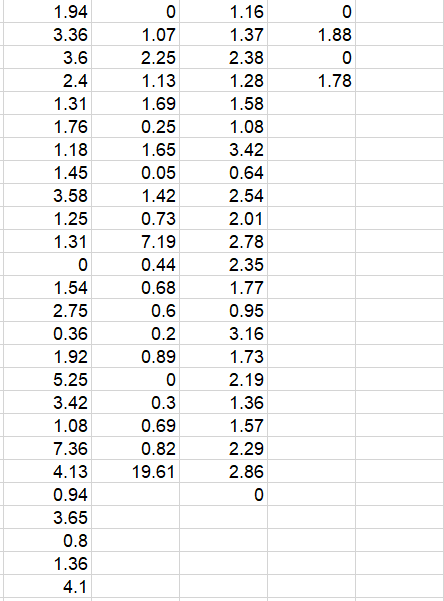
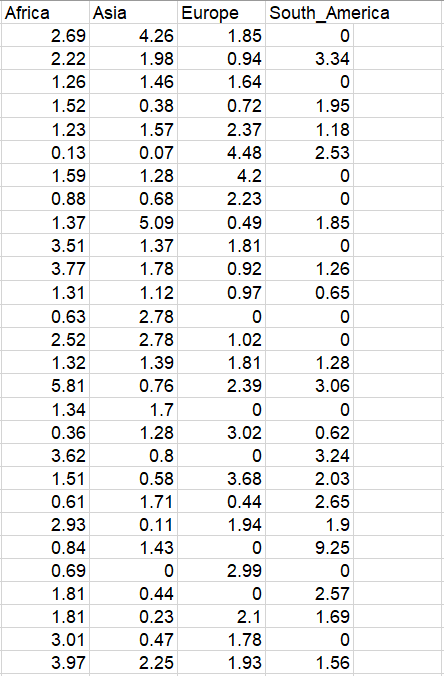
**Tiêu chuẩn kiểm định:** F = = = 1,0989

Vì nên chấp nhận giả thiết H0

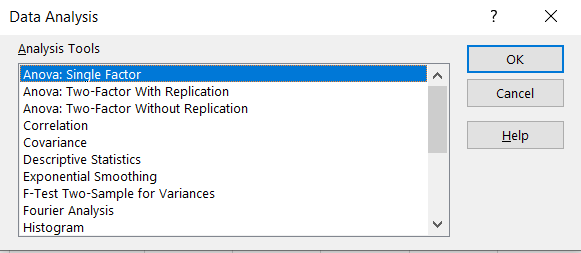
**Kết luận:** Tỉ lệ ca tử vong trên tổng số ca bệnh do nCov-19 của 4 châu lục là như nhau.

**PHẦN 4: GIẢI QUYẾT BÀI TOÁN TRÊN EXCEL**

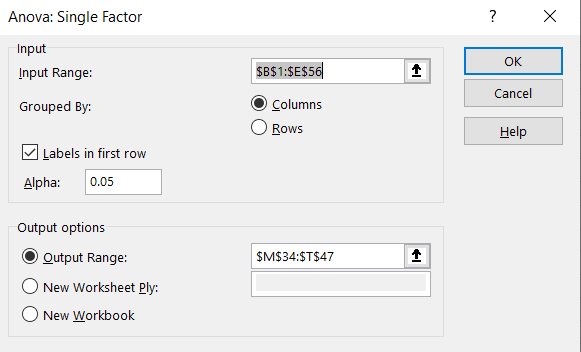
**Bước 1:** Nhập dữ liệu:

****

**Bước 2:** Sử dụng công cụ Data Analysis/Anova: Single Factor:



**Bước 3:** Chọn các mục như hình:



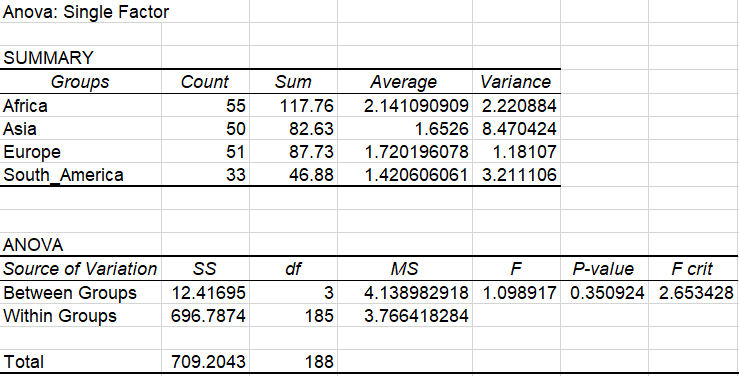
Input Range: Chọn địa chỉ tuyệt đối chứa dữ liệu

Output Range: Chọn vị trí xuất kết quả.

Chọn Labels in first row

Alpha: nhập 0.05 (mức ý nghĩa 5%)

Kết quả:



**Bước 4:** Biện luận:

**Giả thiết H0:** Tỉ lệ ca tử vong trên tổng số ca bệnh do nCov-19 của 4 châu lục là như nhau.

**Giả thiết H1:** Có ít nhất 2 châu lục mà tỉ lệ ca tử vong trên tổng số ca bệnh là khác nhau.

**Tiêu chuẩn kiểm định:** 1,09892

**Miền bác bỏ:** 2.65

Vì nên chấp nhận giả thiết H0

**Kết luận:** Vậy tỉ lệ tử trong trên tổng số ca nhiễm nCov-19 của 4 châu lục là như nhau. Dựa trên quả thu được, ta thấy tỉ lệ tử vong !..!.

**[Câu 4](#De_4): Chọn dữ liệu cho 2 biến để lập bài toán kiểm định so sánh về sự phân tán. Trình bày các bước thực hiện và nhận xét kết quả.**

**PHẦN 1: LẬP BÀI TOÁN**

Việc dự đoán điểm thi giữa kì của 2 sinh viên khoa XD và khoa Điện , ta dựa vào hiệu suất trung bình được đánh giá từ câu 2. Nhưng kết quả của việc đánh giá này bị ảnh hưởng bởi mức độ phân tán của điểm thi giữa kì của sinh viên 2 khoa Điện và khoa XD. Từ bảng lấy số liệu đầu vào dưới đây của sinh viên khoa Điện và sinh sinh viên khoa Xây Dựng , hãy so sánh sự phân tán của điểm thi dự đoán giữa kì của sinh viên khoa Điện và sinh viên khoa Xây Dựng .

**Nhóm 1:** Sinh viên khoa Điện-Điện tử

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 5 | 7 | 7 | 8 | 7 | 6 | 5 | 7 | 5 | 7 | 7 | 5 |
| 5 | 5 | 7 | 6 | 4 | 5 | 5 | 6 | 5 | 10 | 9 | 5 |
| 7 | 3 | 8 | 6 | 3 | 7.5 | 6 | 5 | 6 | 7 | 3 | 6 |
| 7 | 4 | 9.5 | 5 | 8 | 5 | 5 | 10 | 7 | 7 | 10 | 4 |
| 5 | 2 | 4 | 6 | 10 | 9.5 | 6 | 4 | 10 | 6 | 3.5 | 7 |
| 7 | 6 | 2 | 4.5 | 5 | 7 | 2 | 5 | 6 | 7 | 5 | 10 |
| 4.5 | 5 |  |  |  |  |  |  |  |  |  |  |

**Nhóm 2:** Sinh viên khoa Xây Dựng

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 10 | 9 | 7 | 8 | 10 | 4 | 8 | 4 | 4 | 6 | 5 | 5.5 |
| 7 | 7 | 8 | 4 | 4 | 3 | 7 | 7 | 6 | 7 | 3 | 3 |
| 2 | 9 | 6 | 8.5 | 7.5 | 10 | 7 | 7 | 9 | 7 | 6 | 5 |
| 6 | 5 | 5 | 7 | 6.5 | 7 | 6 | 5 | 5 | 8.5 | 8 | 7.5 |
| 5 |  |  |  |  |  |  |  |  |  |  |  |

Thang điểm: từ 0 đến 10 (đơn vị: điểm)

Với mức ý nghĩa 5%, hãy so sánh sự phân tán điểm đánh giá dự đoán điểm giữa kì của Khoa Điện và khoa Xây Dựng.

