Chương 3

PHƯƠNG PHÁP DỰ BÁO ĐỊNH LƯỢNG

3.1. Giới thiệu

Khái niệm và vai trò của dự báo

- **Dự báo** (tiếng Hy Lạp là Prognosis): **sự tiên đoán**, sự thấy trước
- Dự báo (Từ điển Tiếng Việt-Viện ngôn ngữ học- 2006): **Báo trước về tình hình có nhiều khả năng sẽ xảy ra**, dựa trên cơ sở những số liệu, những thông tin đã có.
- Dự báo (Phương pháp dự báo kinh tế căn bản): Dự báo là **tiên đoán khoa học mang tính xác suất và phương án** trong khoảng thời gian hữu hạn về tương lai phát triển của đối tượng kinh tế.
- **Tiên đoán khoa học**: Là những tiên đoán dựa trên việc **phân tích mối liên hệ** qua lại giữa các đối tượng kinh tế và các **phương pháp xử lý thông tin khoa học** nhằm phát hiện ra **tính quy luật** của đối tượng được dự báo.
- Yếu tố quan trọng trong **lập kế hoạch và ra quyết định**.

3.1. Giới thiệu

Phân loại dự báo

- Sử dụng nhiều tiêu chí khác nhau để phân loại dự báo định lượng
- **Phân loại theo thời gian dự báo:**
  - Dự báo ngắn hạn (1-3 năm)
  - Dự báo trung hạn (3-5 năm, <10 năm)
  - Dự báo dài hạn (>10 năm)
- **Phân loại theo đối tượng kinh tế:**
  - Dự báo dân số, dự báo giá cả, dự báo sản lượng tiêu thụ...
- **Phân loại theo kết quả dự báo:**
  - Dự báo điểm và dự báo khoảng
- **Phân loại theo phương pháp tiếp cận đối tượng dự báo:**
  - **Dự báo khảo sát:** Thăm dò trực tiếp đối tượng dự báo
  - **Dự báo mục tiêu:** Tìm phương án tối ưu để đạt được mục tiêu phát triển tương lai, tiếp cận gián tiếp
- **Phân loại theo phương pháp dự báo:**
  - Dự báo bằng phương pháp định tính, phương pháp định lượng

3.1. Giới thiệu

Phân loại phương pháp dự báo

PHƯƠNG PHÁP DỰ BÁO

- PHƯƠNG PHÁP ĐỊNH TÍNH
  - -Lấy ý kiến của ban lãnh đạo
  - -Lấy ý kiến của bộ phận bán hàng
  - -Lấy ý kiến của người tiêu dùng
  - -Phương pháp chuyên gia
- PHƯƠNG PHÁP ĐỊNH LƯỢNG
  - Các mô hình nhân quả
    - -PP hồi quy đơn
    - -PP hồi quy bội
  - Các mô hình chuỗi thời gian
    - -Bình quân đơn giản
    - -Bình quân di động
    - -San bằng số mũ
    - -Chuỗi thời gian
    - -Phương pháp Box- Jenkins

3.1. Giới thiệu

Dự báo định lượng

- **Khái niệm dự báo định lượng:** Phương pháp dự báo định lượng dựa vào các số liệu thống kê và thông qua phương pháp toán học để dự báo cho tương lai.
- **Ưu điểm của phương pháp dự báo định lượng:**
  - Kết quả dự báo là các số liệu cụ thể hỗ trợ tốt cho quản lý, kinh doanh
  - Kết quả dự báo khách quan
  - Phần mềm ứng dụng trong dự báo khá đa dạng, thuận tiện cho sử dụng
  - Có phương pháp đánh giá độ chính xác dự báo
- **Nhược điểm của phương pháp dự báo định lượng:**
  - Yêu cầu cơ sở dữ liệu tốt (Chính xác, đầy đủ, kịp thời, dễ tái lập...)
  - Thường chỉ áp dụng dự báo cho các đối tượng dự báo mang tính định lượng
- Phân loại mang tính tương đối và quy ước, có thể kết hợp các phương pháp khác nhau

3.1. Giới thiệu

Qui trình dự báo

- Bước 1. Xác định mục đích dự báo
- Bước 2. Xác định khoảng thời gian dự báo
- Bước 3. Lựa chọn phương pháp dự báo
- Bước 4. Thu thập và phân tích dữ liệu
- Bước 5. Tiến hành dự báo
- Bước 6. Kiểm chứng kết quả và rút kinh nghiệm

3.2. Phương pháp dự báo theo dãy số thời gian

Khái niệm

- **Các đối tượng kinh tế đều vận động theo quy luật thời gian** (hiện tại chịu ảnh hưởng của quá khứ, tương lai là do quá khứ, hiện tại hình thành theo xu thế phát triển nào đó)
- **Dãy số thời gian:** Dãy các trị số của đối tượng nghiên cứu được sắp xếp theo thứ tự thời gian.

t (thời gian)	**1**	**2**	**3**	**4**	**5...**	**n**
Y(GDP)	$\mathbf{Y_1}$	$\mathbf{Y_2}$	$\mathbf{Y_3}$	$\mathbf{Y_4}$	$\mathbf{Y_5}$	$\mathbf{Y_n}$

- **Dự báo theo chuỗi thời gian:** Phương pháp nghiên cứu phát hiện tính quy luật của đối tượng dự báo trong quá khứ và hiện tại để chuyển sang tương lai
- Phương pháp dự báo chuỗi thời gian đã ngầm hiểu **quy luật phát triển trong quá khứ và hiện tại sẽ được kéo dài trong tương lai**

