Chủ □ề 1: MÔ HÌNH HỒI QUI ĐƠN

VẤN ĐỀ ƯỚC LƯỢNG

I. PHƯƠNG PHÁP BÌNH PHƯƠNG TỐI THIỂU THÔNG THƯỜNG (OLS).

Ta nhắc lại hàm PRF hai biến:

$$Y_i = \beta_1 + \beta_2 X_i + U_i$$

Tuy nhiên, như đã lưu ý trong Chương 1, hàm PRF không thể quan sát trực tiếp được. Ta ước lượng nó từ hàm SRF:

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + e_i$$

Hay: $Y_i = \hat{Y}_i + e_i$

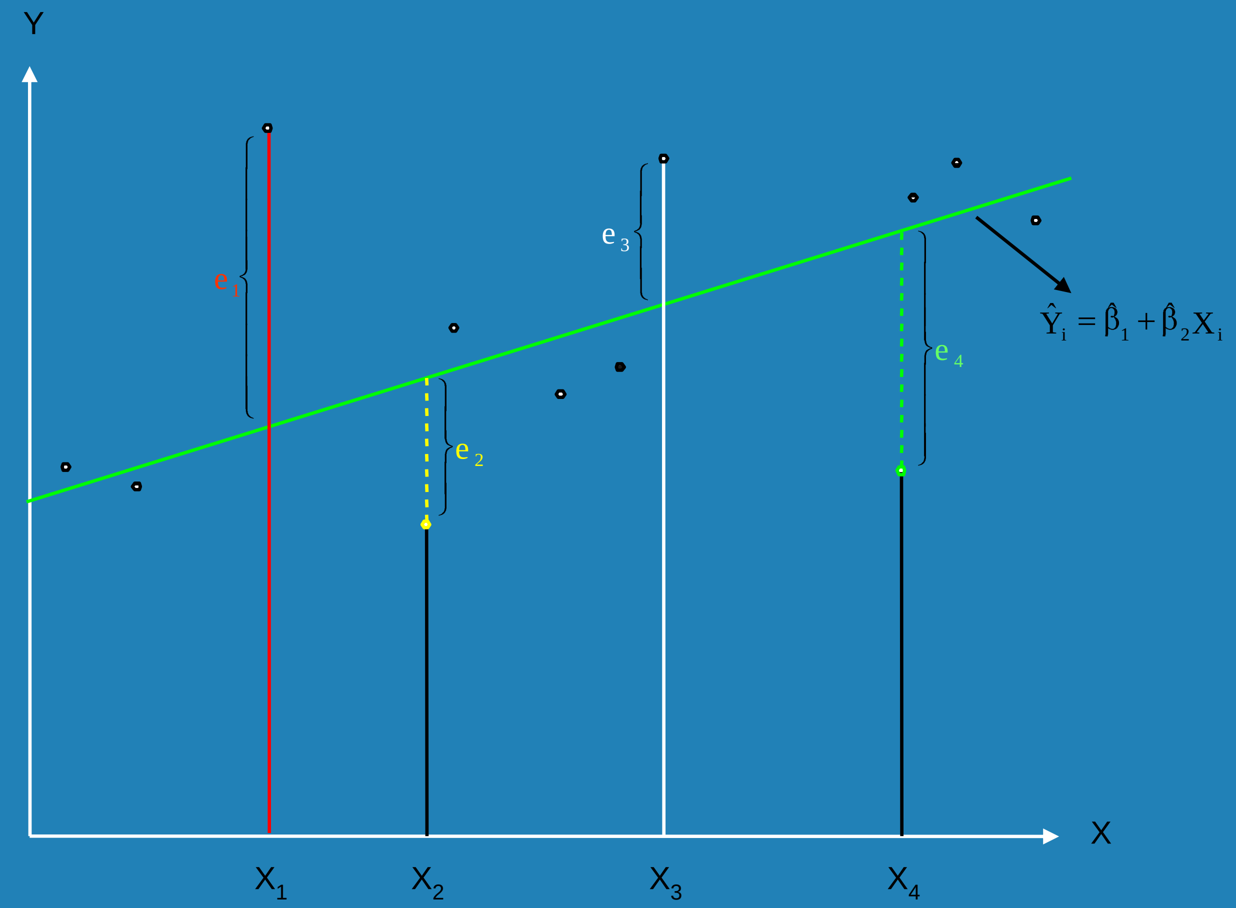
I. PHƯƠNG PHÁP BÌNH PHƯƠNG TỐI THIỂU THÔNG THƯỜNG (OLS).

