**ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH**



**TRƯỜNG ĐẠI HỌC BÁCH KHOA**

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



**BÁO CÁO BÀI TẬP LỚN XÁC SUẤT THỐNG KÊ**

**Lớp L13-Nhóm 12**

**Bài tập lớn 1(1.6)**

**Giảng viên hướng dẫn:** Hoàng Văn Hà

**Sinh viên thực hiện:** Hồ Trung Nghĩa 1911681

Lý Vĩnh Khang 1852443

Võ Đình Thanh 1912041

Trần Thế Quang 1914806

Nguyễn Huy Phúc 1914697

Võ Mạnh Hào 1913238

**TP. Hồ Chí Minh, tháng 12 năm 20****20**

**Mục Lục Trang**

Bài tập lớn 1(1.6)…………………………………………………….3

Câu 1 …………………………………………………………..3

Câu 2 …………………………………………………………..3

Câu 3 …………………………………………………………..5

Câu 4 …………………………………………………………..11

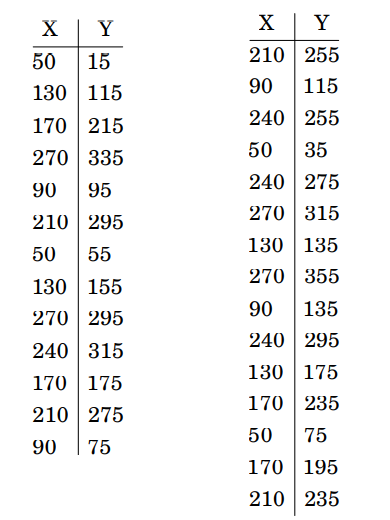
Câu 5 …………………………………………………………..14

**BÀI TẬP LỚN 1**

**Bài 1: Giải lại ví dụ 10 trang 172 và ví dụ 12 trang 181 Sách GT XSTK (N.Đ.HUY) bằng phần mềm thống kê R (Không cần làm )**

**Bài 2:**

**Tỷ số tương quan của Y đối với X và hệ số xác định của tập số liệu sau đây:**



**Có kết luận gì về mối tương quan giữa X và Y(tuyến tính hay phi tuyến)?**

**Bài làm**

**Cơ sở lý thuyết:**

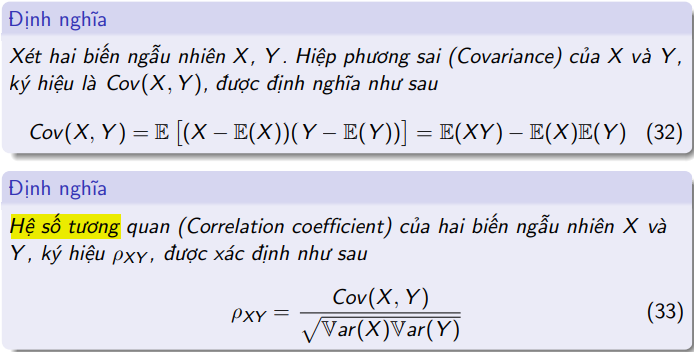
Hệ số tương quan là chỉ số thống kê đo lường mức độ mạnh yếu của mối quan hệ giữa hai biến số. *Trong đó:*

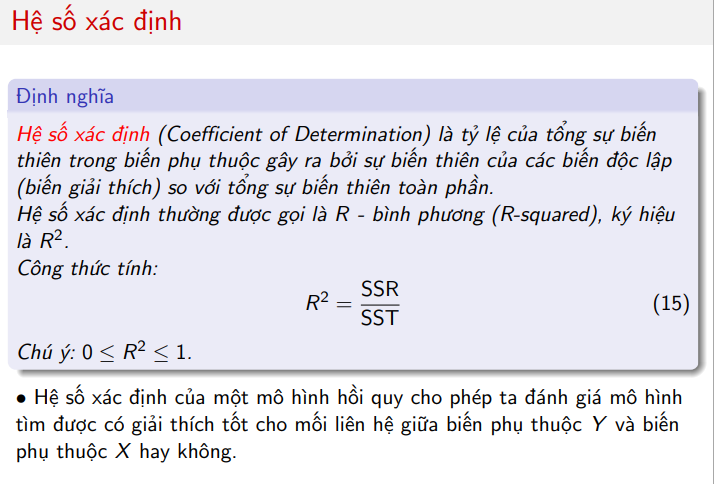
Hệ số tương quan có giá trị từ -1.0 đến 1.0. Kết quả được tính ra lớn hơn 1.0 hoặc nhỏ hơn -1 có nghĩa là có lỗi trong phép đo tương quan.