3.2. Phương pháp dự báo theo dãy số thời gian

Khái niệm

- **Các thành phần của dãy số thời gian:**
  - Tính xu hướng (trend): **T**
  - Tính thời vụ (seasonality): **S**
  - Tính chu kỳ (cycles): **C**
  - Những biến động ngẫu nhiên (random variation): **R**
- **Mô hình số cộng:**

$$Y = T + S + C + R$$

- **Mô hình số nhân:**

$$Y = T * S * C * R$$

- **Dự báo thường sử dụng mô hình số nhân**

4.2. Dự báo dựa trên dữ liệu chuỗi thời gian

Phương pháp dự báo giản đơn

- **Phương pháp dự báo giản đơn là phương pháp dự báo sử dụng giá trị ở thời gian ngay trước làm giá trị dự báo ở ngay sau**
- **Mô hình dự báo:** $F_{t+1} = D_t$
  - F_{t+1} Giá trị dự báo ở kỳ (t+1)
  - D_t Giá trị thực tế ở kỳ (t)
- **Ưu điểm:**
  - Đơn giản, xác định nhanh chóng
- **Nhược điểm:**
  - Mức độ chính xác của dự báo thấp
  - Chỉ dự báo được sau 1 thời kỳ

t (năm)	1	2	3	4	5	6
Y_t (thực tế)	100	150	180	200	210	
F_{t+1} (dự báo)		**100**	**150**	**180**	**200**	**210**

3.2. Phương pháp dự báo theo dãy số thời gian

Dự báo dựa vào lượng tăng (giảm) tuyệt đối bình quân

- Phương pháp này sử dụng khi biến động của hiện tượng có **lượng tăng (giảm) tuyệt đối liên hoàn xấp xỉ nhau.**

  Tăng (giảm) tuyệt đối liên hoàn: $\boldsymbol{\delta_i = y_i - y_{i-1}}$

  Tăng (giảm) định gốc: $\boldsymbol{\Delta_i = y_i - y_1}$

  Tăng (giảm) tuyệt đối bình quân: $\boldsymbol{\Delta = (y_n - y_1)/(n-1) = \Delta_n/(n-1)}$

- **Mô hình dự báo có dạng:**

  $\boldsymbol{y_{n+L} = y_n + \Delta.L}$ **L:** tầm xa dự báo

- Ứng dụng khi cần tính toán nhanh, sơ bộ và ngắn hạn
- Có thể làm sai lệch nếu 2 điểm đầu cuối nằm lệch nhiều so với đường xu thế
- Áp dụng với hiện tượng phát triển theo hàm tuyến tính
- Lãng phí thông tin

3.2. Phương pháp dự báo theo dãy số thời gian

Dự báo dựa vào tốc độ phát triển bình quân

- Phương pháp này sử dụng khi biến động của hiện tượng có **tốc độ phát triển liên hoàn xấp xỉ nhau.**

  Tốc độ phát triển liên hoàn: $\mathbf{t_i = y_i / y_{i-1}}$

  Tốc độ phát triển định gốc: $\mathbf{T_i = y_i / y_1}$

  Tốc độ phát triển bình quân:

$$\bar{t} = \sqrt[n-1]{\frac{y_n}{y_1}}$$

- **Mô hình dự báo có dạng:**

$$y_{n+L} = y_n * (\bar{t})^L$$

  L: tầm xa dự báo

- Phương pháp này áp dụng cho hiện tượng phát triển theo hàm mũ
- Có thể làm sai lệch nếu 2 điểm đầu cuối nằm lệch nhiều so với xu thế các điểm giữa dãy số thời gian
- Lãng phí thông tin

4.2. Dự báo dựa trên dữ liệu chuỗi thời gian

Phương pháp dự báo trung bình đơn giản

- Phương pháp dự báo trên cơ sở lấy trung bình giản đơn của các giá trị quá khứ làm giá trị dự báo cho thời kỳ kế tiếp.
- Công thức:

$$F_{t+1} = \frac{\sum_{i=1}^{t} D_i}{t},$$

  - F_{t+1} Giá trị dự báo cho giai đoạn (t+1)
  - D_i Giá trị thực tế của giai đoạn (i)
  - t Số giai đoạn thực tế
- **Ưu điểm:**
  - Chính xác hơn phương pháp dự báo giản đơn
  - Phù hợp với những dòng yêu cầu đều có xu hướng ổn định.
- **Nhược điểm:**
  - Phải lưu trữ một số lượng dữ liệu khá lớn
  - Chỉ dự báo được một thời kỳ phía sau
  - Phụ thuộc vào mức độ trung bình được tính

4.2. Dự báo dựa trên dữ liệu chuỗi thời gian

Phương pháp dự báo trung bình giản đơn

- Ví dụ 1: Hãy dự báo nhu cầu tháng 6 dựa trên mức bán hàng trung bình thực tế của các tháng trước:

Tháng	Mức bán thực tế (Dt)	Mức bán Dự báo (Ft)
1	100	-
2	110	-
3	120	-
4	130	-
5	140	-
6		**F6 = (100+110+120+130+140)/5= 120**

3.2. Phương pháp dự báo theo dãy số thời gian

Phương pháp dự báo bằng số trung bình động (trung bình trượt)

- Phương pháp dự báo bằng số trung bình trượt dựa trên việc sử dụng số bình quân trượt (số trung bình động) của dãy số thời gian.
- **Phương pháp dự báo bằng số trung bình động:**
  - Số trung bình động không có trọng số
  - Số trung bình động có trọng số
- Phương pháp số trung bình động làm san phẳng sự biến thiên ngẫu nhiên và làm bộc lộ xu thế của hiện tượng nghiên cứu.
- Phương pháp chỉ dự báo được 1 bước về phía trước
- Lãng phí thông tin
- Áp dụng khi biến động quá khứ không lớn
- Không có đột biến trong tương lai

3.2. Phương pháp dự báo theo dãy số thời gian

Phương pháp dự báo bằng số trung bình động (trung bình trượt)