**PHẦN 2: CƠ SỞ LÝ THUYẾT**

**Dạng bài:** Kiểm định phương sai

**Lý thuyết kiểm định phương sai hai tổng thể:**

Khi cần kiểm định hai tổng thể có biến động như nhau hay không chúng ta dùng phương pháp kiểm định phương sai của hai tổng thể độc lập dựa trên một đại lượng F như sau:

F = s12/s22, trong đó:

s12 là phương sai của mẫu thứ nhất, mẫu này có cỡ n1

s22 là phương sai của mẫu thứ hai , mẫu này có cỡ n2

Thông thường, để xác định mẫu nào thứ nhất và mẫu nào là mẫu thứ hai thì, trong khi tính đại lượng F, giá trị phương sai lớn hơn sẽ được đặt ở tử số và như vậy, mẫu tương ứng với phương sai lớn hơn đó là mẫu thứ nhất.

Nếu giả thiết đặt ra là kiểm định hai bên:

**Giả thiết H0:** σ12 = σ22

**Giả thiết H1:** σ12 ≠ σ22

Nếu tỉ số F rất lớn hoặc rất nhỏ: ta có thể suy diễn bằng hai phương sai tổng thể khó mà bằng nhau, ngược lại nếu tỉ số này gần đến 1 ta sẽ có bằng chứng ủng hộ giả thuyết H0. Như vậy tỉ lệ F lớn đến đâu thì xem như là đủ bằng chứng bác bỏ H0 và ngược lại.

Nếu tổng thể lấy mẫu được giả định có phân phối bình thường: tỉ lệ F có phân phối xác suất, gọi tên là phân phối Fisher. Các giá trị tới hạn của phân phối F phụ thuộc và hai giá trị bậc tự do, là bậc tự do tử số (df1 = n1 -1) gắn liền với mẫu thứ nhất và bậc tự do mẫu số gắn liền với mẫu thứ hai (df2 = n2 -1).

Quy tắc thực sự để bác bỏ H0 với kiểm định hai bên (khi df1 =n1 - 1 ; df2 = n2 - 1, mức ý nghĩa α) là:

Giả thiết H0 bị bác bỏ nếu:

Giá trị kiểm định F lớn hơn giá trị tới hạn trên FU = Fdf1;df2;α/2 của phân phối F, tức là Ftt > Fdf1;df2;α/2 h oặc

Giá trị kiểm định F bé hơn giá trị tới hạn dưới FL = Fdf1;df2;α/2  của phân phối F, tức là Ftt < Fdf1;df2;α/2

* Nếu chúng ta kiểm định bên phải:

**Giả thiết H0:** σ12 = σ22

**Giả thiết H1:** σ12 > σ22

Quy tắc bác bỏ H0 : khi Ftt > FU(n1-1;n2-1;α)

* Nếu chúng ta kiểm định bên trái:

**Giả thiết H0:** σ12 = σ22

**Giả thiết H1:** σ12 < σ12

Quy tắc bác bỏ H0: khi Ftt < FU(n1-1;n2-1;α)

Phương pháp kiểm định phương sai tổng thể:

Giả thiết kiểm định H0 và giả thiết đối H1.

Tiêu chuẩn kiểm định: F = S12 / S22

Xác định miền bác bỏ tùy thuộc vào giá trị F tra bảng.

Dựa vào Tiêu chuẩn Kiểm định và Miền bác bỏ Wa , đưa ra kết luận.

**PHẦN 3: LỜI GIẢI TÍNH TAY**

Gọi σ12, σ12 là lần lượt là phương sai của điểm đánh giá mức độ tập trung khi học tập tại nơi ở của sinh viên nam và sinh viên nữ.

n1 = 74; s12 = 3,8965; n2 = 49, s22 = 3,8102

**Giả thiết H0:** σ12 = σ22 ⭤ “Phương sai về điểm dự đoán của 2 sinh viên khoa Điện và khoa Xây Dựng là như nhau.”.

**Giả thiết H1**: σ12 > σ22 ⭤ “Phương sai về điểm dự đoán của sinh viên khoa Điện lớn hơn phương sai của khoa Xây Dựng.

Miền bác bỏ: Wα = ( Fα(n1-1; n2-1); +∞) =

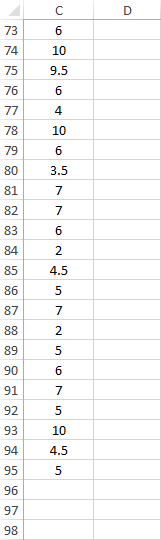
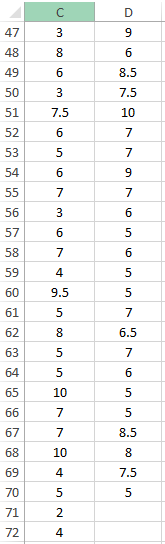
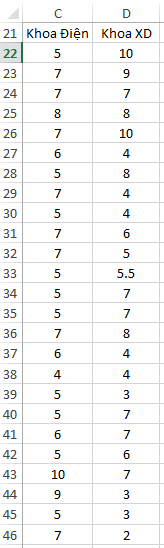
**Tiêu chuẩn kiểm định:** F= = = 1,0227

Vì F= 1.0227 ⇨ chấp nhận giả thiết H0

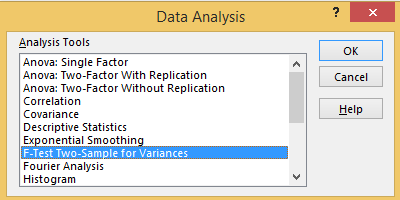
**Kết luận:** Vậy phương sai của điểm dự đoán Khoa Điện như nhau so với phương sai của điểm dự đoán Khoa Xây Dựng.

**PHẦN 4: GIẢI QUYẾT BÀI TOÁN TRÊN EXCEL**

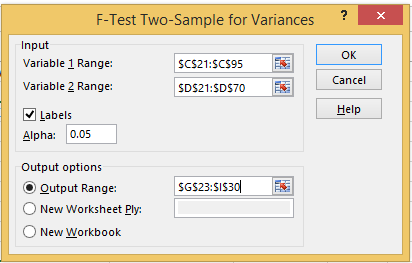
**Bước 1:** Nhập dữ liệu:



**Bước 2:** Sử dụng công cụ Data Analysis/ F – Test Two Samples for Variances



**Bước 3:** Chọn các mục như hình:



Variable 1 Range: Quét địa chỉ chứa dữ liệu 1.

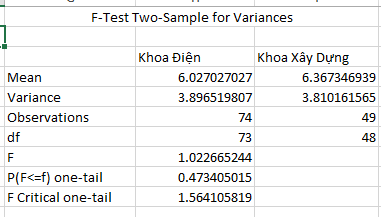
Variable 2 Range: Quét địa chỉ chứa dữ liệu 2.

Output Range: Vị trí xuất kết quả.

Chọn Labels in first row.

