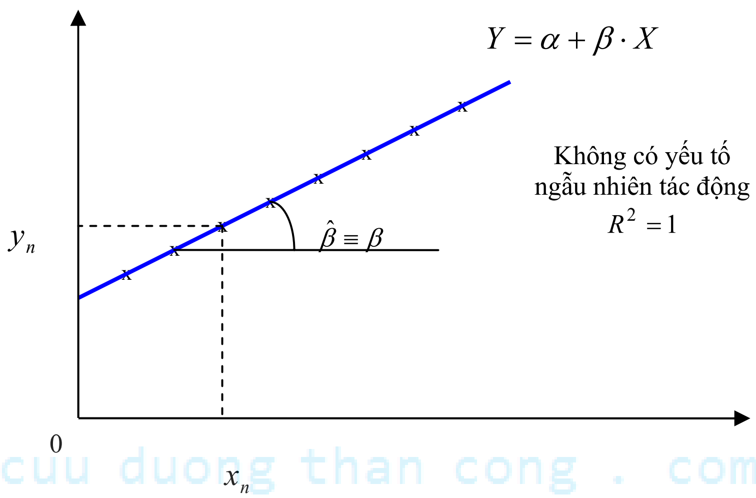
CHƯƠNG 3: HỒI QUI ĐƠN BIẾN

3.1 Bản chất thống kê của mô hình hồi quy đơn biến

Phương pháp ước lượng *LS*, về thực chất, chỉ là vẽ một đường hồi quy đi xuyên qua "đám bụi" dữ liệu, sao cho tổng bình phương các phần dư [hay sai số] ESS là nhỏ nhất. Nhưng việc đo lường mang tính thuần túy đại số đó chưa có gì bảo đảm chắc chắn rằng nó sẽ cho ra những ước lượng $\hat{\alpha}, \hat{\beta}$ tốt nhất của các tham số tổng thể α, β theo những tiêu chuẩn xác định về mặt thống kê. Để có thể những đánh giá cụ thể hơn về độ tốt của ước lượng, chúng ta cần xem xét sâu hơn **bản chất thống kê** của mô hình hồi quy.

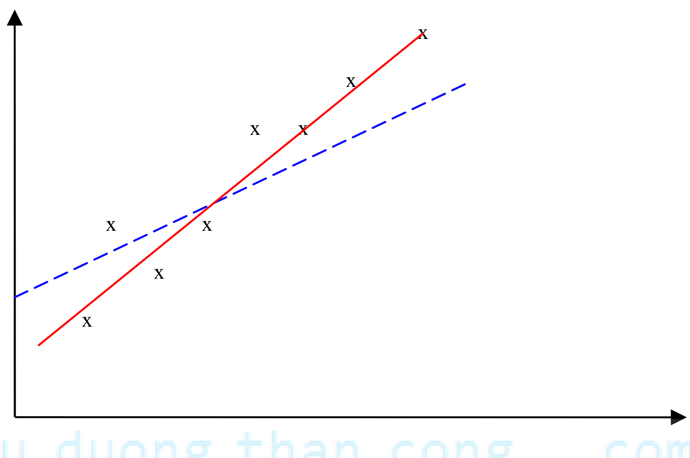
Để dễ hình dung, chúng ta bắt đầu bằng sự giả định phi thực rằng, quan hệ giữa biến X và Y [chẳng hạn như giữa thu nhập và tiêu dùng] chỉ tuân theo quy luật xác định, và hoàn toàn không bị chi phối bởi các yếu tố ngẫu nhiên. Khi đó, các quan sát $\{x_n, y_n\}_{n=1}^{N}$ sẽ nằm gọn trên một đường thẳng mô tả xu thế thực của tổng thể:



Đồ thị 3.1a: quy luật xác định giữa *X* và *Y*.

Khi đó, việc ước lượng trở nên tầm thường, vì ta luôn có $\hat{\alpha} = \alpha, \hat{\beta} = \beta$, và $R^2 = 1$.

Bây giờ, chúng ta cho phép các yếu tố ngẫu nhiên tác động lên quan hệ giữa X, Y. Như đã nêu, các nhân tố này khiến cho các quan sát $\{x_n, y_n\}_{n=1}^{N}$ bị lệch một cách ngẫu nhiên khỏi đường xu thế tổng thể. Vì vậy, thay vì nhìn thấy một đường xu thẳng tuyến tính như trên hình 3.1a, ta chỉ nhìn thấy một đám bụi dữ liệu bám xung quanh một xu thế nào đó mà ta muốn ước lượng.