- Hệ số tương quan có giá trị âm cho thấy hai biến có mối quan hệ nghịch biến hoặc tương quan âm (nghịch biến tuyệt đối khi giá trị bằng -1)

- Hệ số tương quan có giá trị dương cho thấy mối quan hệ đồng biến hoặc tương quan dương (đồng biến tuyệt đối khi giá trị bằng 1)

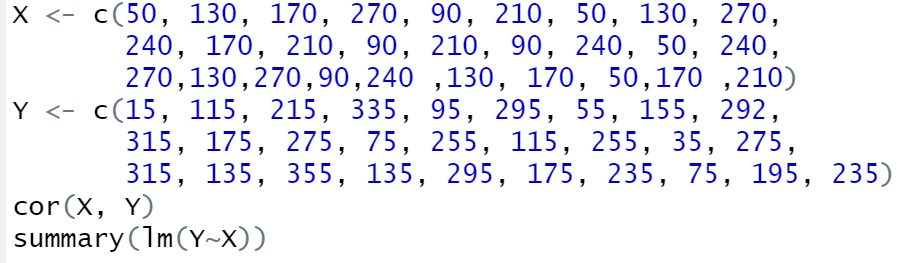
- Tương quan bằng 0 cho hai biến độc lập với nhau.



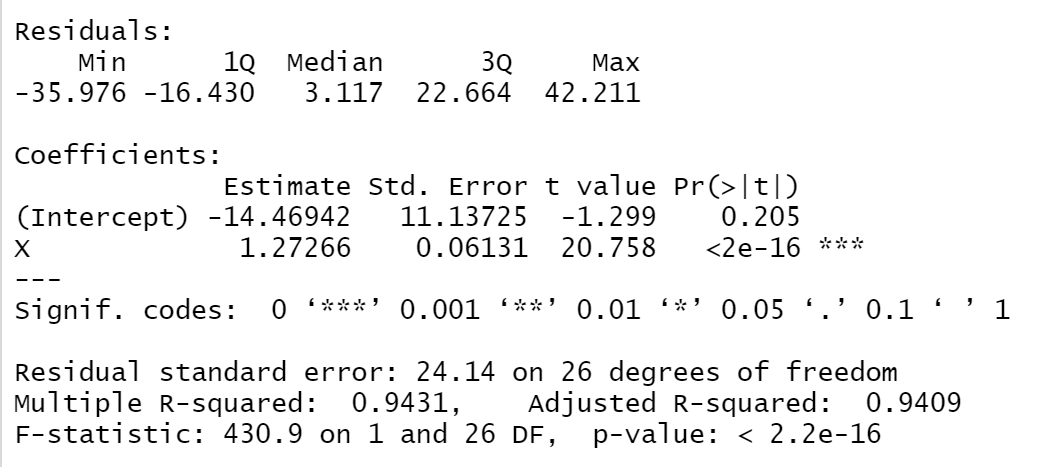


Giá trị R bình phương dao động từ 0 đến 1. R bình phương càng gần 1 thì mô hình đã xây dựng càng phù hợp với bộ dữ liệu dùng chạy hồi quy. R bình phương càng gần 0 thì mô hình đã xây dựng càng kém phù hợp với bộ dữ liệu dùng chạy hồi quy.

**Dùng phần mềm R tính giá trị:**







Hệ số tương quan của Y đối với X là 0.9711311

Hệ số xác định của tập số liệu là: 0.9431

Từ đó kết luận X và Y có mối quan hệ tuyến tính mạnh.

**Bài 3: Một nhà nghiên cứu muốn khảo sát thời gian phản ứng của nam giới và nữ giới đối với các loại tín hiệu khác nhau. Các đối tượng tham gia thí nghiệm được yêu cầu nhấn nút ENTER trên bàn phím máy tính ngay khi nhận biết tín hiệu thời gian (đo bằng giây ) giữa lúc tín hiệu phát ra và lúc đối tượng nhận biết được ghi lại. Sau đây là kết quả trên 15 nam và 15 nữ**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Âm thanh | Ánh sáng | Xung |
| Nam | 10.0  7.2  6.8  6.0  5.0 | 6.0  3.7  5.1  4.0  3.2 | 9.1  5.8  6.0  4.0  5.1 |
| Nữ | 10.5  8.8  9.2  8.1  13.4 | 6.6  4.9  2.5  4.2  1.8 | 7.3  6.1  5.2  2.5  3.9 |