Alpha: Nhập 0.05 (Mức ý nghĩa 5%)

Kết quả:



**Bước 4:** Biện luận

Gọiσ12, σ22 lần lượt là phương sai điểm đánh giá điểm dự đoán của 2 sinh viên khoa Điện và khoa Xây Dựng

**Giả thiết H0:** σ12 = σ22 ⭤ “Phương sai về điểm dự đoán của 2 sinh viên khoa Điện và khoa Xây Dựng là như nhau.”.

**Giả thiết H1:** σ12 > σ22 ⭤ “Phương sai về điểm dự đoán của sinh viên khoa Điện lớn hơn phương sai của khoa Xây Dựng.

Miền bác bỏ : (F Critical one – tail; ) = (1.5641; )

* Kết luận : có 2 cách

**Cách 1 :** Sử dụng giá trị F(tiêu chuẩn kiểm định)

Ta có : F= 1.0227 ∉ ⇨ chấp nhận H0

**Cách 2 :** Sử dụng giá trị P(F <= f) one – tail

* Nếu P(F <= f) one – tail > (mức ý nghĩa) ⇨ chấp nhận H0
* Nếu P(F <= f) one – tail < (mức ý nghĩa) ⇨ bác bỏ H0, chấp nhận H1

Quay lại bài toán : ta có P(F <= f) one – tail = 0.473>

Suy ra Chấp nhận H0

**Bước 5: Kết luận bài toán**

Vậy phương sai của điểm dự đoán Khoa Điện như nhau so với phương sai của điểm dự đoán Khoa Xây Dựng. Nghĩa là độ phân tán điểm dự đoán của Khoa Điện với Khoa Xây Dựng là như nhau.

**[Câu 5](#De_5): Chọn dữ liệu phù hợp để lập bài toán so sánh các tỉ lệ ( hay là bài toán kiểm định tính độc lập). Trình bày các bước thực hiện và nhận xét kết quả.**

**PHẦN 1: LẬP BÀI TOÁN**

Bảng số liệu dưới đây thể hiện tỉ lệ ca tử vong do nCov-19 theo độ tuổi trung bình của từng quốc gia:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Ca tử vong (%)  Tuổi thọ | [0, 1.5) | [1.5, 4.5) | [4.5, 100) | **Tổng hàng** |
| Trung bình  [0,65) | 17 | 19 | 1 | **37** |
| Cao  [65, 75) | 32 | 29 | 5 | **66** |
| Rất cao  [75, + ) | 37 | 47 | 3 | **87** |
| **Tổng cột** | **86** | **95** | **9** | **190** |

!!! Với mức ý nghĩa 5%, hãy cho biết xyz phân bố tỉ lệ như nhau ở từng độ tuổi trung bình hay không?

**PHẦN 2: CƠ SỞ LÝ THUYẾT**

**Dạng bài**: Kiểm định tính độc lập

**Phương pháp kiểm định tính độc lập:**

Đối với một thí nghiệm có hai kết quả (binomial experiment) – thí dụ, đối với một thuốc được kê đơn: có hay không – bạn thường so sánh 2 tỷ số với nhau (thực nghiệm với lý thuyết hay thực nghiệm với thực nghiệm). Song đối với một thí nghiệm bệnh nhân được điều trị bởi thuốc trong một khoảng thời gian – bạn cần so sánh nhiều tỷ số. Trắc nghiệm “khi” bình phương () cho phép bạn so sánh không những hai mà còn nhiều tỷ số (hay tỷ lệ hoặc xác suất), không có tính đối xứng và chỉ có giá trị 0.

Giả sử bạn có một công trình mang một xác suất thực nghiệm là Pi (i = 1, 2, …, k). Nếu gọi Pi,0 là các giá trị ký thuyết tương ứng với Pi thì các tần số lý thuyết sẽ là Ei = N.Pi,0.

Điều kiện để áp dụng trắc nghiệm () một cách thành công là các tần số lý thuyết Ei phải 5.

**Giả thiết:**

H0: P1 = P1,0, P2 = P2,0, …, Pk,0 ⭤ “Các cặp Pi và Pi,0 giống nhau”.

H1: “Ít nhất có một cặp Pi và Pi,0 khác nhau”.

**Giá trị thống kê:**

**Giá trị thống kê:**

Trong đó : các tần số thực nghiệm (observerd frequency);

: các tần số lý thuyết (expeceted frequency)

**Biện luận:** >  bác bỏ giả thiết H0

Trong chương trình MS-EXCEL có ***hàm số CHITEST*** có thể tính:

* Giá trị 2 : tính theo biểu thức:



*Với:*

**Oij** : tần số thực nghiệm của ô thuộc hàng i và cột j.

**Eij** : tần số lý thuyết của ô thuộc hàng i và cột j.

**r** : số hàng

**c** : số cột

* Tính xác suất P(X>), với bậc tự do *DF* = (r – 1).(c – 1); trong đó, r là số hàng và c là số cột trong bảng ngẫu nhiên (contingency table).
* Nếu P(X>) > α 🠚 Chấp nhận giả thiết H0 và ngược lại.

**PHẦN 3: LỜI GIẢI TÍNH TAY**

**Giả thiết H0:** Tỉ lệ ca tử vong trên tổng số ca bệnh do nCov của một quốc gia không phụ thuộc vào độ tuổi dân số trung bình của quốc gia đó.

**Giả thiết H1:** Tỉ lệ ca tử vong trên tổng số ca bệnh do nCov của một quốc gia phụ thuộc vào độ tuổi dân số trung bình của quốc gia đó.

**Miền bác bỏ:** W h – 1)(k – 1);+∞ )= ( (2)(2); +∞) = (19.00; +∞)

**Bảng tần số thực nghiệm :**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Ca tử vong (%)  Tuổi thọ | [0, 1.5) | [1.5, 4.5) | [4.5, 100) | **Tổng hàng** |
| Trung bình  [0,65) | 17 | 19 | 1 | **37** |
| Cao  [65, 75) | 32 | 29 | 5 | **66** |
| Rất cao  [75, + ) | 37 | 47 | 3 | **87** |
| **Tổng cột** | **86** | **95** | **9** | **190** |

Lập bảng tần số lý thuyết Eij

Eij =

E11== 16,7474 E12= = 18,5 E13= = 1,7526

E21= = 29,8737 E22= =33 E23= = 3,1263

E31 = = 39,3790 E32== 43,5 E33= = 4,1211

|  |  |  |
| --- | --- | --- |
| 16.74737 | 18.5 | 1.752632 |
| 29.87368 | 33.0 | 3.126316 |
| 39.37895 | 43.5 | 4.121053 |

Tính giá trị 2 :



- 2= = + + ++++++



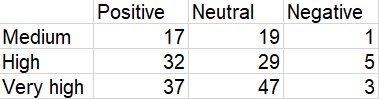
- = = + + + + + + + = 0.586673

Vì ∉ W nên chấp nhận H0 bác bỏ H1

**Kết luận:** Điểm dự kiến của sinh viên học môn Xác suất thống kê kỳ 201 không phụ thuộc vào giảng viên dạy môn học Xác suất thống kê.

**PHẦN 4: GIẢI QUYẾT BÀI TOÁN TRÊN EXCEL**

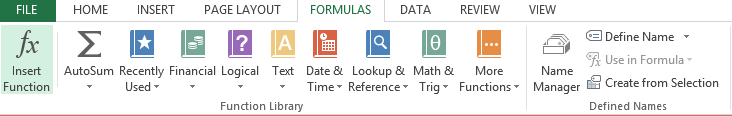
**Bước 1:** Nhập dữ liệu:



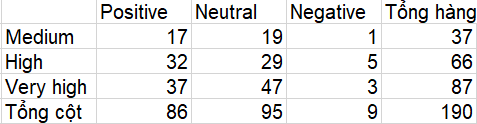
**Bước 2:** Tính tổng hàng, tổng cột

Quét bảng số liệu dư 1 hàng 1 cột.

