**t – test, kiểm định t [i]**

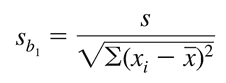
Đặt các giả thuyết như sau, tùy theo mục đích kiểm định:

H0: β1 = 0                                        H0: β1 ≤ 0                                   H0: β1≥ 0

H1: β1 ≠ 0                                        H1: β1 > 0                                   H1: β1 < 0

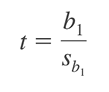
Tuy nhiên để kiểm định cho giá trị β1cho tổng thể phải dựa vào giá trị b1tìm được từ bộ dữ liệu mẫu, do đó phải xem xét phân phối mẫu của b1 có đủ điều kiện để thực hiện kiểm định hay không. Phân phối mẫu của b1 có các tính chất như sau:

* Giá trị kỳ vọng E(b1) = β1
* Độ lệch chuẩn:



* Phân phối của mẫu thuộc dạng phân phối chuẩn

Như vậy sau cùng công thức tổng quát của giá trị kiểm định t có dạng như sau:



Nguyên tắc bác bỏ H0:

Với kiểm định 2 phía: H0: β1 = 0 được bác bỏ khi trị tuyệt đối của t lớn hơn t tra bảng (tα/2, n-2)

Với kiểm định bên phải: H0: β1≤ 0 được bác bỏ khi giá trị t dương lớn hơn giá trị dương của t tra bảng (tα, n-2)

Với kiểm định bên trái: H0: β1 ≥ 0 được bác bỏ khi giá trị t < giá trị âm của t tra bảng (tα, n-2)

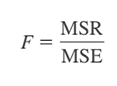
Nếu xét trên giá trị p-value, bác bỏ H0 khi p-value < α. (p-value là mức ý nghĩa nhỏ nhất mà tại đó H0 bị bác bỏ).

**F** – **test, kiểm định F [i]**

Tương tự như kiểm định t, kiểm định F, dựa trên phân phối F – một dạng phân phối xác suất, cũng sẽ xác định bác bỏ hay không bác bỏ giả thuyết H0: β1 = 0. Tuy nhiên khác với kiểm định t, kiểm định F có thể kết luận có hay không có mối quan hệ tuyến tính giữa một biến phụ thuộc y và nhiều biến độc lập x một cách tổng quát**.** Trong bài viết này em chỉ xét trường hợp một biến x.

Kiểm định F ngoài sử dụng MSE, ước lượng phương sai của sai số, mà còn sử dụng thêm **MSR** **(mean square regression)**được tính bằng cách lấy SSR chia cho bậc tự do của phương trình hồi quy, là số biến độc lập có trong phương trình, ở đây em chỉ xét 1 biến x nên bậc tự so bằng 1. MSR chính là giá trị ước lượng phương sai σ2dựa trên SSR

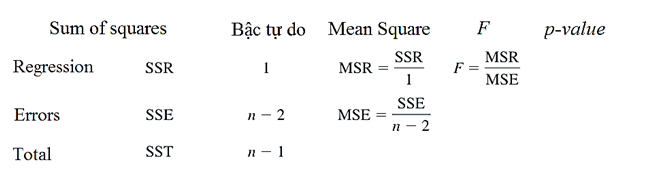
Lưu ý, kiểm định F cho hệ số hồi quy β1chủ yếu là **kiểm định 1 phía**, sử dụng mức ý nghĩa α cho trước. Công thức tổng quát của giá trị kiểm định F là:



Cơ sở bác bỏ H0:

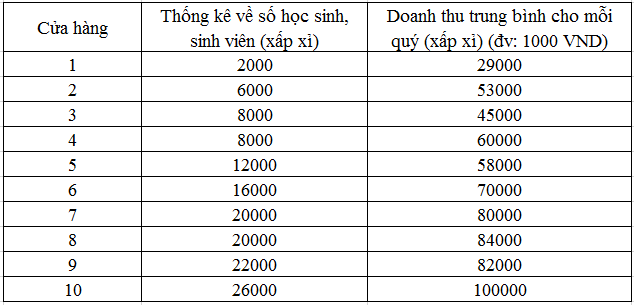
* p – value nhỏ hơn mức ý nghĩa α
* giá trị F tính được phải lớn hơn giá trị F tra bảng phân phối F (với bậc tự do thứ nhất là 1 ở hàng trên cùng, và bậc tự do thứ hai là n – 2 ở cột ngoài cùng, α ở cột thứ 2 tính từ cột ngoài cùng)

Bảng ANOVA (phân tích phương sai) tổng quát như sau:



Ví dụ:

Giá sử một chuỗi cửa hàng gà rán có 10 cửa hàng nằm tại các quận khác nhau trên thành phố Hồ Chí Minh, tại từng khu vực của từng cửa hàng sẽ có số lượng học sinh, sinh viên sinh sống và học tập, chuỗi cửa hàng này muốn biết rằng doanh thu của từng cửa hàng có mối liên hệ nào với số lượng học sinh, sinh viên này không (dựa trên dữ liệu doanh thu trung bình theo quý của mỗi cửa hàng, và dữ liệu thống kê về số lượng học sinh, sinh viên)