**Với mức =5% . Có sự tương tác giữa giới tính và tín hiệu hay không?**

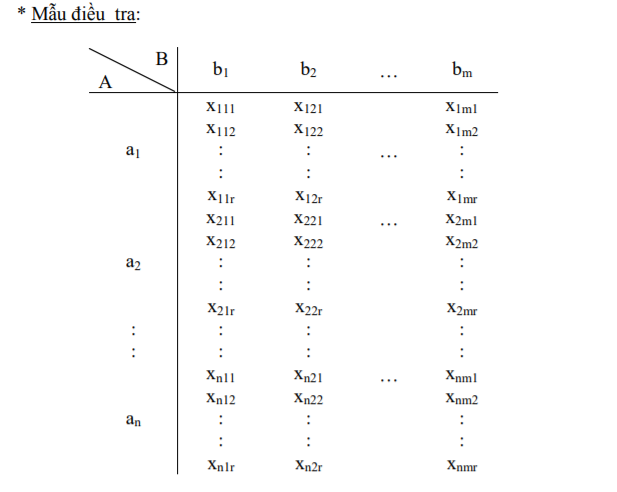
**Bài làm**

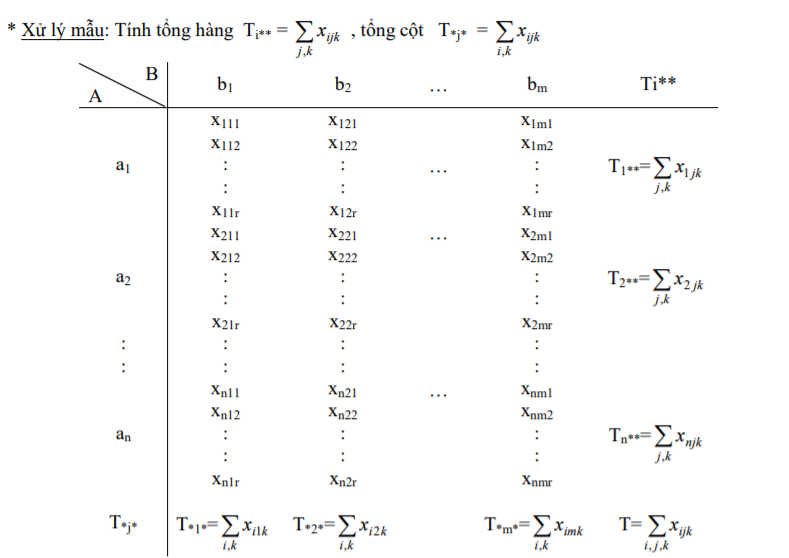
**Dạng bài** :phân tích phương sai hai yếu tố (có lặp)

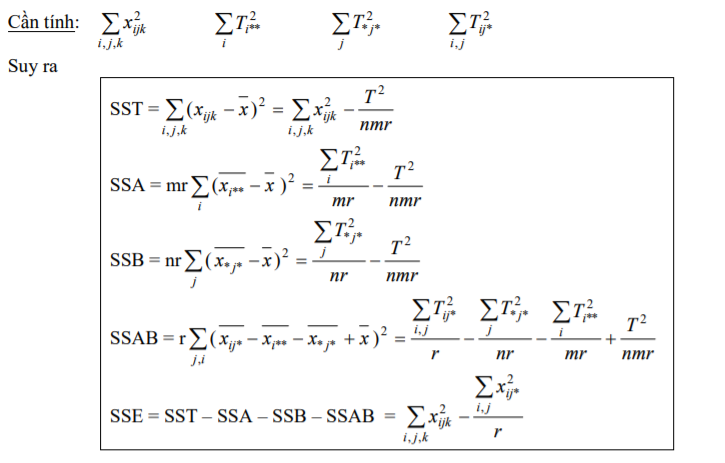
**Phương pháp giải**:Anova

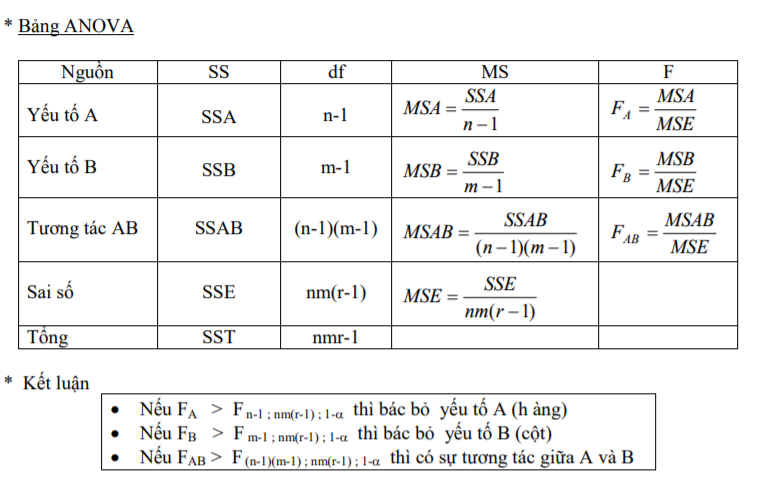
**Công cụ giải** :hộp thoại anova :two factor with replication