- **Phương pháp dự báo bằng số trung bình động không trọng số**
- **Số trung bình động không trọng số:** Số trung bình cộng của một nhóm nhất định các mức độ của dãy số thời gian và không có trọng số đối với các mức độ ở những thời gian khác nhau.
- Số trung bình động không trọng số (**M**oving **A**verage) được tính:

$$MA_t = \sum_{i=t-K}^{t-1} Y_i / K$$

- **K**: Khoảng tính trung bình có thể lẻ hoặc chẵn, thường chọn lẻ, nếu chọn chẵn thường tính 2 lần
- Số trung bình động tính được có thể để ở giữa khoảng tính trung bình hoặc cuối khoảng tính trung bình
- Mô hình dự báo bằng số trung bình động không trọng số: $\mathbf{Y_{t+1} = MA_t}$

3.2. Phương pháp dự báo theo dãy số thời gian

Phương pháp dự báo bằng số trung bình động (trung bình trượt)

- **Ví dụ:** Dự báo nhu cầu cho tháng tới bằng phương pháp trung bình động, với n=3.

Tháng	Mức bán thực tế (Dt)	Dự báo (Ft)
1	100	
2	110	
3	120	
4	115	F4=(120+110+100)/3
5	125	F5=(115+120+110)/3
6		F6=?

3.2. Phương pháp dự báo theo dãy số thời gian

Phương pháp dự báo bằng số trung bình động có trọng số

- **Phương pháp trung bình động có trọng số:** Bản chất là phương pháp trung bình động nhưng có tính đến ảnh hưởng của từng giai đoạn khác nhau đến biến dự báo thông qua sử dụng trọng số
- Trung bình động có trọng số (**W**eighted **M**oving **A**verage)

$$\mathrm{WMA}_t = \frac{\sum_{i=t-K}^{t-1} \alpha_i Y_i}{\sum_{i=t-K}^{t-1} \alpha_i}$$

α_i Trọng số của giai đoạn **(i)**

- **Giá trị dự báo: F_{t+1} = WMA$_t$**
- **Ưu điểm:** Có thể cho kết quả dự báo sát hơn vì tính đến tầm quan trọng của từng giai đoạn thời gian
- **Nhược điểm:** Việc xác định trọng số phức tạp hơn và cũng chỉ dự báo trước 1 thời kỳ

3.2. Phương pháp dự báo theo dãy số thời gian

Phương pháp dự báo bằng số trung bình động (trung bình trượt)

- **Ví dụ:** Dự báo nhu cầu cho tháng tới bằng phương pháp trung bình động có K= 3 và trọng số tương ứng các tháng quá khứ là 1, 2, 3 tương đối theo thời gian với số trung bình trượt.

Tháng	Mức thực tế	Mức dự báo
1	10	
2	12	
3	13	
4	16	F_4= (10*1+12*2+13*3)/6
5	19	F_5 = (12*1+13*2+16*3)/6
6	23	F_6 = (13*1+16*2+19*3)/6
7		F_7 = (16*1+19*2+23*3)/6

3.2. Phương pháp dự báo theo dãy số thời gian

Phương pháp dự báo bằng san bằng hàm số mũ

- Phương pháp này dựa trên quan điểm các mức độ ở thời gian càng xa càng ít ảnh hưởng đến mức độ ở hiện tại và tương lai.
- Trọng số của các giá trị gần tương lai lớn hơn các trọng số giá trị gần quá khứ
- Mô hình dự báo có dạng:

$$\mathbf{F_{t+1} = \alpha D_t + \alpha(1-\alpha) D_{t-1} + \alpha(1-\alpha)^2 D_{t-2} + \alpha(1-\alpha)^3 D_{t-3} + \ldots}$$

$$\mathbf{F_t = F_{t-1} + \alpha(D_{t-1} - F_{t-1}) = \alpha D_{t-1} + (1-\alpha) F_{t-1}}$$

  - F_t , F_{t-1} Dự báo nhu cầu giai đoạn t, t-1
  - D_t, D_{t-1} Nhu cầu thực của giai đoạn t, t-1
  - α Hệ số san bằng hàm số mũ
- Chọn (**α**) thể hiện mức độ ảnh hưởng (tầm quan trọng) của các số liệu hiện tại đến đại lượng dự báo
- Giá trị (**α**) lớn. dãy số dự báo nhạy bén với sự thay đổi của dãy số ban đầu
- Giá trị (**α**) nhỏ, dãy số dự báo kém nhạy bén với thay đổi dãy số ban đầu
- Giá trị (**α**) chọn sao cho kết quả dự báo có sai số là nhỏ nhất

3.2. Phương pháp dự báo theo dãy số thời gian

Phương pháp dự báo bằng san bằng hàm số mũ

- **Ví dụ:** Hãy dự báo nhu cầu của tháng 6 bằng phương pháp san bằng hàm số mũ với số liệu cho trong Bảng sau:

Tháng (t)	Nhu cầu thực tế (Dt)	Nhu cầu dự báo (Ft)			
		α = 0.10		α = 0.40	
		$F_{t,0.1}$	Sai số	$F_{t,0.4}$	Sai số
1	100	-	-	-	-
2	110	?	?	?	?
3	120	?	?	?	?
4	115	?	?	?	?
5	125	?	?	?	?
6		?	?	?	?