Chọn chức năng Auto Sum (tự động tính tổng) trong Formulas:



Kết quả:



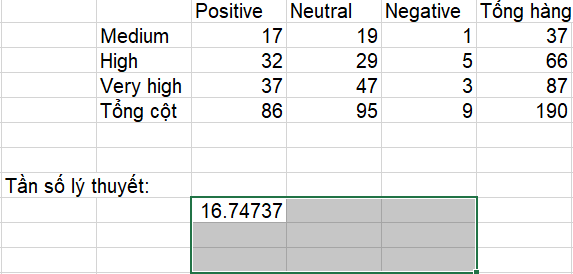
**Bước 3:** Lập bảng tần số lý thuyết:

Tần số lý thuyết được tính theo công thức:

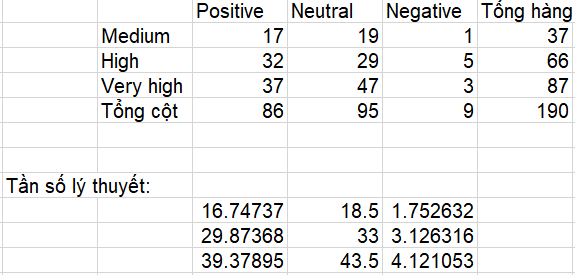
Tính giá trị tần số lý thuyết:

Tính giá trị tần số lý thuyết đầu tiên:

Quét chọn 9 ô chứa các tần số lý thuyết cần tính:

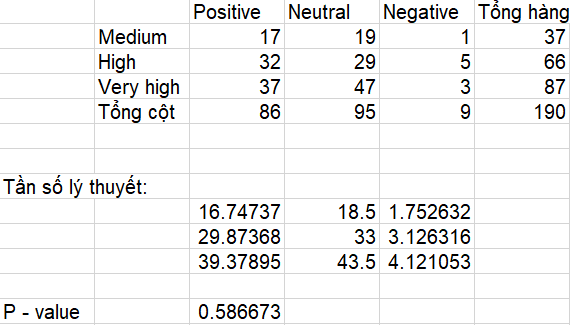


Kết quả:



Sử dụng tổ hợp phím F2 - (ctrl + enter)

**Bước 4:** Tính P – Value bằng hàm “Chitest”:

****

***Bước 5:* Biện luận**

* Giả thiết:

H0 : Mức độ thường xuyên tương tác với giảng viên khi học Hangout có phân bố tỷ lệ như nhau ở 2 khóa sinh viên.

H1 : Mức độ thường xuyên tương tác với giảng viên khi học Hangout có phân bố tỷ lệ không như nhau ở 2 khóa sinh viên.

* **Mức ý nghĩa quan sát :** **P – Value** = 0.58976

Mức ý nghĩa:

Vì **P – Value >** → Chấp nhận giả thiết H0

* **Kết luận:** Điểm dự kiến của sinh viên học môn Xác suất thống kê kỳ 201 không phụ thuộc vào giảng viên dạy môn học Xác suất thống kê.

# **Câu 6: Khảo sát hệ số tương quan giữa 2 biến cụ thể, dự đoán phương trình đường hồi quy tuyến tính giữa chúng (có hình vẽ minh họa) và nhận xét về mối tương quan giữa 2 biến.**

**PHẦN 1: LẬP BÀI TOÁN**

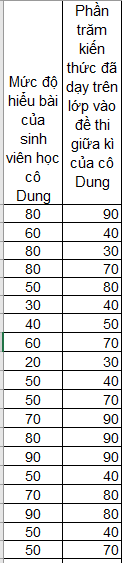
Điểm thi giữa kì cho kết quả thế nào phụ thuộc ở việc mức độ hiểu bài của sinh viên và phần trăm kiến thức đã dạy trên lớp vào đề thi giữa kì của thầy cô.

Để làm rõ vấn đề này ta khảo sát hệ số tương quan giữa 2 biến

X:Mức độ hiểu bài của sinh viên học cô Dung

Y:Phần trăm kiến thức đã dạy trên lớp vào đề thi giữa kì của cô Dung

Bảng số liệu thu được:



dự đoán phương trình đường hồi tuyến tính giữa chúng và nhận xét mối tương của mức độ hiểu bài và phần trăm kiến thức đã dạy trên lớp (mức ý nghĩa 5%).

**PHẦN 2: CƠ SỞ LÝ THUYẾT**

**Dạng bài:** Phân tích tương quan và hồi quy.

**Nội dung 1:** Hệ số tương quan R:

R > 0: X, Y tương quan thuận.

R < 0: X, Y tương quan nghịch.

R = 0: X, Y không tương quan.

Nếu | R | = 1, X, Y có quan hệ hàm bậc nhất.

Nếu | R |→ 1: X, Y có tương quan chặt (tương quan mạnh)

Nếu | R |→ 0: Xi X, Y có tương quan không chặt (tương quan yếu)

**Nội dung 2:** Ước lượng đường hồi quy:

* Phương trình hồi quy tuyến tính: y = a + b.x với:
* Kiểm định hệ số a, b:

**Giả thiết H0:** Hệ số hồi quy không có ý nghĩa (hệ số bằng 0).

**Giả thiết H1:** Hệ số hồi quy có ý nghĩa (hệ số khác 0).

**Trắc nghiệm:** t < tα,n-2  nên chưa bác bỏ H0.

* Kiểm định phương trình hồi quy:

**Giả thiết H0:** Phương trình hồi quy tuyến tính không thích hợp.

**Giả thiết H1:** Phương trình hồi quy tuyến tính thích hợp.

**Trắc nghiệm:** F < Fα,1,n-2 nên chưa bác bỏ H0

**Nội dung 3:** Quan hệ giữa X và Y:

Chúng ta muốn kiểm định các giả thiết liên quan đến giả thiết các giá trị khác nhau của hệ số tương quan tổng thể, ký hiệu , dựa trên phân phối mẫu của hệ số tương quan mẫu R.

**Giả thiết:**

Giả thiết H0: = 0

Giả thiết H1: 0 (hoặc > 0, hoặc < 0)

Người ta chứng minh được rằng với giả thiết H0, phân phối mẫu của R đối xứng; từ đó, thống kê:

T = R~ Student (n – 2)

Trắc nghiệm t được dụng trong trường hợp này

**PHẦN 3: LỜI GIẢI TÍNH TAY**

**Nội dung 1:** Xác định hệ số tương quan

Ta có:

57,60465

65,22093

Hệ số tương quan mẫu:

0,475238399

**Nội dung 2:**

* Dự đoán phương trình đường hồi quy tuyến tính giữa chúng

Phương trình đường hồi quy Y theo X có dạng: y = a + b.x với:

Vậy phương trình hồi quy Y theo X:

* Kiểm định hệ số tự do và hệ số góc có ý nghĩa thống kê hay không?

**Giả thiết H0:** Hệ số tự do và hệ số góc không có ý nghĩa thống kê.

**Giả thiết H1:** Hệ số tự do và hệ số góc có ý nghĩa thống kê.

**Miền bác bỏ:**

( ; ) (); + ) = ( ∞; 1,9889) (1,98896 ; + )

**Tính sai số chuẩn:**

**Tính sai số chuẩn của hệ số tự do:**

**Tính sai số chuẩn của hệ số góc:**

**Tiêu chuẩn kiểm định (ứng với hệ số tự do):**

**Tiêu chuẩn kiểm định (ứng với hệ số góc):**

Vì nên bác bỏ giả thiết H0, chấp nhận giả thiêt H1.