Đồ thị 3.1b: Quan hệ giữa X và Y bị nhiễu bởi các yếu tố ngẫu nhiên

Trên Đồ thị 3.1b, ta thấy các điểm quan sát $\{x_n, y_n\}_{n=1}^{N}$, trước đây nằm trên cùng một đường thẳng trên hình 3.1a, nay bị "thổi bay" lên thành một "đám bụi" dữ liệu, mà việc "chụp ảnh" chúng [tức là đi thu thập dữ liệu], rồi vẽ một đường hồi quy chạy xuyên qua chúng sẽ không nhất thiết là trùng với quy luật tổng thể (mô tả bởi gạch chấm). Điều này gợi ý rằng mỗi ước lượng $\hat{\beta}$ chịu sự quy định bởi tham số tổng thể β, nhưng bị lái đi bởi các biến ngẫu nhiên. [Tương tự, ta có thể nói như vậy về $\hat{\alpha}$]. Vì vậy, $\hat{\beta}$ cũng là một biến ngẫu nhiên. Vấn đề đặt ra là, về trung bình mà nói [tức là sau rất nhiều lần chụp ảnh các đám bụi dữ liệu], liệu ước lượng $\hat{\beta}$ có thể hiện đúng β hay không? Và liệu phương pháp ước lượng bình phương cực tiểu có là hiệu quả nhất hay không?

Về mặt toán học, phương pháp bình phương cực tiểu cho ta ước lượng sau:

$$\hat{\beta} = \frac{S_{XY}}{S_{XX}} = \frac{\sum (x_n - \bar{x})(y_n - \bar{y})}{S_{XX}} \tag{3.1}$$

Hay cũng vậy,

$$\hat{\beta} = \frac{\sum (x_n - \bar{x}) y_n}{S_{XX}} \tag{3.2}$$

[điều này là do $\sum_n (x_n - \bar{x})\bar{y} = 0$, như đã chỉ ra ở chương 1, phần ôn tập].

Trong (3.2), ta đặt $c_n = \dfrac{(x_n - \bar{x})}{S_{XX}}$, và nhận xét rằng, tham số đó chỉ phụ thuộc vào các quan sát $\{x_n\}_{n=1}^{N}$. Do vậy, nó không chịu ảnh hưởng bởi các yếu tố ngẫu nhiên. Khi đó, công thức (3.2) có thể viết lại như sau:

$$\begin{aligned}\hat{\beta} &= \sum_n c_n y_n \\ &= \sum_n c_n [\alpha + \beta x_n + \varepsilon_n] \\ &= \alpha \sum c_n + \beta \sum c_n x_n + \sum c_n \varepsilon_n\end{aligned}$$

Chúng ta có thể dễ dàng chỉ ra rằng, $\sum_n c_n = 0$ và $\sum_n c_n x_n = 1$. Và do vậy:

$$\hat{\beta} = \beta + \sum c_n \varepsilon_n \tag{3.3}$$

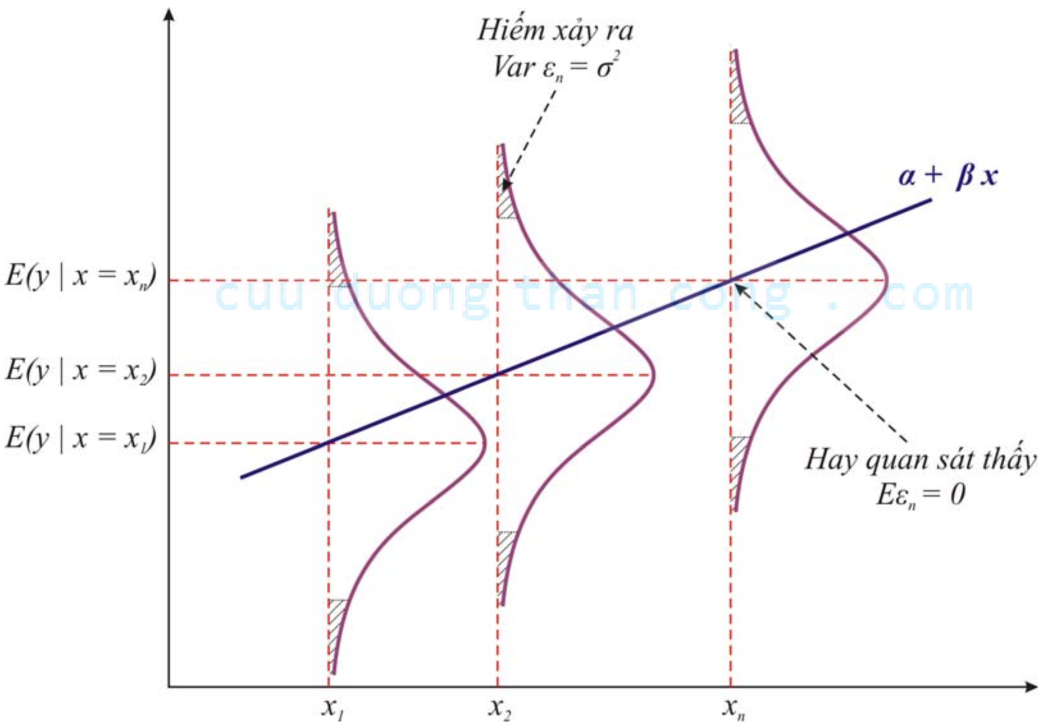
Phương trình (3.3) khẳng định nhận định trước đây về $\hat{\beta}$ là đúng: Ước lượng $\hat{\beta}$ bị ảnh hưởng bởi các yếu tố ngẫu nhiên ε_n, làm giá trị của nó không trùng khít với β tổng thể. Và vì vậy, $\hat{\beta}$ cũng là một biến ngẫu nhiên.