3.2. Phương pháp dự báo theo dãy số thời gian

Phương pháp dự báo bằng san bằng hàm số mũ

- **Giải:** $F_t = F_{t-1} + \alpha(D_{t-1} - F_{t-1}) = \alpha D_{t-1} + (1-\alpha)F_{t-1}$

Tháng (t)	Nhu cầu thực tế (D_t)	Nhu cầu dự báo (F_t)			
		$\alpha = 0.10$		$\alpha = 0.40$	
		$F_{t,0.1}$	Sai số	$F_{t,0.4}$	Sai số
1	100	-	-	-	-
2	110	100	10	100	10
3	120	101	19	104	16
4	115	102.9	12.1	110.4	4.6
5	125	104.11	20.89	112.24	12.76
6		106.20		117.34	

3.2. Phương pháp dự báo theo dãy số thời gian

Phương pháp dự báo bằng hàm xu thế (ngoại suy)

- **Khái niệm:**

  Phương pháp dự báo bằng hàm xu thế chính là việc phát hiện xu thế vận động của đối tượng được dự báo có khả năng tuân theo quy luật hàm số thời gian f(t) nào và dựa vào đó dự báo giá trị của đối tượng trong tương lai.

- **Các bước tiến hành dự báo bằng hàm xu thế:**
  - Xử lý chuỗi thời gian (Phân tích số liệu ban đầu)
  - Phát hiện xu thế (Xây dựng mô hình dự báo)
  - Xây dựng hàm xu thế (Xác định các thông số của mô hình dự báo)
  - Kiểm định hàm xu thế (Đánh giá độ tin cậy của dự báo)
  - Dự báo bằng hàm xu thế (Dự báo điểm và dự báo khoảng)

3.2. Phương pháp dự báo theo dãy số thời gian

Phương pháp dự báo bằng hàm xu thế (ngoại suy)

- **Xử lý chuỗi thời gian:**
  - Thiếu giá trị trong chuỗi thời gian: Trung bình cộng 2 giá trị trước và sau thời điểm thiếu
  - Phương pháp nội suy
  - Xử lý giao động ngẫu nhiên: Làm trơn dãy số (san phẳng) bằng phương pháp trung bình động không có hoặc có trọng số
  - Loại bỏ sai số "thô": Phương pháp kiểm định thống kê toán
    - Tính Độ lệch tiêu chuẩn mẫu hiệu chỉnh $S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2}$
    - Tính giá trị để so sánh: $t_K = \frac{\left| y_K - \bar{y} \right|}{S}$
    - Xác định **$t_n(\alpha)$** Tra Bảng phân phối Student với (n) bậc tự do và xác suất (α) cho trước
    - Nếu **t_K** > **$t_n(\alpha)$** kết luận giá trị **(y_K)** có chứa sai số "thô", loại bỏ và thay bằng giá trị khác đáng tin cậy hơn

3.2. Phương pháp dự báo theo dãy số thời gian

Phương pháp dự báo bằng hàm xu thế (ngoại suy)

- **Phát hiện xu thế:**
  - **Phương pháp đồ thị**
  - **Phương pháp phân tích số liệu**
    - $\hat{y} = a_0 + a_1 t$ — t_i: Cấp số cộng — y_i: Cấp số cộng
    - $\hat{y} = a_0 a_1^t$ — t_i: Cấp số cộng — y_i: Cấp số nhân
    - $\hat{y} = a_0 t^{a1}$ — $\ln t_i$: Cấp số cộng — $\ln y_i$: Cấp số nhân
  - **Phương pháp sai phân**

t_i	1	2	3	4	5	6	7
y_i	2	4	9	19	36	62	99
Δy_i	-	2	5	10	17	26	37
$\Delta^2 y_i$	-	-	3	5	7	9	11
$\Delta^3 y_i$	-	-	2	2	2	2	2

3.2. Phương pháp dự báo theo dãy số thời gian

Phương pháp dự báo bằng hàm xu thế (ngoại suy)

- **Xây dựng hàm xu thế (Xác định các tham số của hàm dự báo)**
  - Sau khi phát hiện khả năng dạng hàm xu thế, cần mô tả dãy số thời gian thông qua các dạng hàm xu thế cụ thể và xác định các tham số của hàm.
- **Phương pháp điểm chọn:**
  - Đơn giản, xác định các tham số bằng xấp xỉ
  - Lãng phí thông tin, độ chính xác không cao, tùy thuộc cách chọn điểm có thể có các bộ tham số khác nhau
  - **Tư tưởng của phương pháp:** Giả định dạng hàm dự báo đã được chọn, chọn các cặp số điểm **(t_i, y_i)** và xác định các tham số của hàm dự báo
  - **Yêu cầu cặp điểm chọn:**
    - Khoảng cách các điểm chọn phải bằng nhau
    - Tổng số các điểm chọn bằng tổng số các tham số
    - Chọn những điểm mà dường biểu diễn hàm xu thế có khả năng đi qua cao nhất

3.2. Phương pháp dự báo theo dãy số thời gian

Phương pháp dự báo bằng hàm xu thế (ngoại suy)

- **Ví dụ:** Cho dãy số thời gian, sử dụng phương pháp điểm chọn để xác định các tham số hàm dự báo. Giả thiết dãy số có thể tuân theo xu thế hàm $\hat{y}= a_0+ a_1t+ a_2t^2$

t_i	1	2	3	4	5	6	7	8	9	10	11	12	13	14
y_i	3	5	6	7	8	9	10	11	13	15	20	22	24	24

- **Giải:** Xác định các a_i bằng phương pháp điểm chọn

Chọn t = 2, 8 và 14

$$\begin{cases} t = 2 & a_0+ 2a_1+ 4a_2 = 5 \\ t = 8 & a_0+ 8a_1+ 64a_2 = 11 \\ t = 14 & a_0+ 14a_1+ 196a_2= 24 \end{cases} \quad \textbf{Giải ra} \quad \begin{cases} a_0 = 4.555556 \\ a_1 = -0.027778 \\ a_2 = 0.097222 \end{cases}$$

- Hàm xu thế có dạng $\hat{y}= 4.56 – 0.027t+ 0.097t^2$

3.2. Phương pháp dự báo theo dãy số thời gian

Phương pháp dự báo bằng hàm xu thế (ngoại suy)