Nhưng ta sẽ xác định hàm SRF như thế nào? Để thấy được điều này, ta hãy tiến hành như sau. Đầu tiên, ta biểu thị SRF thành :

$$e_i = Y_i - \hat{Y}_i$$

$$= Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i$$

I. PHƯƠNG PHÁP BÌNH PHƯƠNG TỐI THIỂU THÔNG THƯỜNG (OLS).



I. PHƯƠNG PHÁP BÌNH PHƯƠNG TỐI THIỂU THÔNG THƯỜNG (OLS).

Theo nguyên lý của phương pháp OLS để tìm SRF, chúng ta phải cực tiểu tổng bình phương các phần dư, có nghĩa là:

$$\sum e_i^2 = \sum (Y_i - \hat{Y}_i)^2$$

$$= \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)^2 \rightarrow Min$$

Chúng ta có thể xem tổng bình phương các phần dư là một hàm theo $\hat{\beta}_1$ và $\hat{\beta}_2$. Nghĩa là:

I. PHƯƠNG PHÁP BÌNH PHƯƠNG TỐI THIỂU THÔNG THƯỜNG (OLS).

$$\sum e_i^2 = f(\hat{\beta}_1, \hat{\beta}_2)$$

Quá trình vi phân cho các phương trình sau để ước lượng $\hat{\beta}_1$ và $\hat{\beta}_2$:

$$\sum Y_i = n\hat{\beta}_1 + \hat{\beta}_2 \sum X_i$$

$$\sum Y_i X_i = \hat{\beta}_1 \sum X_i + \hat{\beta}_2 \sum X_i^2$$

I. PHƯƠNG PHÁP BÌNH PHƯƠNG TỐI THIỂU THÔNG THƯỜNG (OLS).

Trong đó: *n* là cỡ mẫu. Phương trình này được gọi là các **phương trình chuẩn**.

Giải hệ phương trình chuẩn này, ta thu được:

$$\hat{\beta}_1 = \overline{Y} - \hat{\beta}_2 \overline{X}$$

Và:

$$\hat{\beta}_2 = \frac{n\sum X_i Y_i - \sum X_i \sum Y_i}{n\sum X_i^2 - (\sum X_i)^2} = \frac{\sum (X_i - \overline{X})(Y_i - \overline{Y})}{\sum (X_i - \overline{X})^2} = \frac{\sum x_i y_i}{\sum x_i^2}$$

II. CÁC TÍNH CHẤT CỦA PHƯƠNG PHÁP OLS

- SRF đi qua các giá trị *trung bình mẫu* của *Y* và *X*.

- Giá trị trung bình Y ước lượng bằng giá trị trung bình của *Y* thực. Nghĩa là: $\overline{Y} = \overline{\hat{Y}}$

- Giá trị trung bình của các phần dư e_i bằng 0.

- Các phần dư e_i là không tương quan với Y_i ước lượng. Nghĩa là: $\sum e_i \hat{Y}_i = 0$

II. CÁC TÍNH CHẤT CỦA PHƯƠNG PHÁP OLS

- Các phần dư e_i là không tương quan với X_i, nghĩa là: $\sum e_i X_i = 0$

- $\hat{\beta}_1, \hat{\beta}_2$ là duy nhất ứng với n cặp quan sát cho trước.

- $\hat{\beta}_1, \hat{\beta}_2$ lần lượt là các ước lượng điểm của β_1 , β_2 và là các đại lượng ngẫu nhiên.