**Kết luận:** Vậy hệ số tự do có ý nghĩa thống kê.

Vì nên bác bỏ giả thiết H0, chấp nhận giả thiêt H1.

**Kết luận:** Vậy hệ số tự do có ý nghĩa thống kê.

* Kiểm định phương trình hồi quy có thích hợp hay không?

**Giả thiết H0:** Phương trình đường hồi quy không thích hợp.

**Giả thiết H1:** Phương trình đường hồi quy thích hợp.

**Miền bác bỏ:**

**Tiêu chuẩn kiểm định:**

Vì nên bác bỏ giả thiết H0, chấp nhận giả thiêt H1.

**Kết luận:** Phương trình đường hồi quy thích hợp.

**Nội dung 3:** Nhận xét về mối tương quan tuyến tính giữa 2 biến.

**Giả thiết H0:** X, Y không có tương quan tuyến tính.

**Giả thiết H1:** X , Y có tương quan tuyến tính.

**Miền bác bỏ:**

2,2827912,282791

**Tiêu chuẩn kiểm định:**

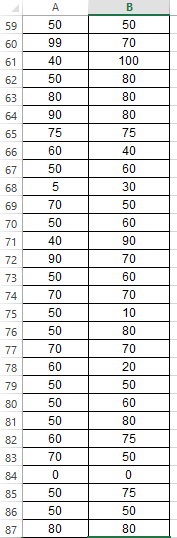
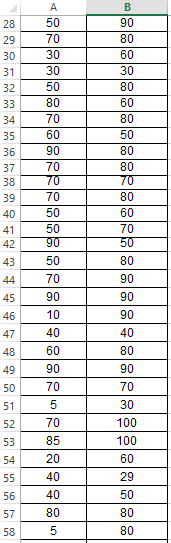
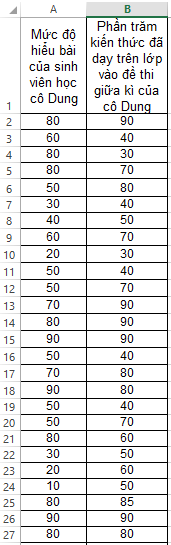
Vì nên bác bỏ giả thiết H0, chấp nhận giả thiêt H1.

**Kết luận:** X , Y có tương quan tuyến tính.

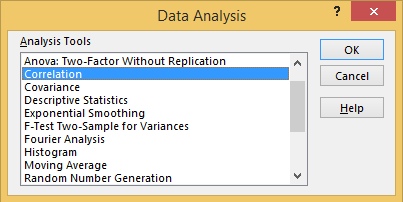
**PHẦN 4: GIẢI QUYẾT BÀI TOÁN TRÊN EXCEL**

**Nội dung 1:** Xác định hệ số tương quan

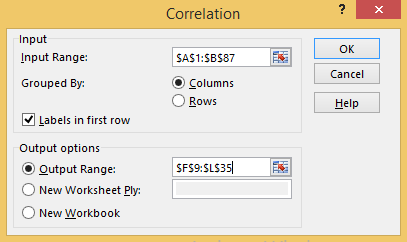
**Bước 1:** Nhập dữ liệu



**Bước 2:** Sử dụng công cụ Data Analysis/ Correlation.



**Bước 3:** Chọn các mục như hình:



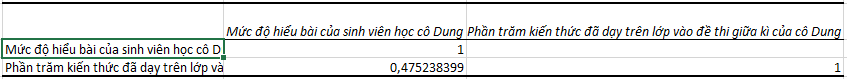
Input Range: Quét địa chỉ chứa dữ liệu

Chọn: Columns

Chọn Labels in first row

Output Range: chọn vị trí cần xuất kết quả

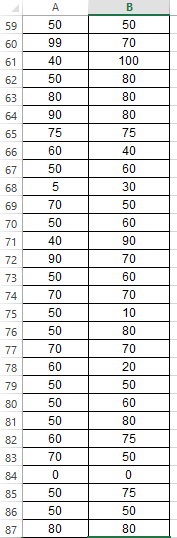
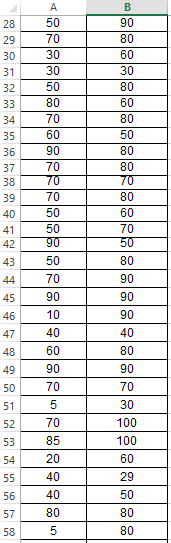
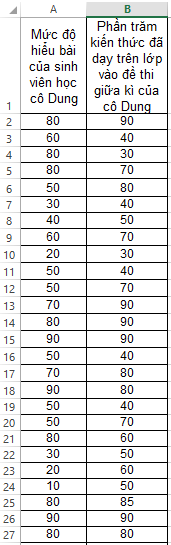
Kết quả:



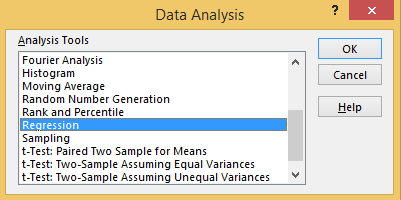
Kết luân: dựa vào kết quả trên ta có hệ số tương quan :R=0.475238399 chứng tỏ giữa mức độ hiểu bài và phần trăm kiến thức có mối tương quan kém và có tương quan thuận.

**2.Xác định phương trình hồi quy tuyến tính:**

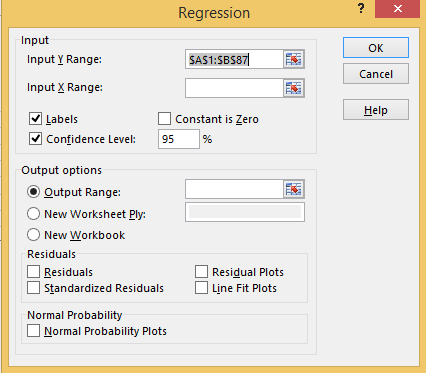
+ Nhập bảng số liệu cần xử lý và excel gồm cột mức độ hiểu bài của sinh viên học cô Dung và phần trăm kiến thức đã dạy trên lớp vào đề thi giữa kì của Dung



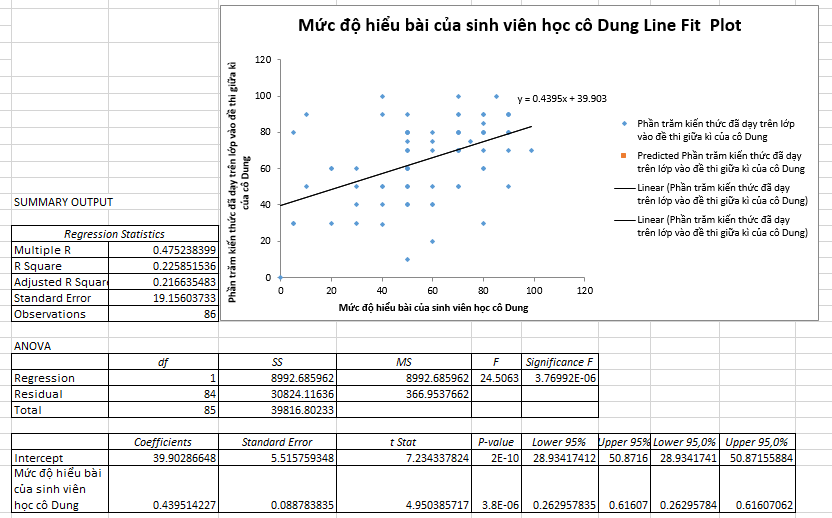
+ Chọn Data analysis => Sử dụng công cụ regression:

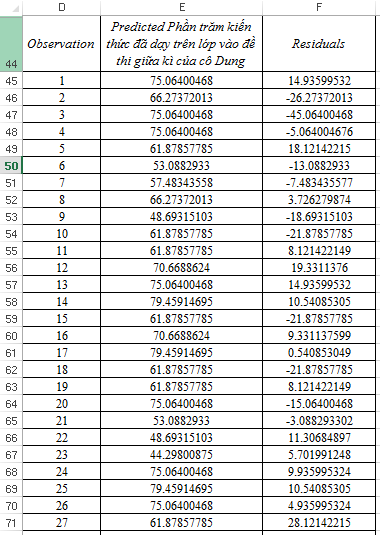
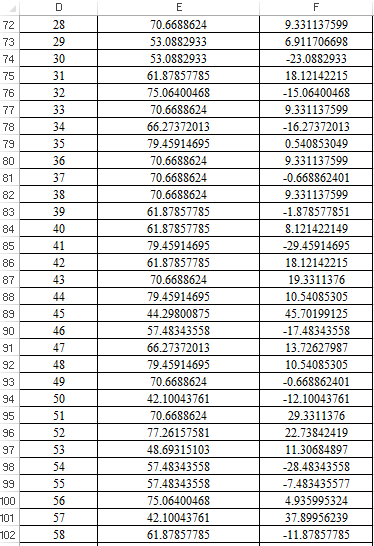


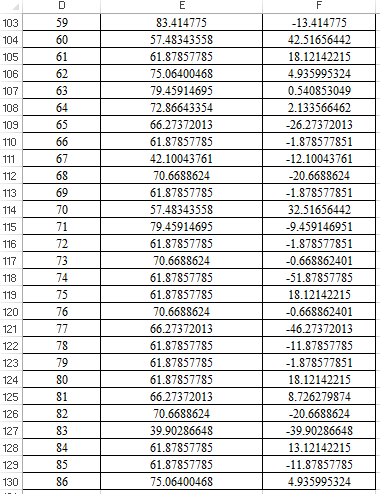
+Chọn địa chỉ tuyệt đối chứa dữ liệu(input range) => chọn vị trí xuất kết quả ( output range) => chọn labels và confidence level, line fit plot:



+ Cuối cùng xuất kết quả và kết luận:







* **Biện luận:**

*Kiểm định hệ số a,b có ý nghĩa thống kê(a là hệ số tự do, b là hệ số góc):*

Giả thiết H0: hệ số hồi quy(a,b) không có nghĩa (=0)

Giả thiết H1: hệ số hồi quy(a,b) có nghĩa(≠0)

P-value= 2,02E-10<0.05 Vậy hệ số tự do a có ý nghĩa thống kê.

P-value= 3,77E-06<0,05Bác bỏ H0, chấp nhận H1. Vậy hệ số góc b có ý nghĩa thống kê.

*Kiểm định đường hồi quy tuyến tính:*

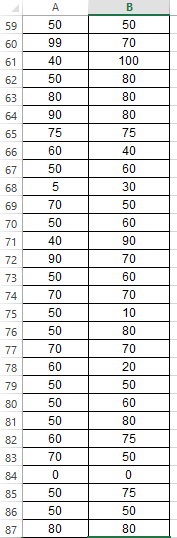
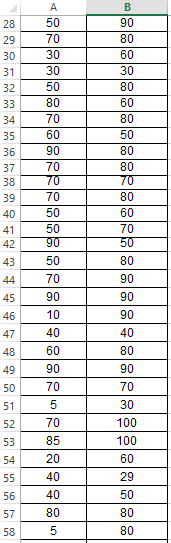
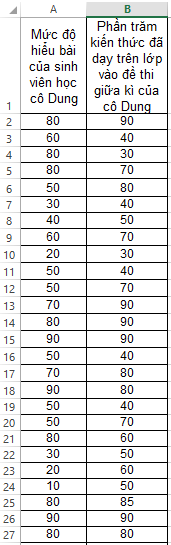
Giả thiết H0: phương trình hồi quy tuyến tính không thích hợp.

Giả thiết H1: phương trình hồi quy tuyến tính thích hợp.

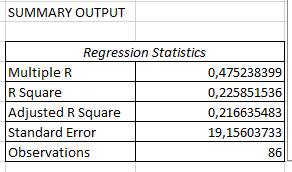
Significance F= 3,77E-06<0,05Bác bỏ H0, chấp nhận H1. Vậy phương trình hồi quy tuyến tính thích hợp.

**3.Nhận xét về sự tương quan giữa mức độ hiểu bài của sinh viên học cô Dung và phần trăm kiến thức đã dạy trên lớp vào đề thi giữa kì của Dung**

+Nhập bảng số liệu cần xử lý và excel gồm cột mức độ hiểu bài của sinh viên học cô Dung và phần trăm kiến thức đã dạy trên lớp vào đề thi giữa kì của Dung



+Tính giá trị quan sát: T=R



Giả thiết H0: mức độ hiểu bài của sinh viên học cô Dung và phần trăm kiến thức đã dạy trên lớp vào đề thi giữa kì của Dung không có sự tương quan tuyến tính

Giả thiêt H1: mức độ hiểu bài của sinh viên học cô Dung và phần trăm kiến thức đã dạy trên lớp vào đề thi giữa kì của Dung có sự tương quan tuyến tính

* >c Bác bỏ H0, chấp nhận H1. Vậy mức độ hiểu bài của sinh viên học cô Dung và phần trăm kiến thức đã dạy trên lớp vào đề thi giữa kì của Dung có sự tương quan tuyến tính

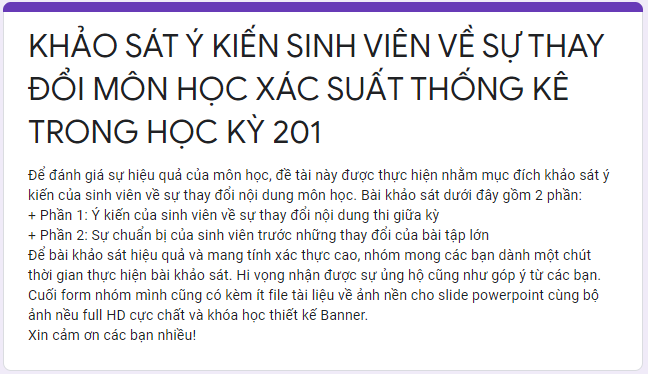
**TÀI LIỆU THAM KHẢO**

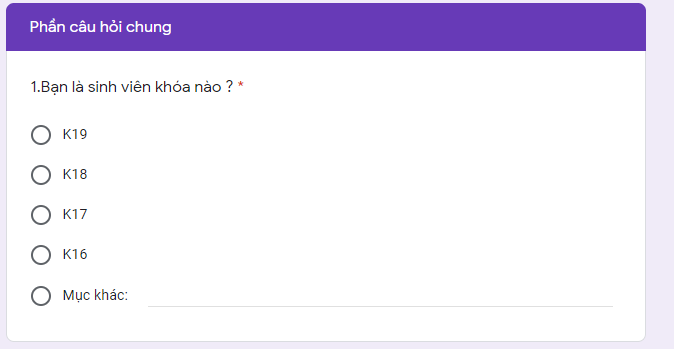
Nguyễn Kiều Dung, Bài Giảng XSTK trên E-Leanrning.