**Cơ sở lý thuyết**:Tương tự như bài toán phân tích Phuong sai 2 yếu tố không lặp, chỉ khác mỗi mức i mức ((ai , bj) đều có sự lặp lại r lần thí nghiệm và ta cần khảo sát thêm sự tương tác (interaction term) FAB giữa 2 nhân tố A và B





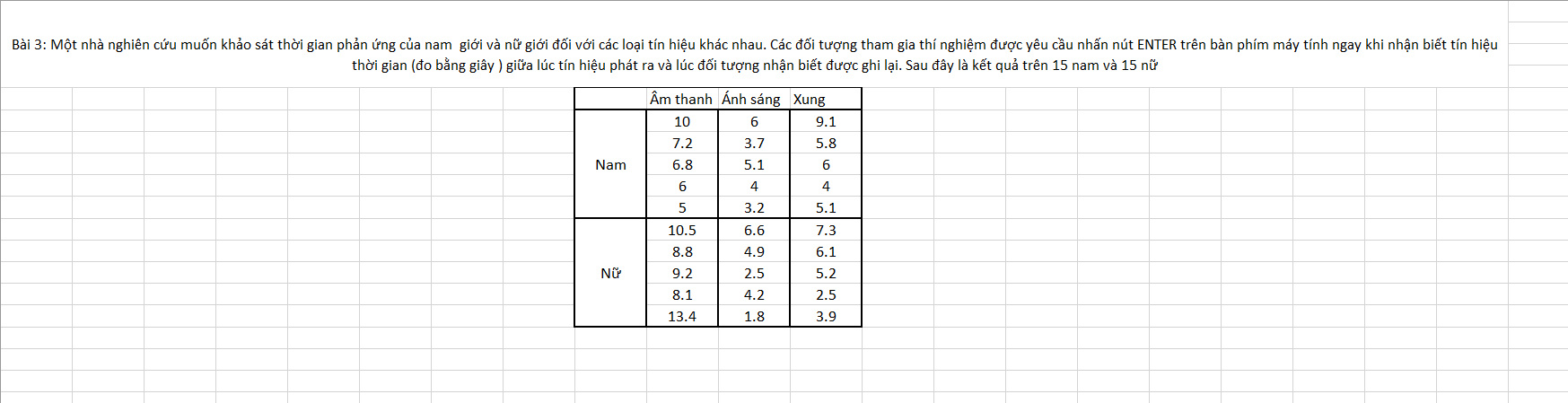