Chúng ta gọi $\hat{\beta}$ là **ước lượng không chệch**, nếu $E\hat{\beta} = \beta$. Và gọi nó là **ước lượng hiệu quả nhất**, nếu sai số ước lượng $Var\hat{\beta} = E(\hat{\beta} - \beta)^2$ là nhỏ nhất trong lớp tất cả các ước lượng tuyến tính, không chệch.

Để trả lời xem $\hat{\beta}$ có phải là ước lượng không chệch và hiệu quả hay không, ta phải xét đến bản chất thống kê của các quá trình ngẫu nhiên $\{\varepsilon_n\}_{n=1}^{N}$ [mà ta đã ví chúng như những "cơn gió", ngẫu nhiên "thổi bay" các quan sát khỏi đường xu thế xác định của tổng thể].

3.2 Các yếu tố ngẫu nhiên

Chúng ta hãy nêu lên giả định về các quá trình ngẫu nhiên. Hãy nhìn vào đồ thị sau:



Đồ thị 3.2: Quy luật phân phối xác suất của các nhiễu $\{\varepsilon_n\}_{n=1}^{N}$

Như đã nhận xét từ các Đồ thị 3.1a và 3.1b, khi không có các tác động ngẫu nhiên, hay $\varepsilon_n = 0$, các quan sát $\{x_n, y_n\}_{n=1}^{N}$ nằm ngay trên đường xu thế của tổng thể. Dưới tác động của yếu tố ngẫu nhiên, các quan sát $\{x_n, y_n\}_{n=1}^{N}$ nằm rải ra, nhưng "bám" xung quanh đường xu thế. Rất hiếm khi có quan sát bị "thổi" mạnh tới nỗi "bay" quá xa so với đường xu thế. Điều đó dẫn đến hai giả thiết sau:

A1 $E\varepsilon_n = 0$, với mọi *n*. [Bụi giữ liệu không thể bay quá xa, mà bám xung quanh đường tổng thể]

A2 $Var\varepsilon_n = \sigma^2$, với mọi *n*. [Độ tán xạ của đám bụi dữ liệu được thể hiện bởi độ lớn của σ^2].

Chúng ta cũng coi rằng quy luật tác động của "cơn gió", tức là phân bố xác suất của yếu tố ngẫu nhiên ε_n là như nhau (*identical)*, và theo phân bố chuẩn. Hơn nữa, các yếu tố ngẫu nhiên đó là độc lập (*independent*). Vì vậy, kết hợp với các giả thiết A1 và A2, ta có:

A3 $\varepsilon_n \overset{iid}{\sim} N(0, \sigma^2)$ với mọi *n*.

Cuối cùng, ta coi ta coi x_n là xác định trước. Từ giả thiết A1 và dạng mô hình $y_n = \alpha + \beta x_n + \varepsilon_n$, điều đó bao hàm rằng:

A4 $E(y_n \mid x_n) = \alpha + \beta x_n$, với mọi *n*.

Hai giả thiết cuối là quan trọng nhất. A3 tóm tắt mọi đặc trưng thống kê của nhiễu ngẫu nhiên, và A4 mô tả xu thế của tổng thể, mà ta ước lượng nó theo phương pháp bình phương cực tiểu.

3.3 Những đặc trưng thống kê của ước lượng bình phương cực tiểu

Bây giờ ta có thể nói đến tính tốt của các ước lượng theo các tiêu chuẩn thống kê .

Từ phương trình (3.3), ta đã có: $\hat{\beta} = \beta + \sum c_n \varepsilon_n$. Bây giờ, hãy áp dụng toán tử kỳ vọng vào hai vế của (3.3):

$$\begin{aligned} E\hat{\beta} &= E(\beta + \sum c_n \varepsilon_n) \\ &= \beta + \sum c_n E\varepsilon_n \\ &= \beta \end{aligned}$$

[ở đây, ta sử dụng giả thiết A1: $E\varepsilon_n = 0$].Ta đi đến kết luận rằng, ước lượng $\hat{\beta}$ là không chệch:

$$E\hat{\beta} = \beta \tag{3.4}$$

Tiếp theo, sử dụng công thức: $Var(x) = Var(x - Ex)$ [xem chương 1, phần ôn tập], và lưu ý (3.3), (3.4), ta có:

$$\begin{aligned} Var\hat{\beta} &= Var(\hat{\beta} - \beta) \\ &= Var(\sum c_n \varepsilon_n) \end{aligned}$$

Sử dụng giả thiết A3 về tính độc lập của các yếu tố ngẫu nhiên, cuối cùng ta nhận được:

$$\begin{aligned} Var\hat{\beta} &= \sum c_n^2 Var\varepsilon_n \\ &= \sigma^2 \sum c_n^2 \text{, hay} \end{aligned}$$

