**[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 tỉ lệ ca tử vong do nCov có 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= = + + ++++++



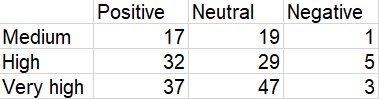
- = = + + + + + + + = 0.586673

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**

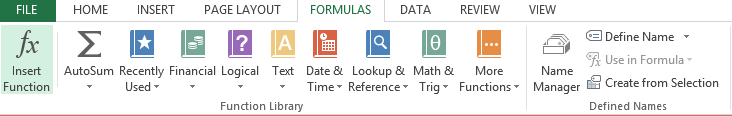
**Bước 1:** Nhập dữ liệu:



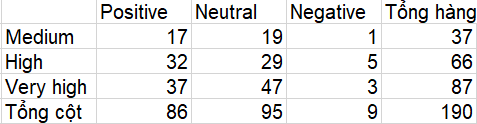
**Bước 2:** Tính tổng hàng, tổng cột

Quét bảng số liệu dư 1 hàng 1 cột.

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



Kết quả:



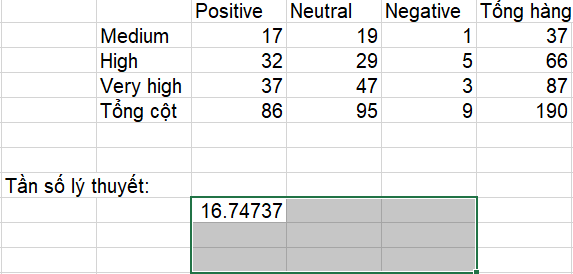
**Bước 3:** Lập bảng tần số lý thuyết:

Tần số lý thuyết được tính theo công thức:

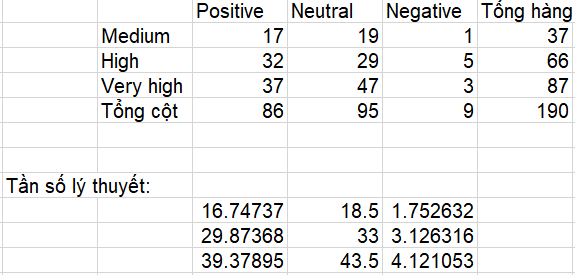
Tính giá trị tần số lý thuyết:

Tính giá trị tần số lý thuyết đầu tiên:

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

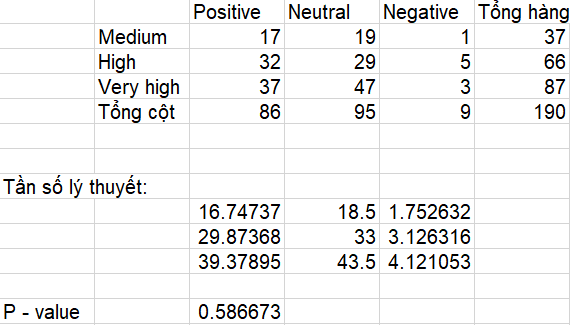


Kết quả:



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

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

****

***Bước 5:* Biện luận**

* Giả thiết:
* **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.
* **Mức ý nghĩa quan sát :** **P – Value** = 0.586673

Mức ý nghĩa:

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

* **Kết luận:** Tỉ lệ ca tử vong trên tổng số ca nhiễm do nCov của một quốc gia không phụ thuộc vào độ tuổi trung bình của quốc gia đó.