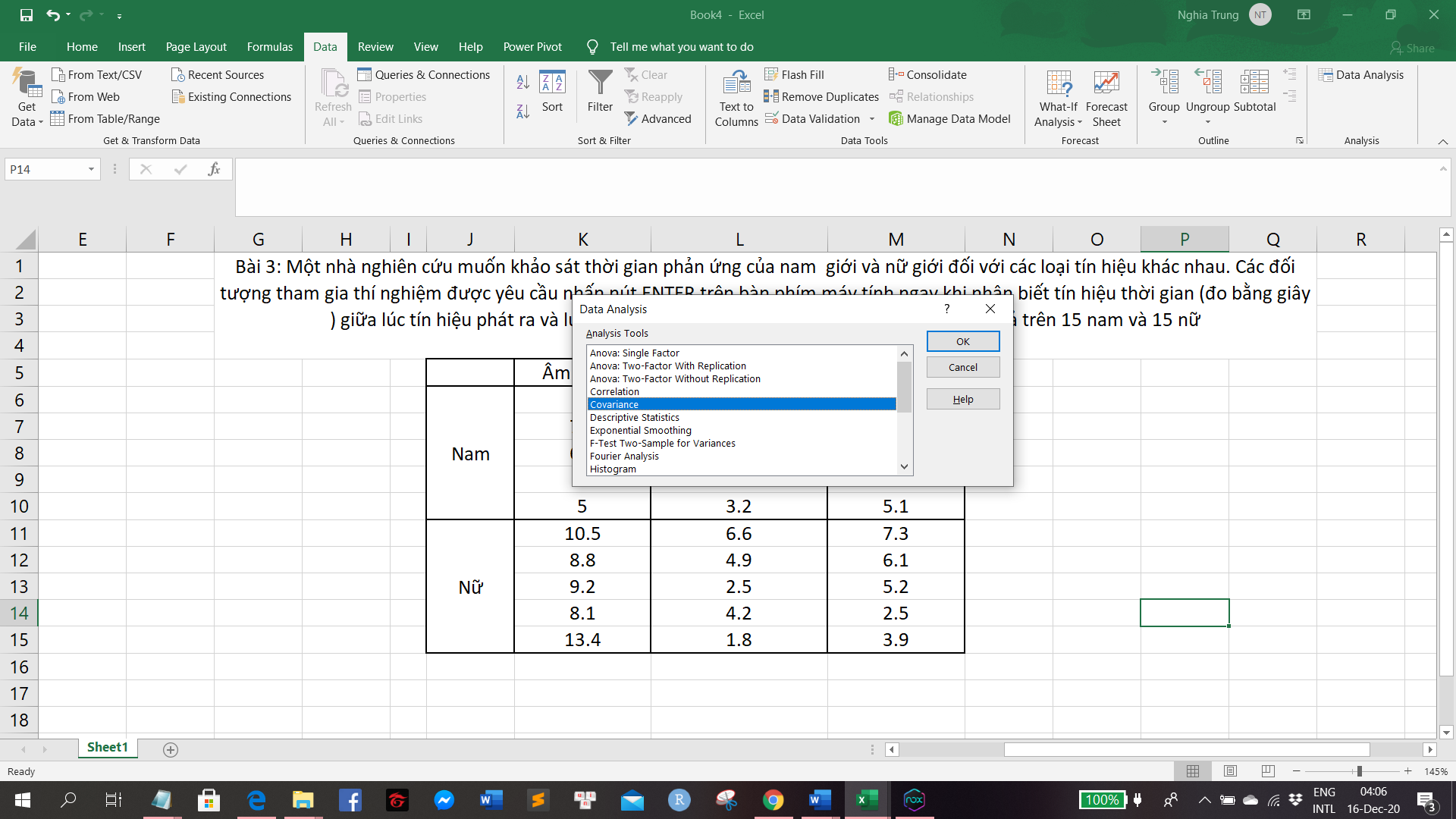
**Dùng excel để tính toán:**

Giả thuyết **H0**=Giới tính có tướng tác với tín hiệu

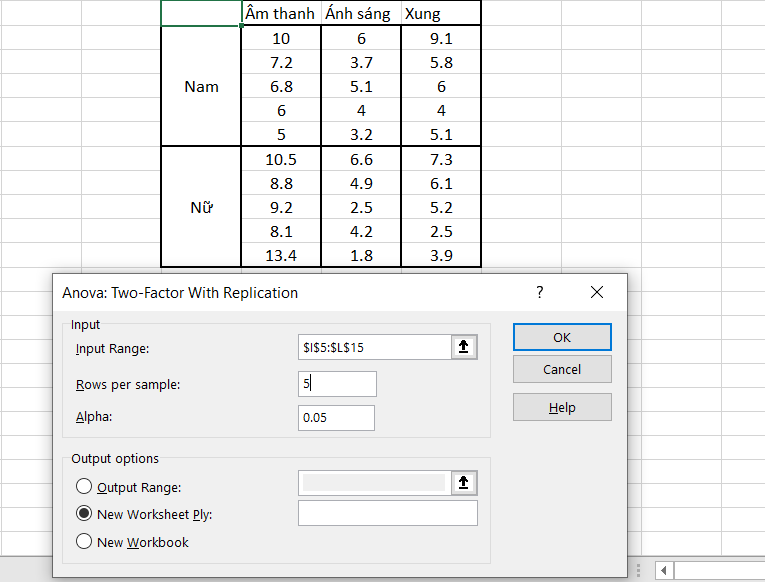
Nhập giá trị vào bảng :



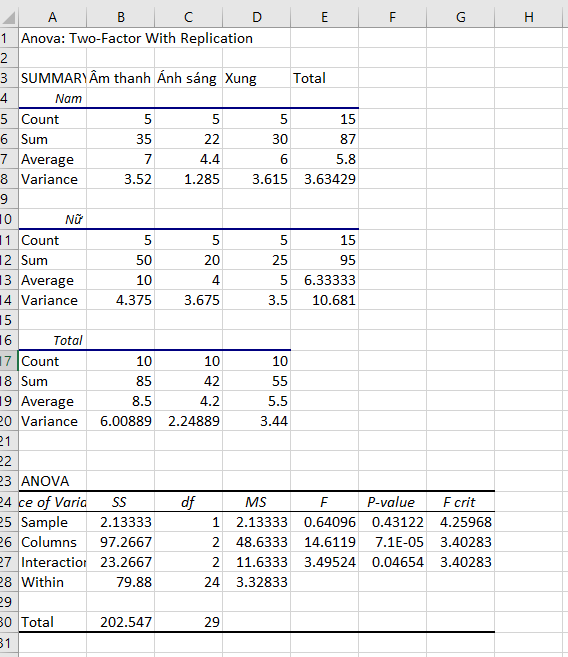
Chọn data->data Analysis, hộp thoại analysis hiện lên