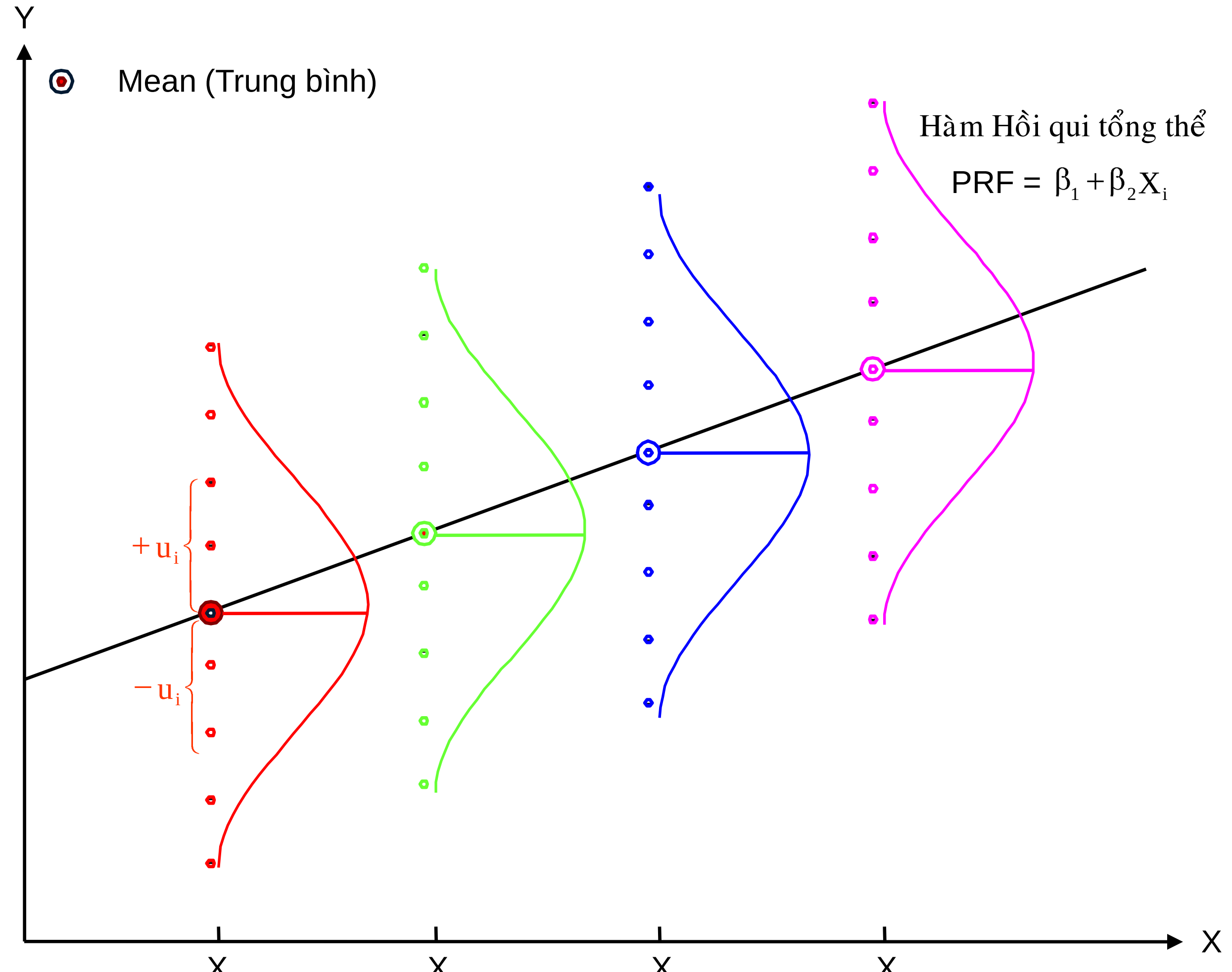
III. CÁC GIẢ THIẾT CỦA PHƯƠNG PHÁP OLS

Giả thiết 1: Mô hình hồi quy tuyến tính. Mô hình hồi quy là **tuyến tính theo các thông số.**

Giả thiết 2: Các giá trị *X* được cố định trong việc lấy mẫu lập lại. Các giá trị rút ra bởi biến hồi qui độc lập *X* được coi là cố định trong các mẫu lập lại. Nói rõ hơn, *X* được giả thiết là *không ngẫu nhiên*.

Giả thiết 3: Giá trị trung bình bằng không của các nhiễu U_i. Cho trước giá trị của *X*, giá trị trung bình hay kỳ vọng của các số hạng nhiễu U_i bằng 0. Nói rõ hơn, giá trị trung bình có điều kiện của U_i là 0. Về mặt ký hiệu, ta có: $E(U_i|X_i)=0$

III. CÁC GIẢ THIẾT CỦA PHƯƠNG PHÁP OLS

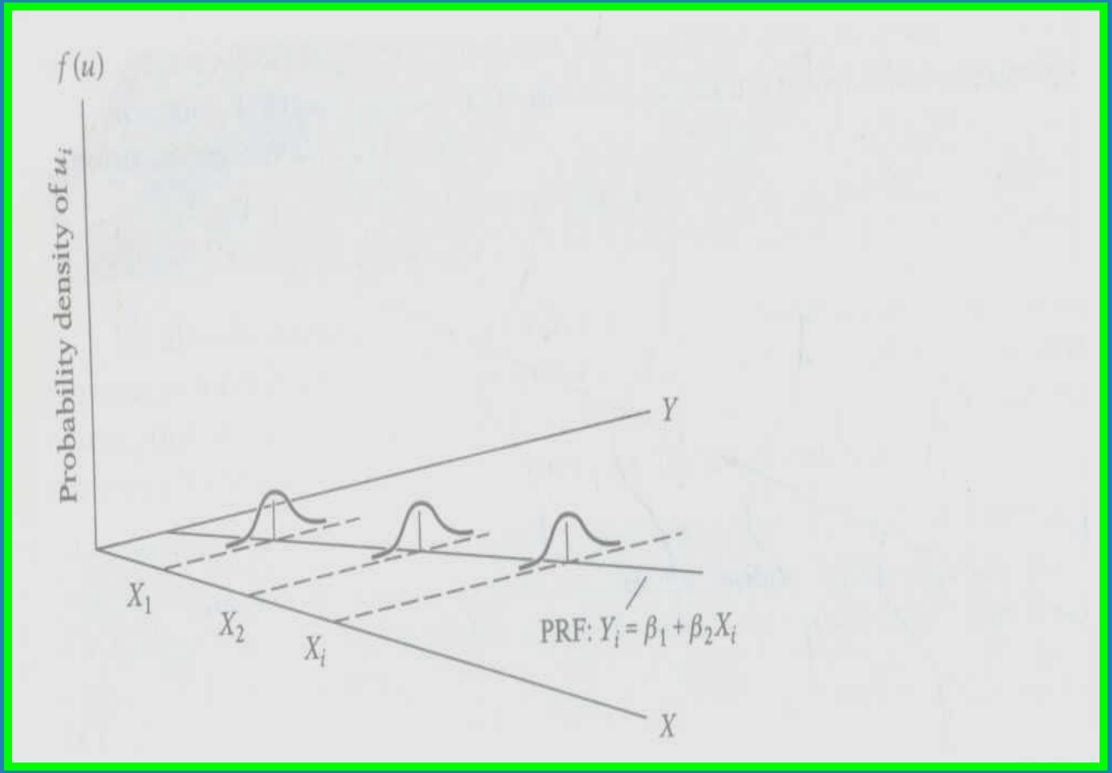


III. CÁC GIẢ THIẾT CỦA PHƯƠNG PHÁP OLS

Giả thiết 4: Phương sai có điều kiện không đổi hay phương sai bằng nhau của U_i. Cho các giá trị của *X*, phương sai của U_i sẽ như nhau đối với tất cả mọi quan sát. Nghĩa là, các phương sai có điều kiện của U_i đều đồng nhất. Về mặt ký hiệu, ta có:

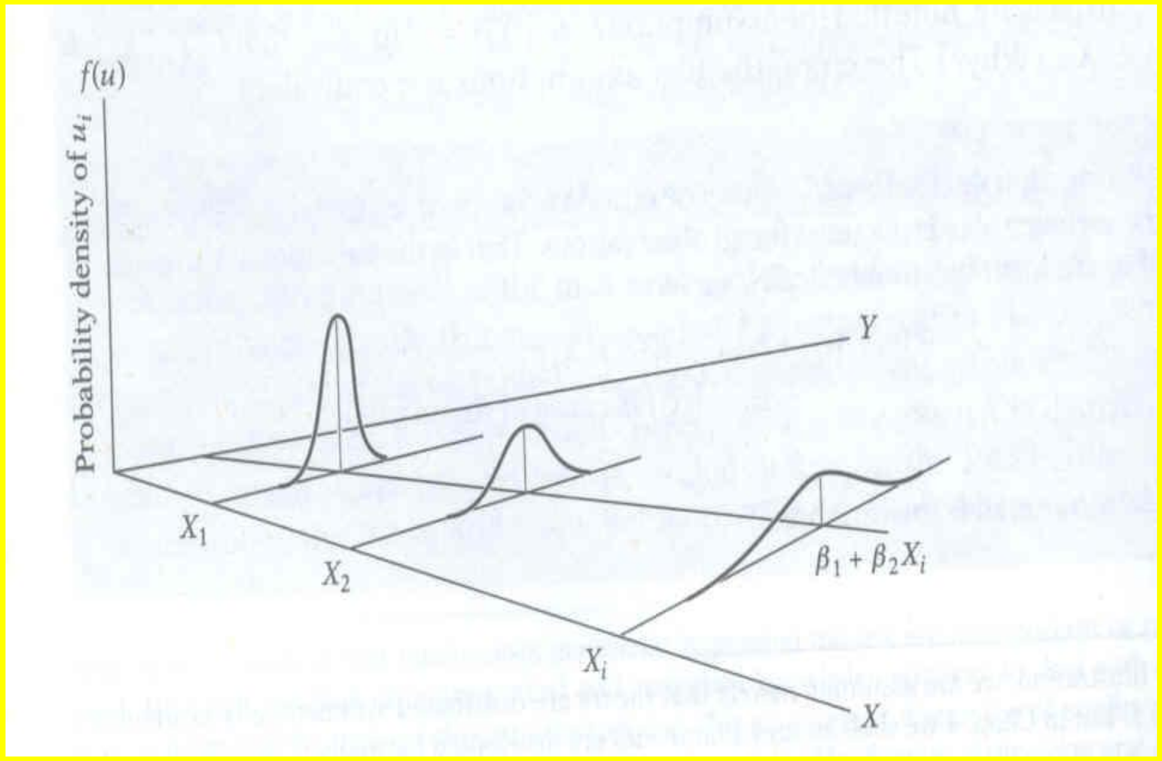
$$\begin{aligned} Var(U_i \mid X_i) &= E[U_i - E(U_i) \mid X_i]^2 \\ &= E(U_i^2 \mid X_i) \\ &= \sigma^2 \end{aligned}$$

III. CÁC GIẢ THIẾT CỦA PHƯƠNG PHÁP OLS



Hình 2.3: ***Phương sai có điều kiện không đổi***

III. CÁC GIẢ THIẾT CỦA PHƯƠNG PHÁP OLS



Hình 2.4: ***Phương sai có điều kiện thay đổi***

III. CÁC GIẢ THIẾT CỦA PHƯƠNG PHÁP OLS

Giả thiết 5: Không có tự tương quan giữa các nhiễu. Cho trước hai giá trị X bất kỳ, X_i và X_j $(i \neq j)$, tương quan giữa U_i và U_j bất kỳ $(i \neq j)$ bằng 0. Về mặt ký hiệu ta có:

$$
\begin{aligned}
Cov(U_i, U_j | X_i, X_j) &= E[U_i - E(U_i) | X_i][U_j - E(U_j) | X_j] \\
&= E(U_i | X_i)(U_j | X_j) \\
&= 0
\end{aligned}
$$

III. CÁC GIẢ THIẾT CỦA PHƯƠNG PHÁP OLS

Giả thiết 6: Đồng phương sai zero giữa U_i và X_i , hay là $E(U_iX_i) = 0$.

Về mặt ký hiệu ta có:

$$
\begin{aligned}
Cov(U_i, X_i) &= E[U_i - E(U_i)][X_i - E(X_i)] \\
&= E[U_i(X_i - E(X_i))] \\
&= E(U_iX_i) - E(X_i)E(U_i) \\
&= E(U_iX_i) \\
&= 0
\end{aligned}
$$

III. CÁC GIẢ THIẾT CỦA PHƯƠNG PHÁP OLS

Giả thiết 7: Số lượng các quan sát n phải lớn hơn số lượng các thông số được ước lượng. Một cách khác, số lượng các quan sát n phải lớn hơn số lượng các biến giải thích.