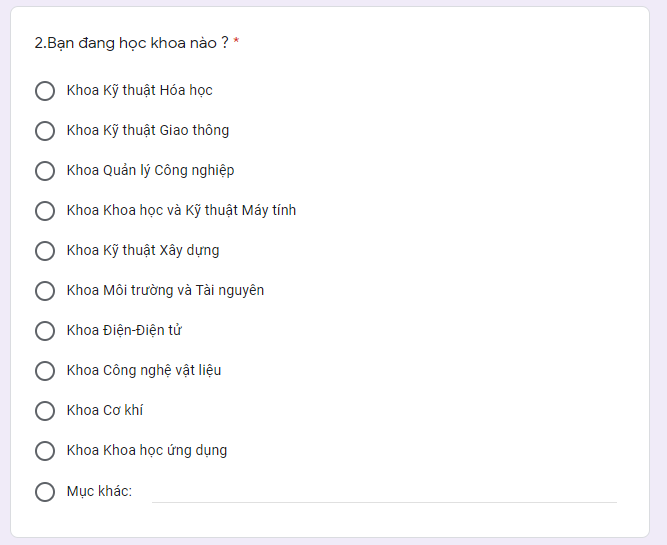
Nguyễn Đình Huy, Đậu Thế Cấp, Lê Xuân Đại (2013). Giáo trình xác suất thống kê, NXB Đại Học Quốc Gia, TPHCM.

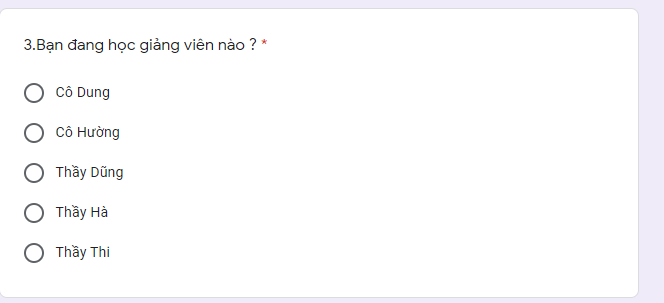
Nguyễn Tiến Dũng, Nguyễn Đình Huy (2019). Xác suất – Thống kê & Phân tích số liệu, NXB Đại Học Quốc Gia, TPHCM.

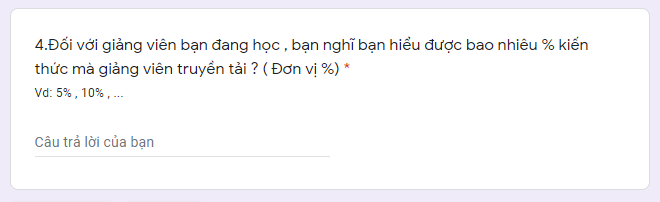
Tập số liệu mẫu thu được từ việc khảo sát thực tế do các sinh viên trong nhóm thực hiện từ ngày 23/11 – 30/11