Nhập input range,row per samble như hình, =0.05



Màn hình sẽ hiện lên kết quả



**Kết quả và biện luận:**

FR=0,64096<Fcrit=4.25968=> chấp nhận giả thiết H0(giới tính)

Fc=14.6119>Fcrit=3.40283=> Bác bỏ giả thiết H0(Tín hiệu)

Vậy chỉ có Giới tính ảnh hưởng đến thời gian phản ứng

**Bài 4: Để nghiên cứu xem quy mô của một công ty có ảnh hưởng đến hiệu quả quảng cáo đối với khách hàng hay không người ta đã điều tra ý kiến của 356 khách hàng và thu kết quả sau:**

|  |  |  |  |
| --- | --- | --- | --- |
| Quy mô công ty | Hiệu quả quảng cáo | |  |
| Mạnh | Vừa phải | Yếu |
| Nhỏ | 20 | 52 | 32 |
| Vừa | 53 | 47 | 28 |
| Lớn | 67 | 32 | 25 |

**Với mức ý nghĩa α= 0,1 có thể cho rằng quy mô của công ty có ảnh hưởng đến hiệu quả của quảng cáo đối với khách hàng hay không?**

**Bài làm**

**1)Dạng bài:** phân tích phương sai hai yếu tố ( không lặp)

# 2)Cơ sở lý thuyết

Sự phân tích này nhằm đánh giá sự ảnh hưởng của hai yếu tố trên các giá trị quan sát Yij ( i=1,2 … r: yếu tố A; j=1,2…c: yếu tố B). Trong đó, mô hình phân tích phương sai có lặp chỉ là sự lặp lại k lần thí nghiệm, mỗi hàng sẽ hiển thị một bản sao của dữ liệu và đầu ra sẽ thêm một đại lượng tương tác F1 giữa A và B

# Mô hình

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Yếu tố A** | **Yếu tố B** | | | | **Tổng cộng** | **Trung bình** |
| 1 | 2 | … | c |
|  | Y11 | Y12 | … | Y1c | Y1 | Y¯1 |
| Y21 | Y22 | … | Y2c | Y2 | Y¯2 |
| … | … | … | … | … | … |
| Yr1 | Yr2 | … | Yr | Yr | Y¯r |
| Tổng cộng trung bình | T1  Y¯ 1 | T2  Y¯2 | …  … | Tc  Y¯c | T…  Y¯… |  |

**Bảng ANOVA**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Nguồn sai số** | **Bậc tự do** | **Tổng số bình phương** | **Bình phương trung bình** | **Giá trị thống kê** |
| Yếu tố A (hàng) | (r - 1) |  |  |  |
| Yếu tố B (cột)  Sai số | (c - 1)  (r-1)(c-1) | SSE= SST-(SSF+SSB) |  |  |
| Tổng cộng | (rc-1) |  |  |  |

Trắc nghiệm

+ Giả thiết: µ1 = µ2 = … = µk↔ “Các giá trị trung bình bằn nhau”. H1: µi G µj ↔ “Ít nhất có hai giá trị trung bình khác nhau”.

+ Giá trị thống kê:

và

+ Biện luận:

Nếu Fr < Fα [r-1, (r-1)(c-1)]}=> chấp nhận H0 (yếu tố A)

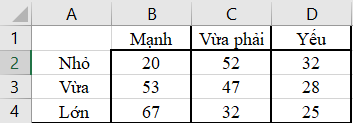
Nếu Fc < Fα [r-1, (r-1)(c-1)]}=> chấp nhận H0 (yếu tố B).

**3)Giải bài toán trên excel:**

**Giả thiết H0**: quy mô của công ty có ảnh hưởng đến hiệu quả của quảng cáo đối với khách hàng

**Đối giả thiết H0**: ngành nghề không ảnh hưởng đến thu thập

Nhập dữ liệu vào bảng tính:



Áp dụng: “Anova Two-Factor Without Replication”

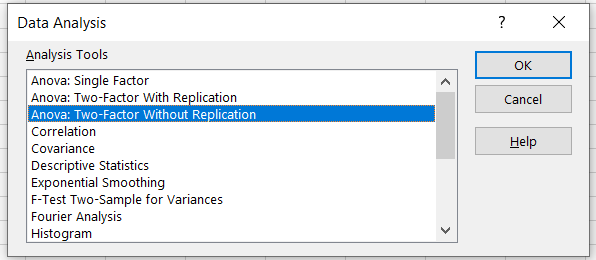
1. Nhập lần lượt đơn lệnh Tools và Data Analysis
2. Chọn chương trình Anova: Two- trong hộp thoại Data Analysis rồi nhấp nút OK
3. Trong hộp thoại Anova: Two – Factor without Replication,lần lượt ấn định các chi tiết:

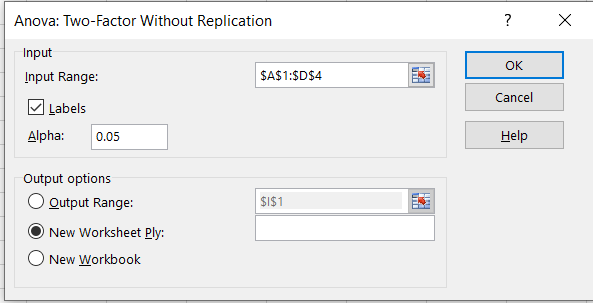
-Phạm vi đầu vào (Input Range)

-Nhân dữ liệu (Labels in First Row/Column)

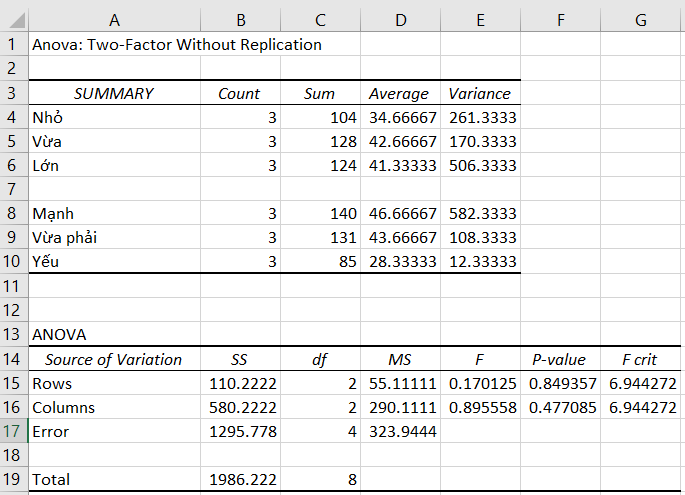
-Ngưỡng tin cậy: Alpha=0,1

-Phạm vi đầu ra (output Range)





Hộp thoại Anova: Two-Factor Without Replication



Kết quả và biện luận:

FR=0,170125<Fcrit=6,944272=> chấp nhận giả thiết H0(yếu tố hiệu quả quảng cáo)

Fc=0,895558< Fcrit=6,944272=> chấp nhận giả thiết H0(yếu tố quy mô công ty)

Vậy quy mô của công ty không ảnh hưởng đến hiệu quả của quảng cáo đối với khách hàng

**Bài 5: Với mức ý nghĩa 1%, Theo dõi số học sinh đến lớp muộn của năm trường PTTH vào các ngày khác nhau trong tuần người ta thu được số liệu về số lượng học sinh trung bình đến lớp của các trường đó váo một ngày tiêu biểu trong tuần như sau:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Ngày trong tuần | Trường PTTH | | | |
| A | B | C | D |
| Thứ hai | 5 | 4 | 5 | 7 |
| Thứ tư | 4 | 5 | 3 | 2 |
| Thứ sáu | 4 | 3 | 4 | 5 |
| Thứ bảy | 4 | 4 | 3 | 2 |

**Bạn có nhận xét gì về số lượng đến lớp đến lớp muộn của các trường Có sự khác biệt gì về số lượng học sinh đến lớp muộn vào các ngày khác nhau trong tuần?**

**Bài làm**

1. **Dạng bài:** phân tích phương sai hai yếu tố ( không lặp)

# Cơ sở lý thuyết

Sự phân tích này nhằm đánh giá sự ảnh hưởng của hai yếu tố trên các giá trị quan sát Yij ( i=1,2 … r: yếu tố A; j=1,2…c: yếu tố B). Trong đó, mô hình phân tích phương sai có lặp chỉ là sự lặp lại k lần thí nghiệm, mỗi hàng sẽ hiển thị một bản sao của dữ liệu và đầu ra sẽ thêm một đại lượng tương tác F1 giữa A và B

# Mô hình

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Yếu tố A** | **Yếu tố B** | | | | **Tổng cộng** | **Trung bình** |
| 1 | 2 | … | c |
|  | Y11 | Y12 | … | Y1c | Y1 | Y¯1 |
| Y21 | Y22 | … | Y2c | Y2 | Y¯2 |
| … | … | … | … | … | … |
| Yr1 | Yr2 | … | Yr | Yr | Y¯r |
| Tổng cộng trung bình | T1  Y¯ 1 | T2  Y¯2 | …  … | Tc  Y¯c | T…  Y¯… |  |

**Bảng ANOVA**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Nguồn sai số** | **Bậc tự do** | **Tổng số bình phương** | **Bình phương trung bình** | **Giá trị thống kê** |
| Yếu tố A (hàng) | (r - 1) |  |  |  |
| Yếu tố B (cột)  Sai số | (c - 1)  (r-1)(c-1) | SSE= SST-(SSF+SSB) |  |  |
| Tổng cộng | (rc-1) |  |  |  |

Trắc nghiệm

+ Giả thiết: µ1 = µ2 = … = µk↔ “Các giá trị trung bình bằn nhau”. H1: µi G µj ↔ “Ít nhất có hai giá trị trung bình khác nhau”.