Giả thiết 8: Các giá trị *X* trong một mẫu cho trước không thể tất cả đều bằng nhau. Nói theo từ ngữ kỹ thuật, var(*X*) phải là một số dương hữu hạn.

$$Var(X) = \frac{\sum (X_i - \bar{X})^2}{n-1}$$, trong đó *n* là cỡ mẫu.

III. CÁC GIẢ THIẾT CỦA PHƯƠNG PHÁP OLS

Giả thiết 9: Mô hình hồi quy được xác định một cách đúng đắn. Nói cách khác, các mô hình được sử dụng trong phân tích thực nghiệm không có **độ thiên lệch hoặc sai số đặc trưng**.

Giả thiết 10: Không có tính đa cộng tuyến hoàn toàn. Nghĩa là *không có các mối tương quan tuyến tính hoàn toàn trong các biến giải thích.*

III. CÁC GIẢ THIẾT CỦA PHƯƠNG PHÁP OLS

Định lý Gauss-Markov: Cho trước các giả thiết của mô hình hồi quy tuyến tính cổ điển, các hàm ước lượng bình phương tối thiểu, trong nhóm các hàm ước lượng tuyến tính không thiên lệch, có phương sai nhỏ nhất, nghĩa là chúng là các hàm ước lượng không thiên lệch tuyến tính tốt nhất. Hay còn gọi là BLUE.

BLUE

Best

Linear

Unbiased

Estimators

IV. TÍNH CHÍNH XÁC HAY LÀ CÁC SAI SỐ CHUẨN CỦA CÁC ƯỚC LƯỢNG BÌNH PHƯƠNG TỐI THIỂU

Các phương sai và sai số chuẩn của các ước lượng bình phương tối thiểu thông thường OLS như sau:

$$Var(\hat{\beta}_2) = \sigma^2_{\hat{\beta}_2} = \frac{\sigma^2}{\sum x_i^2} \Rightarrow se(\hat{\beta}_2) = \sigma_{\hat{\beta}_2} = \frac{\sigma}{\sqrt{\sum x_i^2}}$$

$$Var(\hat{\beta}_1) = \sigma^2_{\hat{\beta}_1} = \frac{\sum X_i^2}{n\sum x_i^2}\sigma^2 \Rightarrow se(\hat{\beta}_1) = \sigma_{\hat{\beta}_1} = \sigma\sqrt{\frac{\sum X_i^2}{n\sum x_i^2}}$$

IV. TÍNH CHÍNH XÁC HAY LÀ CÁC SAI SỐ CHUẨN CỦA CÁC ƯỚC LƯỢNG BÌNH PHƯƠNG TỐI THIỂU

Trong đó: **var** là phương sai, **se** là sai số chuẩn và σ^2 là phương sai có điều kiện không đổi hay phương sai hằng số của U_i, trong giả thiết 4.

Trừ đại lượng σ^2, tất cả các số lượng nhập vào công thức trên đều có thể tính từ dữ liệu, σ^2 tự nó được tính bằng công thức sau:

$$\hat{\sigma}^2 = \frac{\sum e_i^2}{n-2}$$

IV. TÍNH CHÍNH XÁC HAY LÀ CÁC SAI SỐ CHUẨN CỦA CÁC ƯỚC LƯỢNG BÌNH PHƯƠNG TỐI THIỂU

Trong đó: $\hat{\sigma}^2$ là ước lượng không chệch của σ^2.

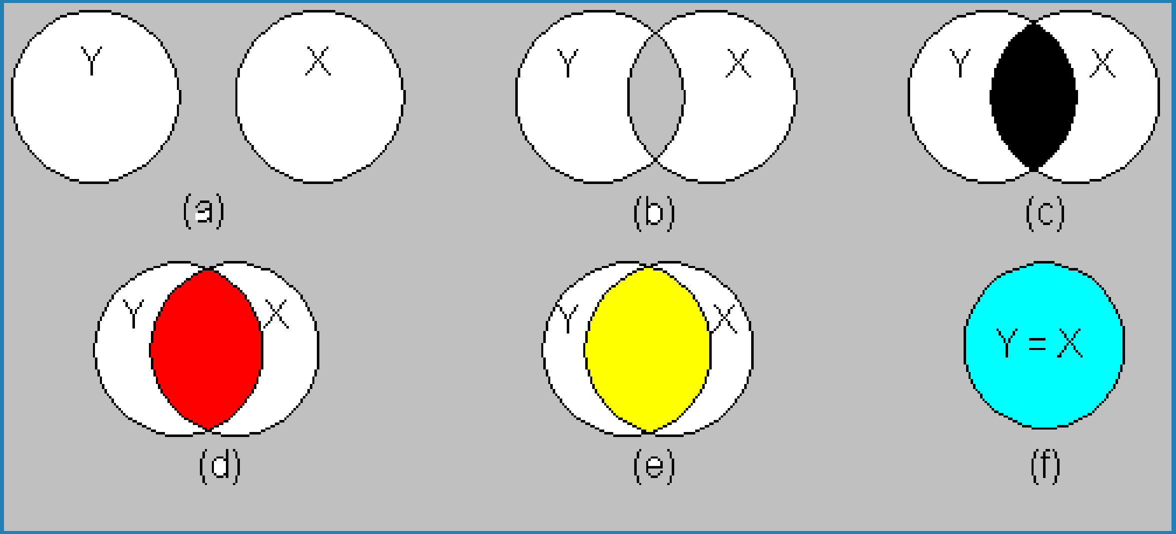
Do đó, trong tính toán thường người ta thay σ^2 bằng $\hat{\sigma}^2$. Khi đó, phương sai và sai số chuẩn ước lượng của các ước lượng OLS sẽ là:

$$\widehat{Var}(\hat{\beta}_2) = \hat{\sigma}^2_{\hat{\beta}_2} = \frac{\hat{\sigma}^2}{\sum x_i^2} \Rightarrow \widehat{se}(\hat{\beta}_2) = \hat{\sigma}_{\hat{\beta}_2} = \frac{\hat{\sigma}}{\sqrt{\sum x_i^2}}$$

$$\widehat{Var}(\hat{\beta}_1) = \hat{\sigma}^2_{\hat{\beta}_1} = \frac{\sum X_i^2}{n\sum x_i^2}\hat{\sigma}^2 \Rightarrow \widehat{se}(\hat{\beta}_1) = \hat{\sigma}_{\hat{\beta}_1} = \hat{\sigma}\sqrt{\frac{\sum X_i^2}{n\sum x_i^2}}$$

V. HỆ SỐ XÁC ĐỊNH (r^2) : ĐẠI LƯỢNG ĐO "SỰ THÍCH HỢP".

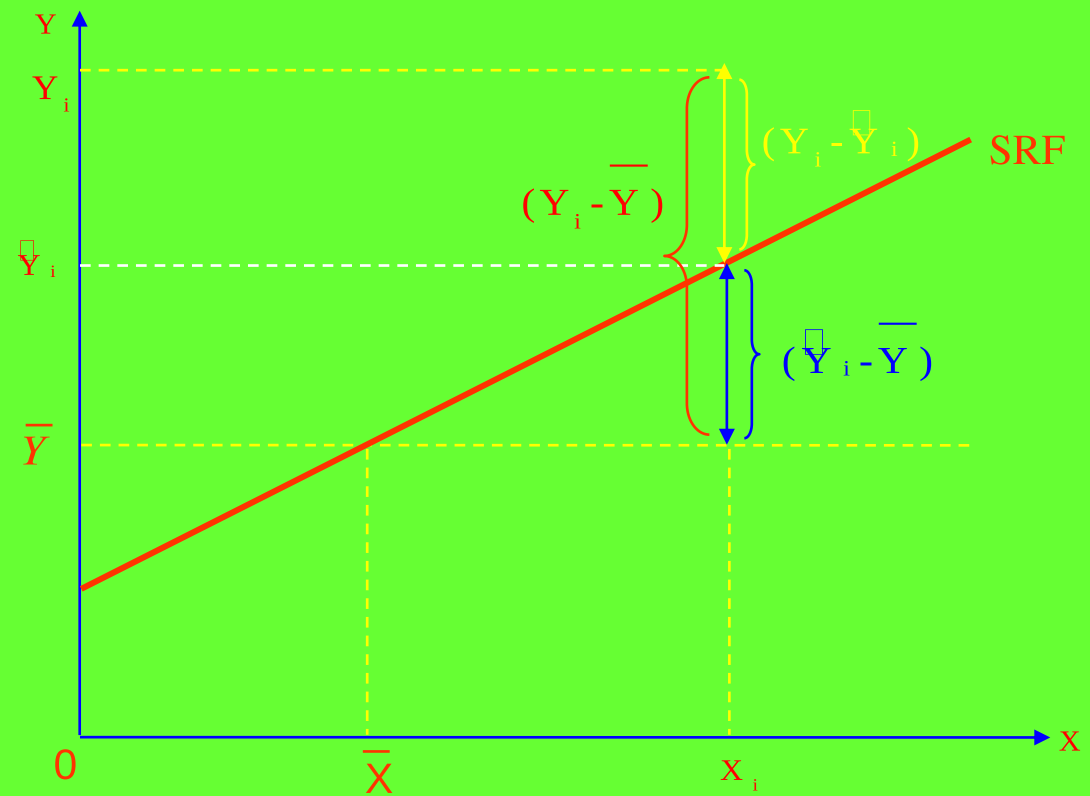
Trước khi chỉ rõ r^2 được tính như thế nào ta hãy xét sự giải thích có tính khai phá đối với r^2 bằng đồ thị, đó là **phương pháp đồ thị Venn, hay là Ballentine**, như trên hình 2.5 sau.