- **Phương pháp tổng bình phương bé nhất:**
  - Phương pháp được ứng dụng rộng rãi để xác định tham số hàm xu thế
  - Mức độ chính xác của phương pháp thể hiện **"Tổng bình phương độ lệch giữa giá trị lý thuyết của hàm xu thế và giá trị thực tế của dãy số thời gian là nhỏ nhất"**
  - *(**S**um of **S**quared **E**rror)* $SSE = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$

    y_i: Giá trị thực tế của dãy thời gian

    $\hat{y}_i$: Giá trị lý thuyết của hàm xu thế

    n: Số mức độ của dãy số thời gian
  - Tùy thuộc vào đặc điểm dãy số mà hàm xu thế được chọn khác nhau: **tuyến tính, bậc 2, bậc 3, parabol...**
  - Hàm phi tuyến được tuyến tính hóa
  - Vấn đề là xác định các tham số của hàm xu thế sao cho **SSE nhỏ nhất**

3.2. Phương pháp dự báo theo dãy số thời gian

Phương pháp dự báo bằng hàm xu thế (ngoại suy)

- **Xác định tham số của hàm xu thế bằng phương pháp tổng bình phương bé nhất**
- Giả sử hàm xu thế có dạng $\hat{y} = a_0 + a_1 t$
- Xác định các a_i sao cho $SSE = \sum(y_i - \hat{y})^2 \rightarrow \min \leftrightarrow \sum(y_i - a_0 - a_1 t)^2 \rightarrow \min$
- Lấy đạo hàm bậc nhất theo a_0 và a_1 của biểu thức trên và cho bằng 0

$$\begin{cases} \sum y_i = n.a_0 + a_1 \sum t_i \\ \sum y_i . t_i = a_0 \sum t_i + a_1 \sum t_i^2 \end{cases}$$

- Giải hệ phương trình tìm a_o và a_1:

$$SS_t = \sum (t_i - \bar{t})^2 = \sum t_i^2 - \frac{(\sum t_i)^2}{n}$$

$$SS_y = \sum (y_i - \bar{y})^2 = \sum y_i^2 - \frac{(\sum y_i)^2}{n}$$

$$SS_{ty} = \sum (t_i - \bar{t})(y_i - \bar{y}) = \sum t_i y_i - \frac{\sum t_i \sum y_i}{n}$$

$$\boxed{\begin{aligned} a_1 &= SS_{ty} / SS_t \\ a_0 &= \bar{y} - a_1 \bar{t} \end{aligned}}$$

3.2. Phương pháp dự báo theo dãy số thời gian

Phương pháp dự báo bằng hàm xu thế (ngoại suy)

- Dạng hàm bậc 2 làm tương tự
- Dạng hàm phi tuyến, cần tìm cách tuyến tính hóa
  - $\hat{y} = a_0 + a_1/t$ Đặt $T = 1/t \rightarrow \hat{y} = a_0 + a_1 T$
  - $\hat{y} = a_0 a_1^t$ Lấy lg 2 vế: $\lg\hat{y} = \log a_0 + t.\lg a_1 \leftrightarrow \hat{Y} = A_0 + A_1.t$
  - $\hat{y} = a_0 t^{a1}$ Lấy lg 2 vế: $\lg\hat{y} = \lg a_0 + a_1 \lg t \leftrightarrow \hat{Y} = A_0 + a_1.T$

- **Ví dụ:** Có số liệu về giá một loại hàng hóa như sau. hãy sử dụng phương pháp bình phương cực tiểu để xác định hàm xu thế của giá hàng hóa đó? Biết rằng hàm xu thế có dạng đường bậc 2.

Thời gian	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Giá	79	128	170	206	235	257	273	282	284	279	267	249	224	192

3.2. Phương pháp dự báo theo dãy số thời gian

Phương pháp dự báo bằng hàm xu thế (ngoại suy)

- **Giải:** Giả sử hàm xu thế có dạng $\hat{y} = a_0 + a_1 t + a_2 t^2$
- Chọn t_i sao cho $\sum t_i = 0$ (t_i = -13, -11, -9...-1, 1, 3, 5, 7, 9...13)
- **Hệ phương trình :**

$$\begin{cases} \sum y_i = n a_0 + a_1 \sum t_i + a_2 \sum t_i^2 \\ \sum y_i t_i = a_0 \sum t_i + a_1 \sum t_i^2 + a_2 \sum t_i^3 \\ \sum y_i t_i^2 = a_0 \sum t_i^2 + a_1 \sum t_i^3 + a_2 \sum t_i^4 \end{cases} \qquad \begin{cases} \sum y_i = n a_0 + a_2 \sum t_i^2 \\ \sum y_i t_i = a_1 \sum t_i^2 \\ \sum y_i t_i^2 = a_0 \sum t_i^2 + a_2 \sum t_i^4 \end{cases}$$

- **Thay số và giải hệ phương trình tìm a_0, a_1 và a_2:**

  $a_0 = 278.074$

  $a_1 = 4.366$

  $a_2 = -0.844$

- **Hàm dự báo có dạng: $\hat{y} = 278.074 + 4.366t - 0.844t^2$ (Với t= -13, -11...)**
- **Hàm dự báo có dạng: $\hat{y} = 22.61538 + 59.39808t - 3.377747t^2$ (Với t=1,2...)**

3.2. Phương pháp dự báo theo dãy số thời gian

Phương pháp dự báo bằng hàm xu thế (ngoại suy)

- **Kiểm định hàm xu thế (Đánh giá độ tin cậy của dự báo)**
- Hàm xu thế chỉ mang tính "có khả năng", cần kiểm tra nhằm đánh giá việc lựa chọn xu thế tối ưu
- **Các tiêu thức kiểm định:**