+ Giá trị thống kê:

và

+ Biện luận:

Nếu Fr < Fα [r-1, (r-1)(c-1)]}=> chấp nhận H0 (yếu tố A)

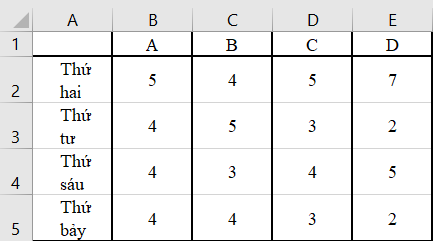
Nếu Fc < Fα [r-1, (r-1)(c-1)]}=> chấp nhận H0 (yếu tố B).

1. **Giải bài toán trên excel:**

**Giả thiết H0**: ngày và trường PTTH ảnh hưởng tới số lượng học sinh đi muộn

**Đối giả thiết H0**: ngày và trường PTTH ảnh hưởng tới số lượng học sinh đi muộn

Nhập dữ liệu vào bảng tính:



Áp dụng: “Anova Two-Factor Without Replication”

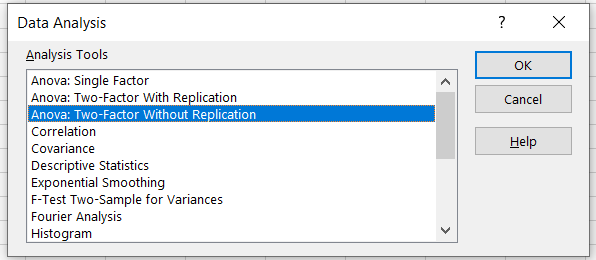
1. Nhập lần lượt đơn lệnh Tools và Data Analysis
2. Chọn chương trình Anova: Two- trong hộp thoại Data Analysis rồi nhấp nút OK
3. Trong hộp thoại Anova: Two – Factor without Replication,lần lượt ấn định các chi tiết:

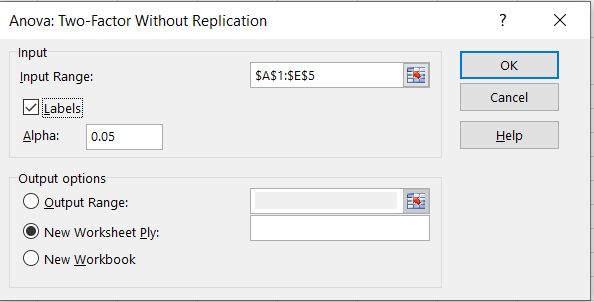
-Phạm vi đầu vào (Input Range)

-Nhân dữ liệu (Labels in First Row/Column)

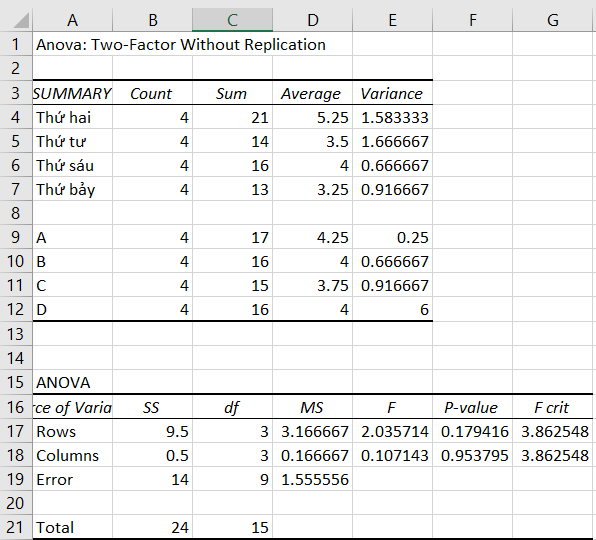
-Ngưỡng tin cậy: Alpha=0,1

-Phạm vi đầu ra (output Range)





Hộp thoại Anova: Two-Factor Without Replication



**Kết quả và biện luận:**

FR=2,035714<Fcrit=3,862548=> chấp nhận giả thiết H0(yếu tố ngày)

Fc=0,107143< Fcrit=6,944272=> chấp nhận giả thiết H0(yếu tố trường PTTH)

Vậy cả yếu tố ngày và yếu tố trường PTTH đều không ảnh hưởng tới số lượng học sinh đi muộn.