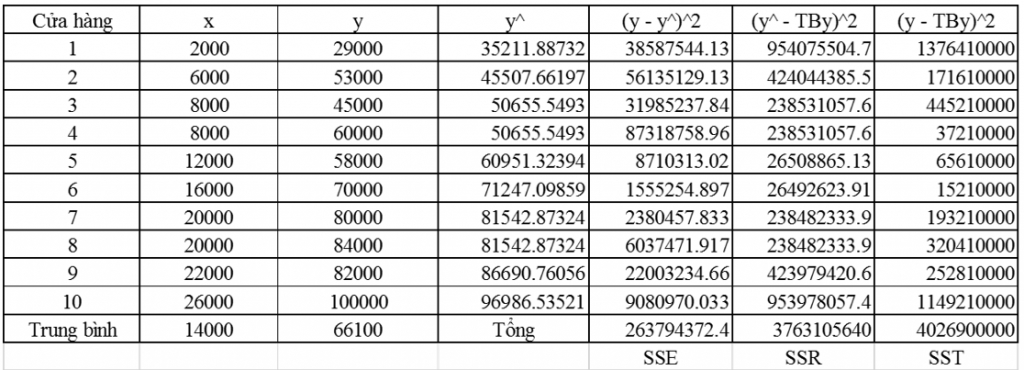


Gọi x là số lượng HS, SV, y là doanh thu, là biến mục tiêu dự báo, em sẽ sử dụng công thức tính bo và b1 để lập phương trình, chúng ta tính được trung bình x: TBx = 14000, trung bình của y: TBy = 66100 Significance F là p-value, các bạn cũng có thể sử dụng hàm F.DIST trong excel để tính.

Với b1= 2.57. Với số lượng HS, SV tăng 1000 thì doanh thu của 1 cửa hàng sẽ tăng 2570000 VND, và b0 = 30064.

Phương trình hồi quy tuyến tính đơn giản: Y^ = 2.57X + 30064

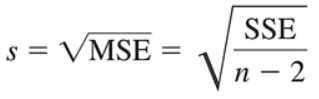
Thay từng giá trị x vào phương trình mới tìm được để tính giá trị Y^ dự báo, sau đó tính SSE, SSR dựa trên Y^.



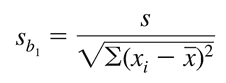
Đặt giả thuyết:     H0: β1 = 0 : Không có mối quan hệ giữa số HS, SV và doanh thu  mỗi cửa hàng

H1: β1 ≠ 0  Có mối quan hệ giữa số HS, SV và doanh thu  mỗi cửa hàng.

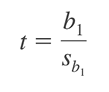
Kiểm định t:



S = √(SSE/n – 2) = √(263794372.4/8) = 5742.3



Sb = 0.24



t = b/Sb = 2.57/0.24 = 10.7

t tra bảng với mức ý nghĩa α = 5%, bậc tự do là 8, tα/2,8 = 2.306

Như vậy với t > tα/2,8 bác bỏ giả thuyết H0, tức có mối quan hệ giữa 2 biến số lượng học sinh sinh viên trong khu vực và doanh thu của mỗi nhà hàng trong khu vực ấy.

Tương tự giá trị p-value = 0.000005 (có thể dùng hàm T.DIST trong excel để tìm p-value) tính được nhỏ hơn rất nhiều so với mức ý nghĩa, nên em bác bỏ H0.

Còn với kiểm định F:

MSE = SSE/(n – 2) = 263794372.4/8 = 32974295.77

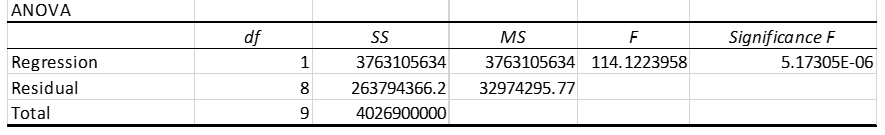
MSR = SSR/1 = 3763105640

F = MSR/MSE = 114

F tra bảng với mức ý nghĩa 5%, Fα,1,8 = 5.32

F > Fα,1,8, bác bỏ H0 tức có mối quan hệ giữa 2 biến số lượng học sinh, sinh viên trong khu vực và doanh thu của mỗi nhà hàng trong khu vực ấy.

Bảng ANOVA có được:



Nguồn: [PHƯƠNG PHÁP KIỂM ĐỊNH TRONG TƯƠNG QUAN VÀ HỒI QUY TUYẾN TÍNH ĐƠN BIẾN - Big Data Uni](https://bigdatauni.com/tin-tuc/kiem-dinh-trong-tuong-quan-va-hoi-quy-don-bien.html) [i]