Sai số tuyệt đối

$$S_{\hat{y}} = \sqrt{\frac{\sum_{1}^{n}(y_i - \hat{y})^2}{n-2}}$$

Sai số tương đối

$$V_{y\%} = \frac{S_{\hat{y}}}{\bar{y}}.100 = \frac{S_{\hat{y}}}{\frac{1}{n}\sum_{1}^{n} y_i}.100$$

- $\mathbf{V_{y\%} > 10\%}$ Không sử dụng hàm f(t) cho dự báo
- $\mathbf{V_{y\%} \leq 10\%}$ Có thể sử dụng hàm f(t) cho dự báo

- **Kiểm tra cập nhật hàm dự báo:** $\mathbf{Vy_{tđối\%} = |y_i - \hat{y}_i| / y_i * 100}$
- $\mathbf{Vy_{tđối\%} > 10\%}$ Không sử dụng hàm f(t) cho dự báo
- $\mathbf{Vy_{tđối\%} \leq 10\%}$ Có thể sử dụng hàm f(t) cho dự báo

3.2. Phương pháp dự báo theo dãy số thời gian

Phương pháp dự báo bằng hàm xu thế (ngoại suy)

- **Dự báo bằng hàm xu thế:** Dự báo điểm và dự báo khoảng
- **Dự báo điểm:**
  - Dự báo giá trị tương lai tại 1 điểm
  - Khoảng cách từ điểm cuối cùng của dãy số đến điểm dự báo-tầm xa dự báo hoặc khoảng cách dự báo
  - Khoảng cách dự báo phụ thuộc vào mức độ ổn định của đối tượng được dự báo
  - Tầm xa dự báo $L_{max} \leq n/3$ (n: số mức độ của dãy số thống kê)
  - Dự báo điểm với khoảng cách dự báo được xác định:

  $$\hat{y}^{DBĐ}{}_{(n+L)} = f(t) = f(n+L)$$

3.2. Phương pháp dự báo theo dãy số thời gian

Phương pháp dự báo bằng hàm xu thế (ngoại suy)

- **Dự báo khoảng:**
- Tìm giá trị dự báo rơi vào khoảng nhất định với xác suất cho trước
- Dự báo khoảng với xác suất cho trước

$$\hat{y}^{DBK}_{(n+L)} = f(n+L) \pm t_{\alpha/2,\, n-p} Se_{(y-\hat{y})} = y^{DBĐ}_{(n+L)} \pm t_{\alpha/2,\, n-p} Se_{(y-\hat{y})}$$

- $\hat{y}^{DBK}_{(n+L)}$: Hàm dự báo khoảng
- $t_{\alpha/2,\, n-p}$:Giá trị (t) trong Bảng phân phối Student với (n-p) bậc tự do và với độ tin cậy α
- **p** : Số tham số của mô hình
- **Hàm tuyến tính (p = 2)**
  - **$Se_{(y-\hat{y})}$**: Sai số dự báo
  - **SSE:** Tổng bình phương sai số giữa giá trị quan sát và giá trị hàm xu thế

$$Se_{(y-\hat{y})} = \sqrt{\frac{SSE}{n-2}} \sqrt{1 + \frac{1}{n} + \frac{3(n+2L-1)}{n(n^2-1)}}$$

$$\sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n-2}}$$

3.2. Phương pháp dự báo theo dãy số thời gian

Phương pháp dự báo bằng hàm xu thế (ngoại suy)

- **Hàm đa thức bậc 2, bậc 3:** $Se_{(y-\hat{y})} = \sqrt{\frac{SSE}{n-p}}\sqrt{1+\frac{1}{n}+T_L(T^{-1})T_L^T}$
  - **T_L:** Ma trận vectơ dòng các giá trị lũy thừa của (t) tại thời điểm (n+L)
    - Hàm bậc 2: $T_L = (1, t_L, t_L^2)$
    - Hàm bậc 3: $T_L = (1, t_L, t_L^2, t_L^3)$
  - **T^{-1}:** Ma trận nghịch đảo của ma trận hệ phương trình chuẩn

Hàm bậc 2 $$T^{-1} = \begin{pmatrix} n & \sum t & \sum t^2 \\ \sum t & \sum t^2 & \sum t^3 \\ \sum t^2 & \sum t^3 & \sum t^4 \end{pmatrix}^{-1}$$

Hàm bậc 3 $$T^{-1} = \begin{pmatrix} n & \sum t & \sum t^2 & \sum t^3 \\ \sum t & \sum t^2 & \sum t^3 & \sum t^4 \\ \sum t^2 & \sum t^3 & \sum t^4 & \sum t^5 \\ \sum t^3 & \sum t^4 & \sum t^5 & \sum t^6 \end{pmatrix}^{-1}$$

  - **T_L^T:** Ma trận chuyển vị của ma trận **T_L**
- **Hàm xu thế dạng khác** $$Se_{(y-\hat{y})} = \sqrt{\frac{SSE}{n-p}}$$

3.2. Phương pháp dự báo theo dãy số thời gian

Phương pháp dự báo bằng mô hình số nhân

- Phương pháp dự báo bằng mô hình số nhân dựa trên cơ sở phân tích các thành phần của dãy số thời gian.
- Mô hình dự báo có dạng: **$\hat{Y}$ = T * S * C * R**
  - Tính xu hướng (trend): **T**
  - Tính thời vụ (seasonality): **S**
  - Tính chu kỳ (cycles): **C**
  - Những biến động ngẫu nhiên (random variation): **R**
- Xác định lần lượt các thành phần trong mô hình và tổng hợp lại
- Các bước tiến hành phương pháp dự báo bằng mô hình số nhân:
  - Xác định phương trình hồi quy lý thuyết **(T)**
  - Xác định chỉ số thời vụ **(S)**
  - Xác định chỉ số chu kỳ **(C)** và bất thường **(R)**
  - Dự báo mức độ tương lai bằng cách kết hợp các yếu tố **(T, S, C, R)**