Quan điểm Ballentine đối với r^2: (a) $r^2 = 0$; (f) $r^2 = 1$

V. HỆ SỐ XÁC ĐỊNH (r^2) : ĐẠI LƯỢNG ĐO "SỰ THÍCH HỢP".

Để tính r^2, ta làm như sau:



V. HỆ SỐ XÁC ĐỊNH (r^2) : ĐẠI LƯỢNG ĐO "SỰ THÍCH HỢP".

- Gọi TSS :độ lệch tổng cộng của giá trị thực của *Y* so với trung bình mẫu của chúng, nó có thể được gọi là **tổng bình phương toàn phần.**

$$TSS = \sum y_i^2 = \sum (Y_i - \overline{Y})^2 = \sum Y_i^2 - n\overline{Y}^2$$

- Gọi RSS : Tổng bình phương tất cả các sai lệch giữa giá trị ước lượng của *Y* với trung bình của chúng.

$$RSS = \sum \hat{y}_i^2 = \sum (\hat{Y}_i - \overline{\hat{Y}})^2 = \sum (\hat{Y}_i - \overline{Y})^2 = \beta_2^2 \sum x_i^2$$

V. HỆ SỐ XÁC ĐỊNH (r^2) : ĐẠI LƯỢNG ĐO “SỰ THÍCH HỢP”.

Gọi RSS: là biến thiên **không giải thích** của giá trị *Y* với đường hồi quy, hay đơn giản là **tổng bình phương phần dư.**

$$ESS = \sum e_i^2 = \sum (Y_i - \hat{Y}_i)^2$$

Người ta đã chứng minh được rằng:

TSS = ESS + RSS

V. HỆ SỐ XÁC ĐỊNH (r^2) : ĐẠI LƯỢNG ĐO "SỰ THÍCH HỢP".

Ta định nghĩa r^2 là:

$$r^2 = \frac{RSS}{TSS} = \frac{\sum (\hat{Y}_i - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2}$$

Hay ta có thể viết:

$$r^2 = 1 - \frac{ESS}{TSS} = 1 - \frac{\sum e_i^2}{\sum (Y_i - \bar{Y})^2}$$

V. HỆ SỐ XÁC ĐỊNH (r^2) : ĐẠI LƯỢNG ĐO "SỰ THÍCH HỢP".

Ta lưu ý 2 tính chất của r^2:

- Nó là số không âm (Tại sao?)

- Giới hạn của nó $0 \le r^2 \le 1$. r^2 bằng 1 nghĩa là hoàn toàn phù hợp, nghĩa là $\hat{Y}_i = Y_i$ với mỗi i. Ở đầu khác, $r^2 = 0$ nghĩa là dù thế nào đi nữa (nghĩa là $\beta_2 = 0$) cũng không có liên quan giữa biến hồi qui phụ thuộc và biến hồi qui độc lập.

VI. HỆ SỐ TƯƠNG QUAN MẪU (r)

Hệ số tương quan đo mức độ kết hợp tuyến tính giữa hai biến. Nó được xác định theo công thức sau:

$$r_{X,Y} = \frac{\sum (X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\left[\sum (X_i - \overline{X})^2\right]\left[\sum (Y_i - \overline{Y})^2\right]}}$$

$$= \frac{\sum x_i y_i}{\sqrt{(\sum x_i^2)(\sum y_i^2)}}$$

VI. HỆ SỐ TƯƠNG QUAN MẪU (r)

Giữa hệ số tương quan và hệ số xác định có mối quan hệ với nhau. Chúng ta có thể tính hệ số tương quan theo cách sau:

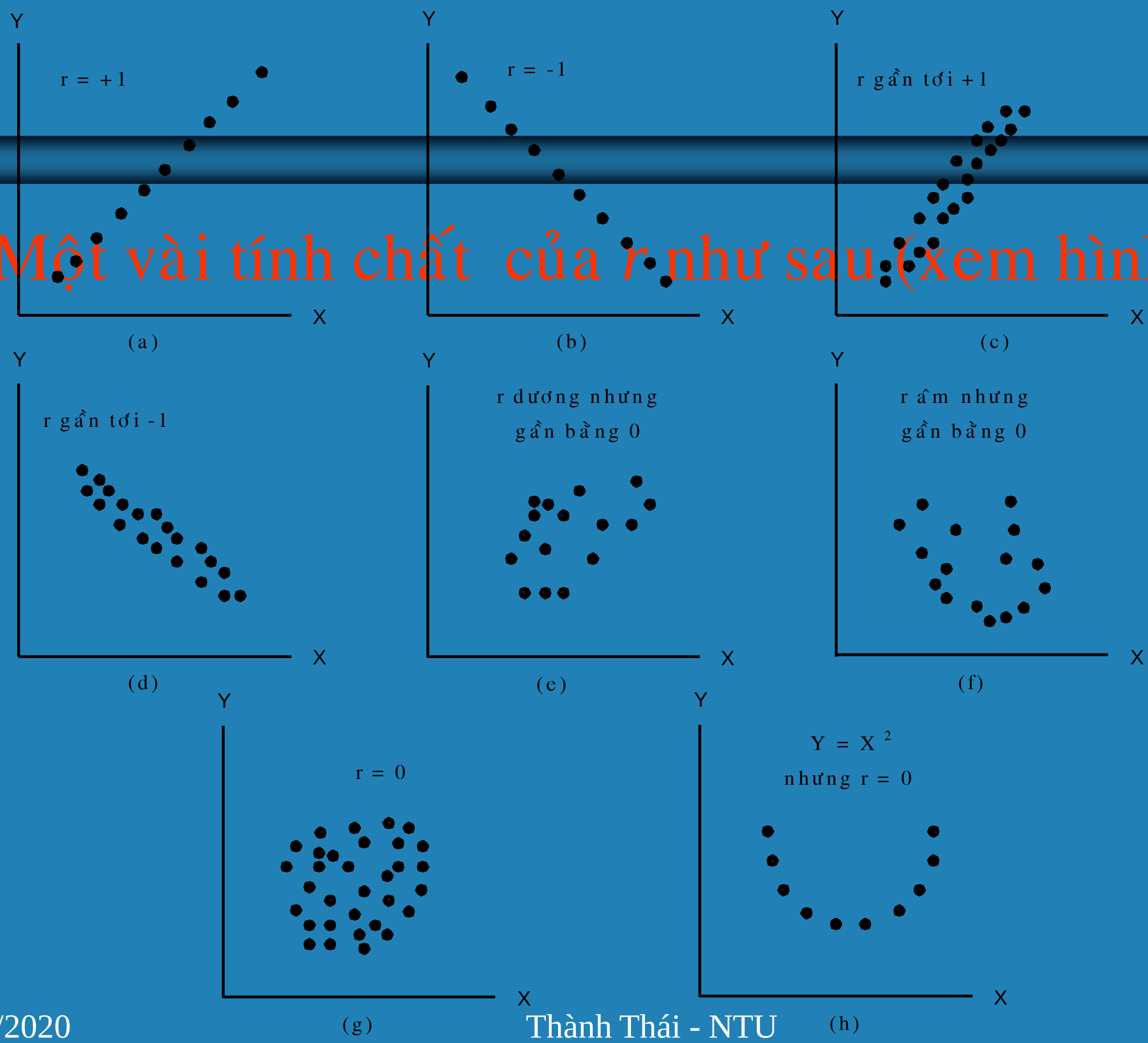
$$r = \pm\sqrt{r^2}$$

- Dấu của r phụ thuộc vào dấu của tử số hay dấu của hệ số góc.

Chú ý: Giá trị bằng số thì quan hệ rất gần, nhưng về khái niệm, r^2 khác xa với **hệ số tương quan.**

VI. HỆ SỐ TƯƠNG QUAN MẪU (r)

Một vài tính chất của *r* như sau (xem hình 3.6):



VI. HỆ SỐ TƯƠNG QUAN MẪU (r)

Một vài tính chất của *r* như sau (xem hình 3.6):

- *r* có thể dương hoặc âm, dấu của *r* phụ thuộc vào dấu của số hạng trong tử số.

- *r* nằm từ –1 đến +1 , nghĩa là: $-1 \le r \le +1$

- Bản chất của *r* là đối xứng ; nghĩa là hệ số tương quan giữa *X* và *Y* (r_{XY}) cũng bằng hệ số đó giữa *Y* và *X* (r_{YX}).

VI. HỆ SỐ TƯƠNG QUAN MẪU (r)

Một vài tính chất của *r* như sau (xem hình 3.6):

- Nếu *X* và *Y* là độc lập theo quan điểm thống kê, hệ số tương quan giữa chúng bằng 0; nhưng nếu $r = 0$, điều đó không có nghĩa là hai biến này độc lập. Nói cách khác, **hệ số tương quan zero không ngụ ý là có tính độc lập** (xem hình 3.6(h)).

- *r* chỉ là đại lượng đo sự *kết hợp tuyến tính* hay là *phụ thuộc tuyến tính*; *r* không có ý nghĩa để mô tả quan hệ phi tuyến tính. Vì vậy, trong hình 3.6(h), $Y = X^2$ là một quan hệ chính xác nhưng $r = 0$. (Tại sao?)

VI. HỆ SỐ TƯƠNG QUAN MẪU (r)

Một vài tính chất của *r* như sau (xem hình 3.6):

- Mặc dù *r* là đại lượng đo sự kết hợp tuyến tính giữa hai biến, *r* không ngụ ý là có bất kỳ mối liên quan nhân quả nào.

✓ Trong nội dung hồi quy, r^2 là đại lượng có đủ ý nghĩa hơn *r*, nó cho ta biết tỷ lệ độ biến thiên trong biến phụ thuộc được giải thích bởi (các) biến giải thích và do đó, nó cũng cho ta thước đo toàn diện của phạm vi mà độ biến thiên trong một biến xác định độ biến thiên trong các biến khác.

VII. PHÂN PHỐI XÁC SUẤT CỦA CÁC ƯỚC LƯỢNG OLS

Giả thiết 11: $U_i \sim N(0, \sigma^2)$

Khi đó các ước lượng OLS tự chúng sẽ có các tính chất sau:

- BLUE

$$\hat{\beta}_2 \sim N(\beta_2, \sigma^2_{\hat{\beta}_2}) \quad \rightarrow \quad Z = \frac{\hat{\beta}_2 - \beta_2}{\sigma_{\hat{\beta}_2}} \sim N(0,1)$$

$$\hat{\beta}_1 \sim N(\beta_1, \sigma^2_{\hat{\beta}_1}) \quad \rightarrow \quad Z = \frac{\hat{\beta}_1 - \beta_1}{\sigma_{\hat{\beta}_1}} \sim N(0,1)$$

VII. PHÂN PHỐI XÁC SUẤT CỦA CÁC ƯỚC LƯỢNG OLS

$$\chi^2 = \frac{(n-2)\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-2)$$

$$Y_i \sim N(\beta_1 + \beta_2 X_i, \sigma^2)$$