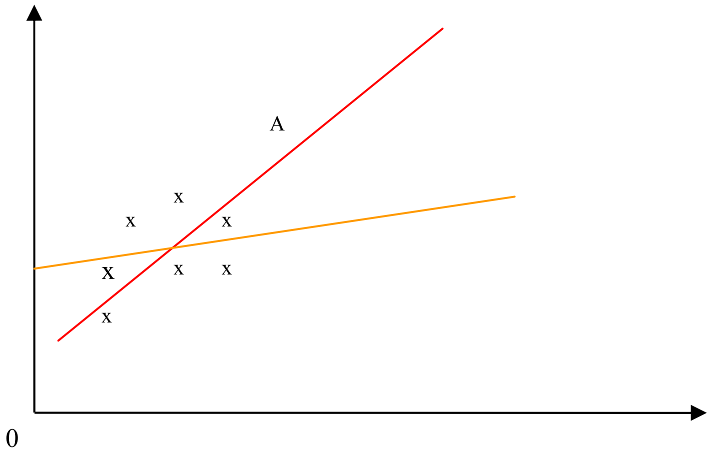
$$Var\hat{\beta} = \frac{\sigma^2}{S_{XX}} \tag{3.5}$$

(ở đây, ta sử dụng cái điều là $\sum c_n^2 = \sum \left[\frac{(x_n - \bar{x})}{S_{XX}} \right]^2 = \frac{S_{XX}}{S_{XX}^2} = \frac{1}{S_{XX}}$)

Định Lý Gauss - Markov: Phương pháp bình phương cực tiểu có sai số ước lượng, đo lường bởi $Var\hat{\beta}$, là nhỏ nhất trong lớp tất cả các ước lượng tuyến tính và không chệch.

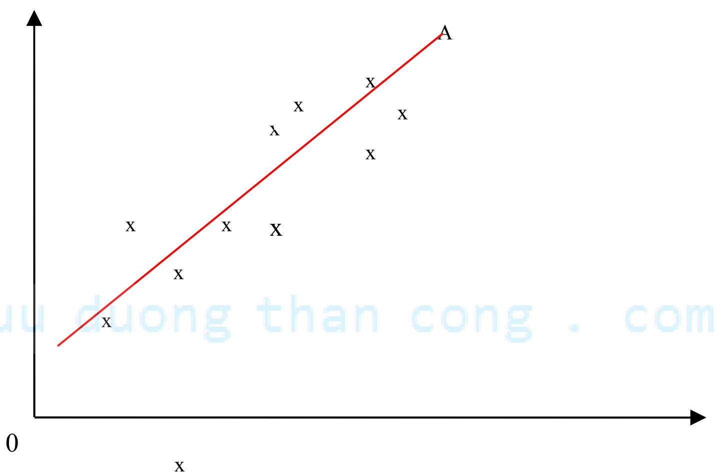
Định lý Gauss-Markov là hết sức quan trọng. Nó nêu lên rằng, chúng ta có được những tính chất rất tốt cho ước lượng theo phương pháp bình phương cực tiểu, mà chỉ đòi hỏi có trung bình bằng zero, tính độc lập, và phương sai giống nhau của các yếu tố ngẫu nhiên – tức là giả thiết A3.

Chúng ta cũng nên nói thêm là, phương trình (3.5) có một ý nghĩa thực tiễn đáng lưu ý. Nó nói rằng sai số của ước lượng $Var\hat{\beta}$ sẽ nhỏ đi, hay hiệu quả ước lượng sẽ tăng lên, nếu độ đa dạng của thông tin quan sát, đo bởi S_{XX}, tăng lên. Điều đó bao hàm rằng, khi làm nghiên cứu, ta không cứ nhất thiết phải tăng rất lớn số quan sát (*sample size*) *N*. Nếu giả thiết về tính tuyến tính của đường hồi quy là đúng, thì việc tăng độ đa dạng của thông tin quan sát, hay biên độ giao động của biến giải thích, $S_{XX} = \sum_n (x_n - \bar{x})^2$, sẽ làm cho ước lượng có độ chính xác cao hơn. Hãy xét các ví dụ sau:



Đồ thị 3.3a: Ước lượng có độ chính xác thấp, do S_{XX} nhỏ.

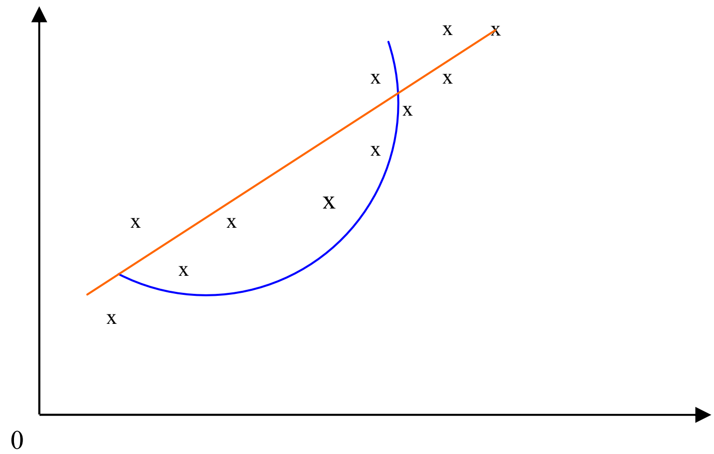
Trên Đồ thị 3.3a, giả sử ta có số quan sát *N* rất lớn, nhưng với biên độ giao động S_{XX} nhỏ. Khi đó, chỉ cần bỏ đi một quan sát như ứng với điểm A thôi, thì cũng đủ làm các hệ số ước lượng $\{\hat{\alpha}, \hat{\beta}\}$ thay đổi rất mạnh [từ đường mầu đỏ chuyển sang đường tô mầu da cam]. Điều đó chứng tỏ sai số ước lượng, đo bởi $Var\,\hat{\beta}$, là lớn. Ta sẽ xét kỹ hơn vấn đề này trong chương 7 về **đa cộng tuyến** (*multicollinearity*).