3.2. Phương pháp dự báo theo dãy số thời gian

Phương pháp dự báo bằng mô hình số nhân

- **Chỉ số thời vụ:**
- Sự biến động của một số hiện tượng kinh tế mang tính thời vụ (Trong từng thời gian nhất định trong năm sự biến động được lặp đi lặp lại qua các năm)
- Nghiên cứu biến động thời vụ giúp chủ động trong quản lý, kinh doanh
- **Phương pháp chỉ số thời vụ:** 2 phương pháp
  - Phương pháp 1: Đối với dãy số không có xu hướng rõ rệt qua thời gian
  - Phương pháp 2: Đối với dãy số có xu hướng tăng qua thời gian

3.2. Phương pháp dự báo theo dãy số thời gian

Phương pháp dự báo bằng mô hình số nhân

- **Phương pháp 1:**
- Chỉ số thời vụ: $S_i = \dfrac{\bar{y}_i}{\bar{y}_0}.100$

$\bar{y}_i$ Số bình quân các mức độ của cùng thời gian (i)

$\bar{y}_0$ Số bình quân tất các mức độ của dãy số thời gian

- **Ví dụ:**

Quý	**Sản lượng hàng hóa tiêu thụ theo năm (tấn)**					$\sum y_i$	$\bar{y}_i$	$S_i = \dfrac{\bar{y}_i}{\bar{y}_0}$
	1	**2**	**3**	**4**	**5**	**(1+2+3+4+5)**	**$=\sum y_i/5$**	
I	1861	1921	1834	1837	2073	9526	1905.2	**90.81**
II	2203	2343	2154	2025	2414	11139	2227.8	**106.19**
III	2415	2514	2098	2304	2339	11670	2334.0	**111.25**
IV	1908	1986	1799	1965	1967	9625	1925.0	**91.75**
						41960	$\bar{y}_0$=**2098**	400.00

3.2. Phương pháp dự báo theo dãy số thời gian

Phương pháp dự báo bằng mô hình số nhân

- **Phương pháp 2:** 3 bước
  - Tính chỉ số thời vụ cá biệt cho từng tháng (quý, mùa...)

$$S_i = \frac{y_i}{\bar{y}_m}$$

y_i Mức độ thực tế của dãy thời gian

$\bar{y}_m$ Số trung bình động tương ứng theo tháng (quý, mùa...)

  - Tính chỉ số thời vụ đại diện cho tháng (quý, mùa...)

    Chỉ số thời vụ đại diện cho tháng (quý, mùa...) bằng **trung bình cộng của các chỉ số thời vụ cá biệt**

  - Hiệu chỉnh chỉ số nếu có sai biệt

3.2. Phương pháp dự báo theo dãy số thời gian

Phương pháp dự báo bằng mô hình số nhân

Năm	Quý	TTế	TBT (4)	TBT(2)	Si =TT/TBT(2)
	I	1861	-	-	-
1	II	2203	-	-	-
	III	2415	2096.75	2104.25	114.77
	IV	1908	2111.75	2129.25	89.61
	I	1921	2146.75	2159.13	88.97
2	II	2343	2171.50	2181.25	107.42
	III	2514	2191.00	2180.13	115.31
	IV	1986	2169.25	2145.63	92.56
	I	1834	2122.00	2070.00	88.60
3	II	2154	2018.00	1994.63	107.99
	III	2098	1971.25	1971.63	106.41
	IV	1799	1972.00	1955.88	91.98
	I	1837	1939.75	1965.50	93.46
4	II	2025	1991.25	2012.00	100.65
	III	2304	2032.75	2062.25	111.72
	IV	1965	2091.75	2140.38	91.81
	I	2073	2189.00	2193.38	94.51
5	II	2414	2197.75	2198.00	109.83
	III	2339	2198.25	-	-
	IV	1967	-	-	-

3.2. Phương pháp dự báo theo dãy số thời gian

Phương pháp dự báo bằng mô hình số nhân

Năm	S_i			
	I	II	III	IV
1	-	-	114.77	89.61
2	88.97	107.415	115.31	92.5604
3	88.599	107.99	106.41	91.9793
4	93.4622	100.646	111.723	91.8063
5	94.5119	109.827	-	-
I_S	**91.39**	**106.47**	**112.05**	**91.49**

- Tổng chỉ số thời vụ 4 quý: 91.39+106.47+112.05+91.49 = 401.40% Mức trung bình mỗi quý là cơ sở để so sánh nên tổng trên phải bằng 100%. Nếu có sai biệt phải có hiệu chỉnh.
- **Hệ số hiệu chỉnh: 400/401.40 = 0.9965**
- Chỉ số thời vụ đại diện từng quý sau khi điều chỉnh:
- **S_{Q1} = 91.39*0.9965 = 91.07; S_{Q2} = 106.10; S_{Q3} = 111.66; S_{Q4}= 91.17**

3.2. Phương pháp dự báo theo dãy số thời gian

Phương pháp dự báo bằng mô hình số nhân

- **Ví dụ:** Hãy sử dụng mô hình số nhân để dự báo về sản lượng tiêu thụ một mặt hàng có số liệu thống kê trong Bảng

Năm / Quý	1	2	3	4	5
I	1861	1921	1834	1837	2073
II	2203	2343	2154	2025	2114
III	2415	2514	2054	2304	2339
IV	1908	1986	1799	1965	1967

3.2. Phương pháp dự báo theo dãy số thời gian

Phương pháp dự báo bằng mô hình số nhân

- Xác định hàm xu thế **(T):** Giả sử để đơn giản chọn hàm $\hat{y} = a_0 + a_1 t$
  - Xác định các tham số của hàm xu thế:
    - $a_0 = \sum y_i / n = 41960/20 = 2098$
    - $a_1 = \sum y_i t_i \sum t_i^2 = 2260/2660 = 0.85$