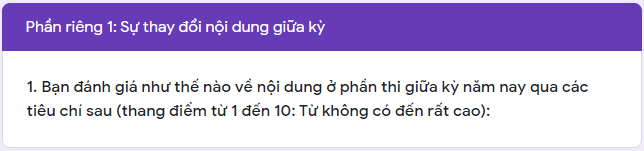


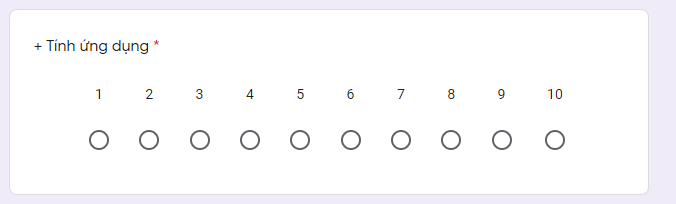




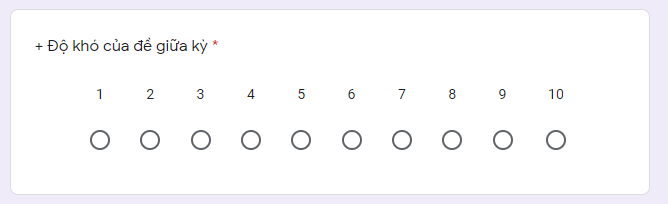


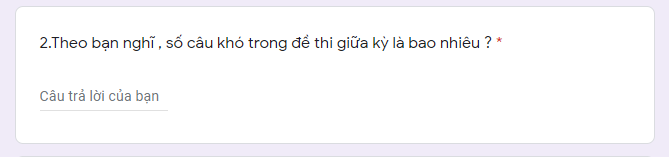


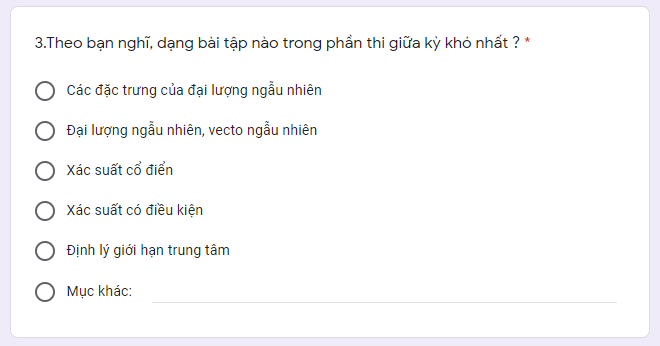


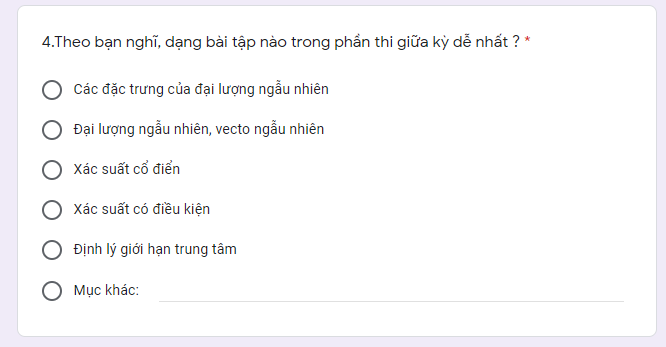


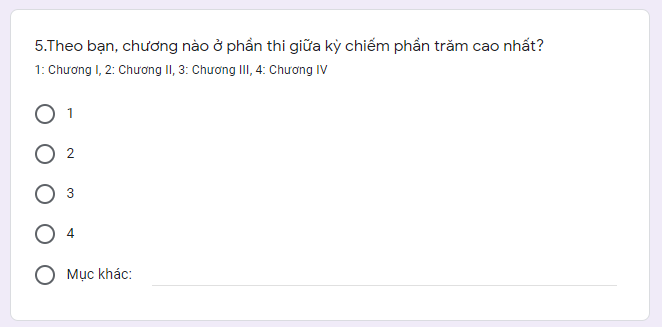


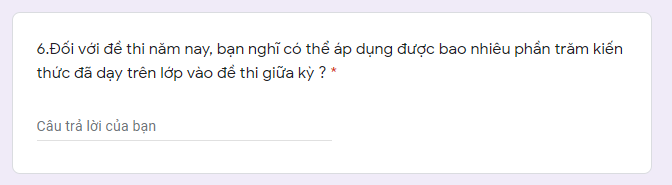


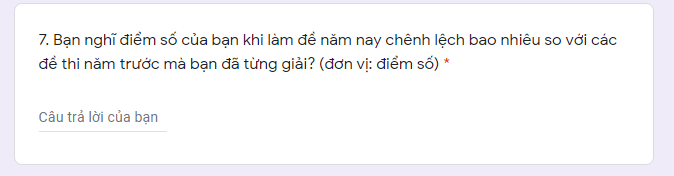


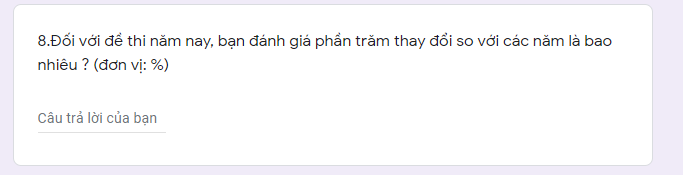


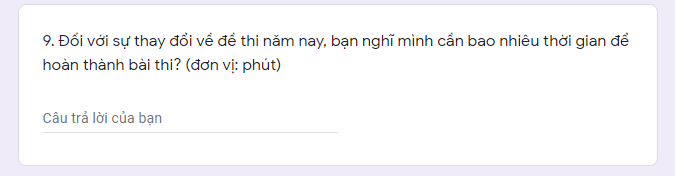


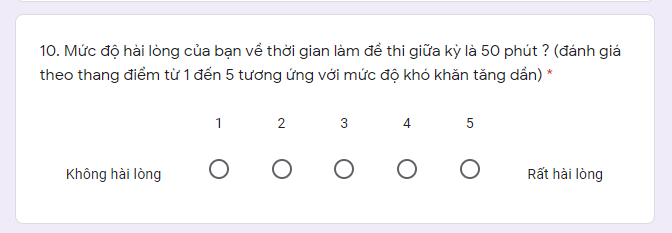


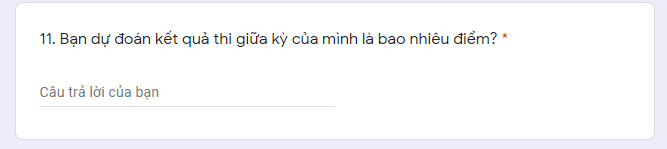


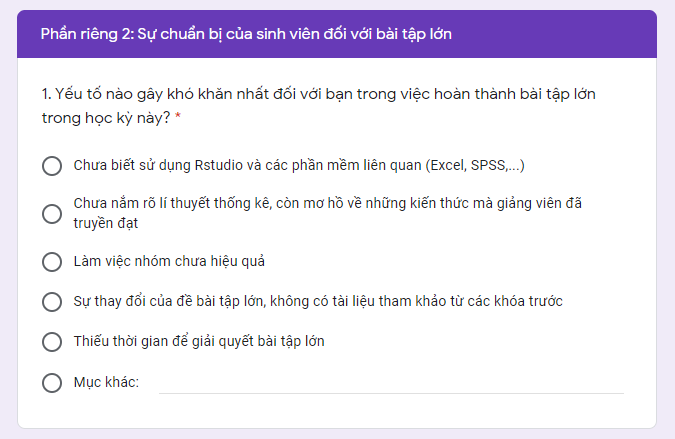


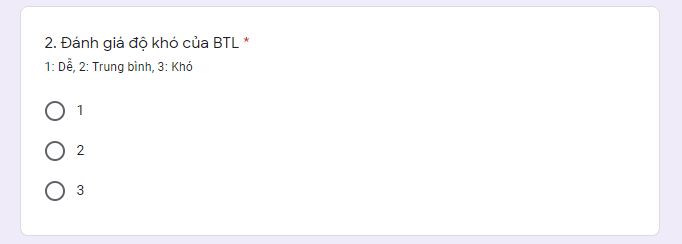


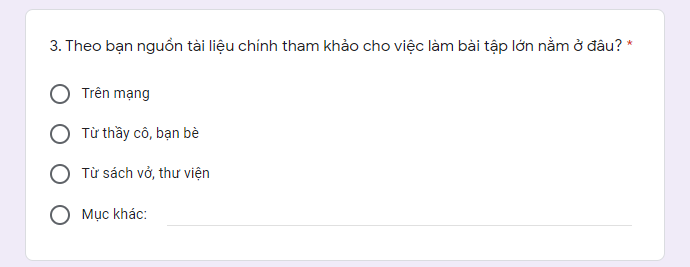


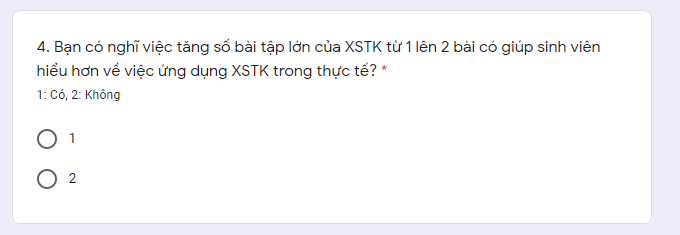


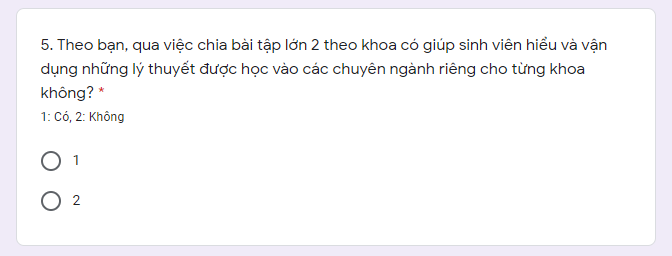


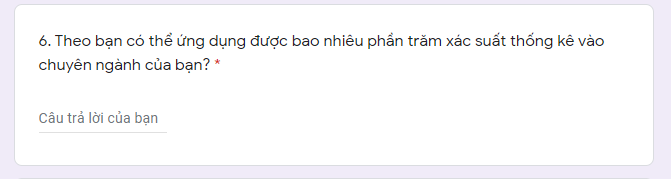


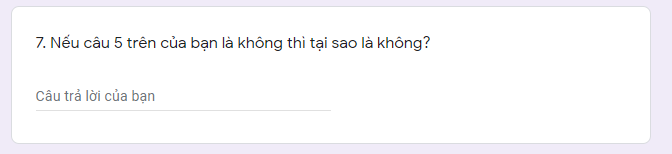


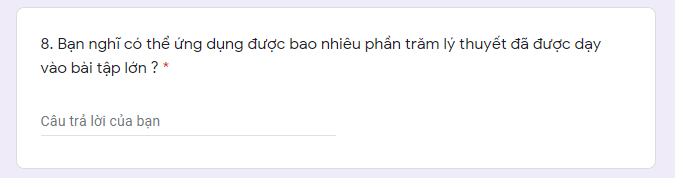


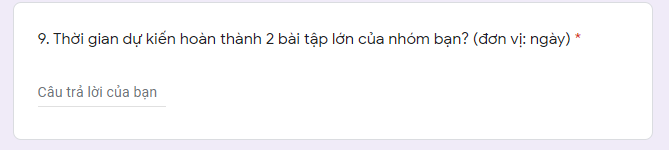


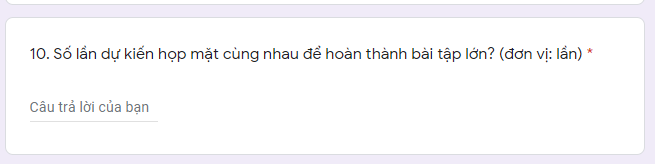


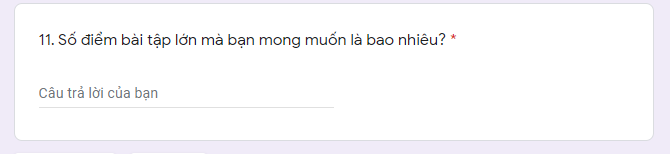












Nhóm chúng em xin cảm ơn Cô đã dành thời gian đọc hết bài Báo Cáo môn học của chúng em. Chúng em xin chúc Cô nhiều sức khỏe, tinh thần luôn thoải mái tươi vui để thực hiện công việc giảng dạy của mình ngày càng nhiều hiệu quả hơn nữa ạ.

Chúng em xin dừng bài Báo Cáo tại đây.