Đồ thị 3.3b: Ước lượng có độ chính xác cao hơn, ứng với S_{XX} lớn hơn.

Trên Đồ thị 3.3b, việc loại bỏ đi một vài quan sát, như điểm A, sẽ ít làm thay đổi các hệ số ước lượng. Kết quả ước lượng có độ ổn định cao hơn và chính xác hơn.

Tuy nhiên, những nhận xét trên chỉ đúng, khi giả thuyết **tuyến tính** của đường hồi quy là đúng. Đôi khi, giá trị rất lớn của S_{XX} lại hàm ý rằng giả thuyết tuyến tính là đáng nghi vấn:



Đồ thị 3.3c: Quy luật tổng thể không phải là tuyến tính (gây nên S_{XX} lớn)

Đồ thị 3.3c thể hiện rằng, việc hiểu sai về bản chất kinh tế đã gây nên việc áp dụng sai mô hình hồi quy tuyến tính. Những sai lầm kiểu như vậy dẫn đến yêu cầu phải kiểm định giả thuyết thống kê về tính **có ý nghĩa** của các tham số của mô hình. Đó là chủ đề của phần 3.4.2 của chương này. Việc sử dụng các dạng hàm khác nhau (*functional forms*) để mô tả quy luật chi phối các dữ liệu quan sát $\{x_n, y_n\}_{n=1}^{N}$ là một chủ đề khác nữa, mà nó cũng sẽ được đề cập trong chương 6.

3.4 Kiểm định giả thuyết thống kê

Để có màu sắc kinh tế, ta hãy xét vấn đề kiểm định thông qua một ví dụ cụ thể.

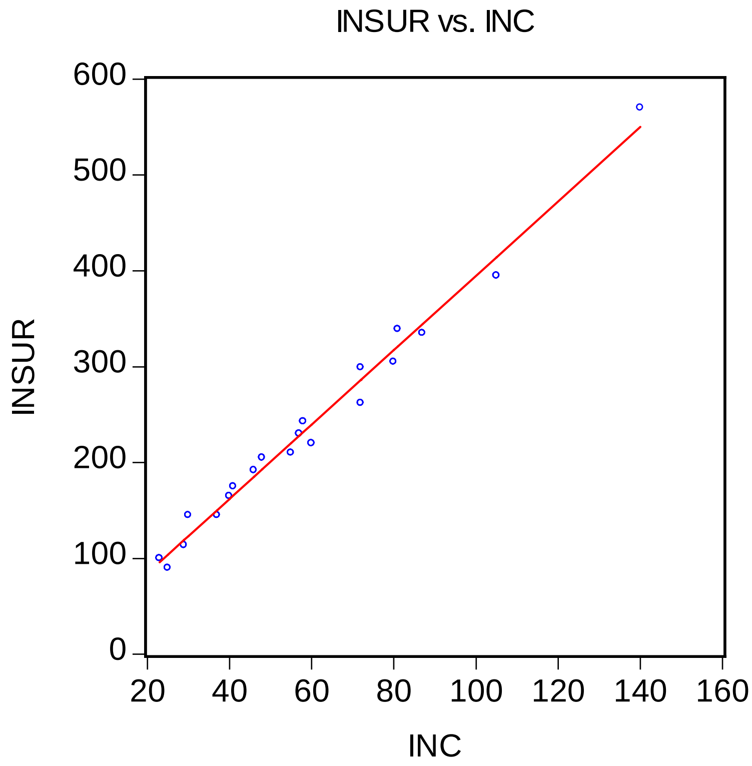
Ví dụ 3.5: Một công ty bảo hiểm ở Mỹ muốn kinh doanh bảo hiểm nhân thọ. Họ tiến hành nghiên cứu tiềm năng của thị trường sở tại. Lý luận kinh tế đã chỉ ra rằng, yêu cầu về mua bảo hiểm tăng lên cùng với khả năng xảy ra rủi ro, với quy mô về tổn thất tài chính khi rủi ro xảy ra, và với tâm lý ngại rủi ro của cá nhân. Họ nhận định rằng, gia đình càng giầu có nhờ kinh doanh, thì người chủ gia đình càng chịu nhiều stress. Tức là, những người lệ thuộc càng ngại rủi ro gây nên bởi stress cho người chủ gia đình, hơn là tại những gia đình có thu nhập thấp, ít tham dự vào kinh doanh. Vì vậy, ban nghiên cứu thị trường của công ty bảo hiểm này đề xuất mô hình sau:

$$INS = \alpha + \beta INC$$

Trong đó, *INS* là giá trị hợp đồng bảo hiểm, được trả cho bên mua bảo hiểm, nếu xảy ra rủi ro. Và *INC* là thu nhập. Cả hai biến lượng đều tính bằng nghìn dollars. Dữ liệu điều tra và kết quả ước lượng được ghi lại trong các bảng dưới đây

obs	INSUR	INC	obs	INSUR	INC
1	90	25	11	230	57
2	165	40	12	262	72
3	220	60	13	570	140
4	145	30	14	100	23
5	114	29	15	210	55
6	175	41	16	243	58
7	145	37	17	335	87
8	192	46	18	299	72
9	395	105	19	305	80
10	339	81	20	205	48

Bảng 3.1: Số liệu điều tra về nhu cầu mua bảo hiểm



Đồ thị 3.4: Nhu cầu mua bảo hiểm

Sử dụng eviews, chúng ta nhận được kết quả hồi quy dưới đây:

Dependent Variable: INSUR
Method: Least Squares
Date: 04/21/07 Time: 21:41
Sample: 1 20
Included observations: 20

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	6.854991	7.383473	0.928424	0.3655
INC	3.880186	0.112125	34.60601	0.0000

R-squared	0.985192	Mean dependent var	236.9500
Adjusted R-squared	0.984370	S.D. dependent var	114.8383
S.E. of regression	14.35730	Akaike info criterion	8.261033
Sum squared resid	3710.375	Schwarz criterion	8.360606
Log likelihood	-80.61033	F-statistic	1197.576
Durbin-Watson stat	3.175965	Prob(F-statistic)	0.000000

Bảng 3.2: Kết quả ước lượng các tham số của mô hình

Kết quả ước lượng được tóm tắt lại như sau:

$$INS = 6.85 + 3.88INC \qquad (3.6)$$
$$(7.38) \quad (0.11)$$
$$N = 20, \qquad R^2 = 0.98, \qquad ESS = 3710$$

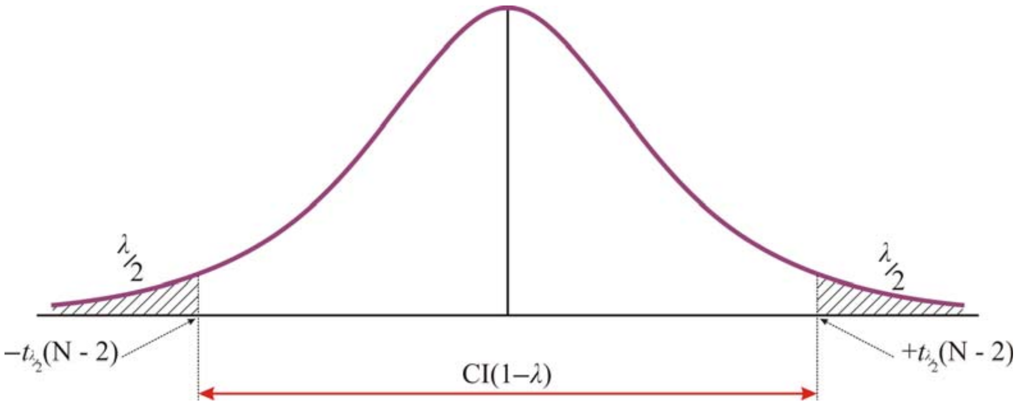
Vấn đề tiếp theo của các nhà hoạch định chiến lược của công ty là liệu họ có thể nói gì về sức mua bảo hiểm tương ứng với từng lớp thu nhập. Điều đó sẽ giúp cho công ty ra quyết định kinh doanh. Ví dụ, nếu thu nhập gia đình tăng thêm một ngàn dollars, thì chi cho bảo hiểm sẽ tăng lên trong khoảng từ 3 ngàn tới 5 ngàn dollars với độ tin cậy là bao nhiêu? Nghĩa là công ty cần xác định được khoảng tin cậy của β tổng thể.

3.4.1 Khoảng tin cậy

Chúng ta sẽ sử dụng các đặc trưng thống kê của ước lượng $\hat{\alpha}, \hat{\beta}$ để đánh giá về các tính chất của tham số thực (tổng thể) α, β.

Từ quan hệ (3.3), $\hat{\beta} = \beta + \sum c_n \varepsilon_n$, và giả thuyết A3 về tính phân bố chuẩn của các yếu tố ngẫu nhiên ε_n, ta đã biết rằng $\hat{\beta}$ có phân bố chuẩn. Hơn thế nữa, từ các đánh giá về trung bình và phương sai của $\hat{\beta}$, ghi trong phương trình (3.4) và (3.5), ta có thể viết lại rằng: $\hat{\beta} \sim N(\beta, \frac{\sigma^2}{S_{XX}})$. Điều đó có nghĩa là, sau khi chuẩn hóa, $Z = \frac{\hat{\beta} - \beta}{\sqrt{\sigma^2 / S_{XX}}} \sim N(0,1)$.

Để công thức này có ý nghĩa ứng dụng, ta thay thế σ^2, bởi ức lượng không trệch của nó là $s^2 = \frac{1}{N-2} \sum_n e_n^2 = \frac{1}{N-2} ESS$. Khi đó, thống kê Z chuyển thành thống kê $t = \frac{\hat{\beta} - \beta}{\sqrt{s^2 / S_{XX}}} = \frac{\hat{\beta} - \beta}{se(\hat{\beta})} \sim t(N-2)$. Đồ thị phân bố của thống kê t, trông rất tương tự như thống kê Z:



Đồ thị 3.5: Phân bố $t = \dfrac{\hat{\beta} - \beta}{se(\hat{\beta})} \sim t(N-2)$

Như đã chỉ ra trên Đồ thị 3.5, khoảng tin cậy (Confidence interval) $(1-\lambda)\%$ của thống kê $t = \dfrac{\hat{\beta} - \beta}{se(\hat{\beta})}$ là vùng mà t sẽ rơi vào khoảng đó với xác suất là $(1-\lambda)$. Tức là:

$$\Pr ob\{-t_{\lambda/2}(N-2) \le \frac{\hat{\beta} - \beta}{se(\hat{\beta})} \le t_{\lambda/2}(N-2)\} = (1-\lambda).$$

Nói khác đi, ta có:

$$\beta \in \{\hat{\beta} \pm se(\hat{\beta}) t_{\lambda/2}(N-2)\} \qquad \text{với độ tin cậy } (1-\lambda)\% \tag{3.7}$$

Chẳng hạn, trong ví dụ về công ty bảo hiểm (3.6), ta có: $\hat{\beta} = 3.88$; $se(\hat{\beta}) = 0.112$. Lưu ý rằng $t_{0.025}[18] = 2.101$, độ tin cậy 95% của β tổng thể là:

$$\beta \in \{3.88 \pm 0.112 \times 2.101\} \tag{3.8}$$

3.4.2 Kiểm định giả thuyết thống kê

Thông thường, kết quả ước lượng mô hình (3.6) và đánh giá độ tin cậy (3.8) sẽ được đính kèm trong bản báo cáo đưa lên cho ban giám đốc công ty để ra quyết định về chiến lược kinh doanh. Tuy nhiên, công việc nghiên cứu thị trường không chỉ dừng lại tại đó. Chúng ta tiếp tục ví dụ bảo hiểm bằng việc nói rằng, ban giám đốc công ty họp để đánh giá bản báo cáo này. Sau đây là những ghi chép được từ cuộc họp:

Nhà quản lý M1 nói rằng, theo kinh nghiệm của ông, thu nhập đã được thể chế hóa qua các tài sản tài chính, như cổ phiếu, địa ốc, vân vân. Và ảnh hưởng của thu nhập bằng tiền mặt tới chi tiêu cho bảo hiểm nhân thọ là rất yếu.

Thành viên khác của ban giám đốc, nhà quản lý M2 lại cho rằng, thu nhập bằng tiền có ảnh hưởng rất mạnh tới nhu cầu mua bảo hiểm nhân thọ. Kinh nghiệm làm ăn của ông cho thấy, cứ 1000 dollars tăng thêm về thu nhập sẽ kéo theo giá trị gói bảo hiểm mua bởi hộ gia đình tăng lên 5000 dollars.

Cuối cùng, ông M3 nêu lại rằng, thu nhập bằng tiền đúng là có ảnh hưởng, nhưng không mạnh tới như vậy. Cứ 1000 dollars tăng thêm về thu nhập chỉ kéo theo nhu cầu về bảo hiểm tăng lên 4000 dollars.

Vậy ai trong số họ là đúng? Và nếu nhận định của nhà quản lý M1 đúng, thì thật là rất đáng tiếc. Vì vậy, chúng ta cần tiến hành kiểm định lại những nhận định này.

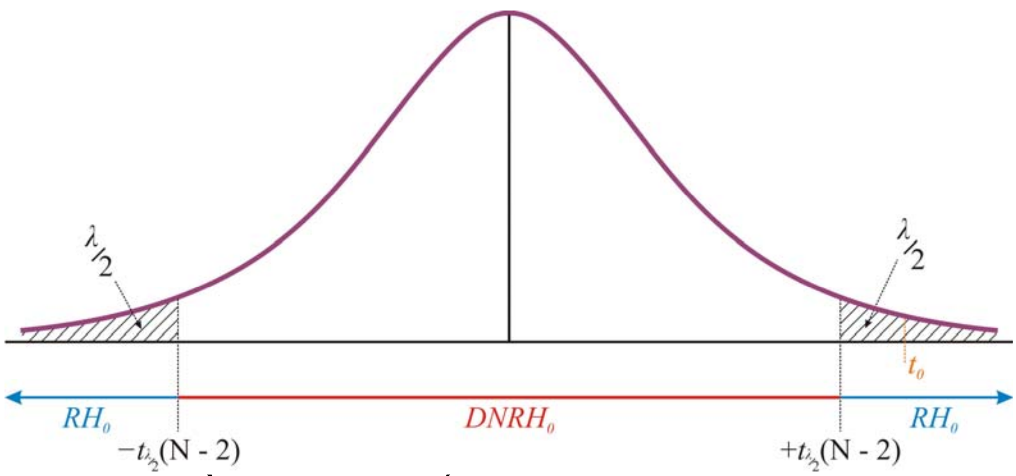
Một cách tổng quát, ta tiến hành kiểm định giả thiết thống kê như sau:

$$H_0 : \beta = \beta_0 \quad .\text{vs.} \quad H_1 : \beta \neq \beta_0$$

Ví dụ, theo nhận định của nhà quản lý công ty M1, ta có:

$$H_0 : \beta = 0 \quad .\text{vs.} \quad H_1 : \beta \neq 0$$

Logic chung của vấn đề kiểm định giả thuyết là như sau: Nếu như nhận định của anh là đúng, thì nó phải phù hợp với phần lớn trường hợp quan sát thấy trên thực tế. Tức là, giá trị thống kê $t_0 = \dfrac{\hat{\beta} - \beta_0}{se(\hat{\beta})}$ phải rơi vào khoảng tin cậy, chẳng hạn là 95%. Trong trường hợp đó, ta không bác bỏ giả thuyết H_0 (hay ký hiệu bằng tiếng Anh: $DNRH_0$). Nếu giá trị $t_0 = \dfrac{\hat{\beta} - \beta_0}{se(\hat{\beta})}$ nằm ngoài khoảng tin cậy, tức là rơi vào vùng hiếm quan sát thấy trên thực tế, khi đó ta bác bỏ H_0 (hay ký hiệu là RH_0).



Đồ thị 3.6: Vùng chấp nhận và bác bỏ H_0

Đồ thị 3.6 thể hiện rằng, chúng ta sẽ bác bỏ (RH_0), nếu $|t_0| = \left|\dfrac{\hat{\beta} - \beta}{se(\hat{\beta})}\right| \geq t_{\lambda/2}(N-2)$, và chúng ta sẽ không bác bỏ ($DNRH_0$), nếu $\left|\dfrac{\hat{\beta} - \beta}{se(\hat{\beta})}\right| \leq t_{\lambda/2}(N-2)$.

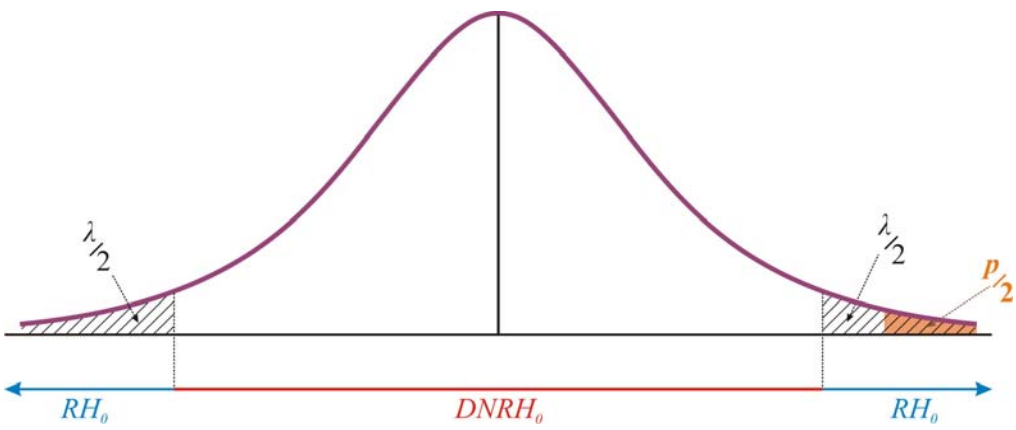
Trong ví dụ nêu trên, đối với nhận định của nhà quản lý M1, ta tiến hành kiểm định như sau:

$$|t_0| = \left|\frac{3.88}{0.112}\right| = 34.6 \geq 2.01 = t_{0.025}[18]$$

Như vậy, dựa trên kết quả kiểm định, ta có thể bác bỏ mạnh mẽ giả định của nhà quản lý M1. Bây giờ chúng ta hãy thử tự kiểm định xem nhận định của các nhà quản lý M2 và M3 có đúng không.

Cuối cùng, để cho tiện sử dụng, trong các software ứng dụng như eviews, người ta thường cho biết giá trị **p-value,** được định nghĩa như sau:

$$P - value = \Pr ob\{|t(N-2)| \geq |t_0|\}$$



Đồ thị 3.7: biểu diễn của p-value

Vì vậy, chúng ta sẽ bác bỏ giả thuyết (RH_0), nếu: $p - value \leq \lambda$, [như chỉ ra trên đồ thị 3.7]. Và chúng ta sẽ không bác bỏ giả thuyết đó ($DNRH_0$), nếu $p - value \geq \lambda$.