  Hàm xu thế có dạng: $\hat{y} = 2098 + 0.85t$ **(t=-19,-17,...-1,1,...19)**

  Hoặc hàm có dạng: $\hat{y} = 2080.16 + 1.699t$ **(t=1,2...20)**
- Xác định chỉ số thời vụ **(S)** (Cách 2):

$S_{Q1} = 91.39 * 0.9965 = 91.07$; $S_{Q2} = 106.10$; $S_{Q3} = 111.66$; $S_{Q4} = 91.17$

- Xác định chỉ số chu kỳ **(C)** và bất thường **(R)**

  Tính **C.R** $= y_i / T.S$ cho từng quý trong từng năm

  Tính **C.R** đại diện chung cho từng quý của tất cả các năm (trung bình cộng)
- Dự báo giá trị tương lai theo mô hình nhân: **Y = T.S.C.R**

3.2. Phương pháp dự báo theo dãy số thời gian

Phương pháp dự báo bằng mô hình số nhân

- Tính **C.R và** dự báo **Y = T.S.C.R**:

Quý \ Năm	1	2	3	4	5	C.R
I	98.2	101.0	96.1	96.0	107.9	99.8
II	104.4	105.7	96.8	90.7	103.2	100.2
III	103.7	107.6	89.9	98.0	99.2	99.7
IV	100.3	104.0	98.6	102.3	102.1	101.5

- Dự báo quý I năm thứ 6: $\mathbf{Y_{Q1}}$**= (2098+0.85*21)*91.07%*99.8% =1923 tấn**
- Dự báo quý II năm thứ 6: $\mathbf{Y_{Q2}}$**= (2098+0.85*23)*106.1%*100.2%=2251 tấn**
- Dự báo quý III năm thứ 6:$\mathbf{Y_{Q3}}$**= (2098+0.85*25)*116.6%*99.68%=2358 tấn**
- Dự báo quý IV năm thứ 6:$\mathbf{Y_{Q4}}$**= (2098+0.85*27)*91.17%*101.46%=1962 tấn**

3.3. Dự báo bằng phương pháp hồi quy

Khái niệm

- Dự báo bằng phương pháp hồi quy là việc tìm mối quan hệ phụ thuộc của một biến (Y-biến phụ thuộc) với một biến độc lập (X) hoặc nhiều biến độc lập khác (X_1, X_2,...X_n). Dựa vào mối quan hệ để dự báo giá trị biến phụ thuộc trong tương lai khi biết các biến độc lập.
- **Phương pháp tương quan** được dùng để nghiên cứu **mối quan hệ tuyến tính giữa 2 biến ngẫu nhiên**. Hai biến ngẫu nhiên này được coi là "ngang nhau" không phân biệt biến độc lập hay biến phụ thuộc.
- Phương pháp tương quan nhằm nghiên cứu khuynh hướng, mức độ của liên quan tuyến tính giữa 2 biến ngẫu nhiên.
- Để đánh giá mức độ, chiều hướng của quan hệ tương quan sử dụng hệ số tương quan tổng thể **($-1 \le \rho \le 1$)**

3.3. Dự báo bằng phương pháp hồi quy

Phân tích tương quan

- Giá trị của ρ cho biết mức độ và chiều hướng mối quan hệ tuyến tính giữa 2 biến ngẫu nhiên X và Y
  - $\rho < 0$ X và Y có mối liên hệ nghịch
  - $\rho > 0$ X và Y có mối liên hệ thuận
  - $\rho = 0$ X và Y không có mối liên hệ tuyến tính
  - $|\rho| \to 1$ Mối liên hệ giữa X và Y càng chặt chẽ
- Không thể có **giá trị tương quan tổng thể ρ**, nên cần **ước lượng ρ** từ **hệ số tương quan mẫu (r)** qua dữ liệu mẫu có được.
- Gọi (x_1, y_1); (x_2, y_2);... (x_n, y_n) là mẫu các cặp giá trị quan sát được của biến ngẫu nhiên X và Y. Hệ số tương quan mẫu (r) được tính:

$$r_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} = \frac{\sum x_i y_i - n\bar{x}\,\bar{y}}{\sqrt{\left(\sum x_i^2 - n\bar{x}^2\right)\left(\sum y_i^2 - n\bar{y}^2\right)}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

3.3. Dự báo bằng phương pháp hồi quy

Phân tích tương quan

- **Kiểm định giả thuyết về mối liên hệ tương quan**
- Giả sử có (n) cặp quan sát chọn ngẫu nhiên từ X và Y có phân phối chuẩn N~N(μ,σ^2)
- **Giả thuyết H_0: $\rho = 0$** (X và Y không có liên hệ tương quan)

  H_1: $\rho \neq 0$ (X và Y có liên hệ tương quan)

  - Giá trị kiểm định

$$t = \frac{r_{xy}}{\sqrt{\dfrac{1 - r_{xy}^2}{n-2}}}$$

- Quy tắc quyết định: **$t < -t_{n-2, \alpha/2}$** hoặc **$t > t_{n-2, \alpha/2}$** **($|t| > t_{n-2, \alpha/2}$)**
- **r_{xy}** có thể được tính từ bảng số liệu theo công thức hoặc ***Excel-Tool-Data Analysis-Correlation***
- **t** được tra từ **Bảng phân phối Student** hoặc ***Excel Insert-function TINV*** với mức ý nghĩa **α (Hai phía)** và **(n-2) bậc tự do**

3.3. Dự báo bằng phương pháp hồi quy

Hồi qui đơn biến

- Khái niệm
- Xác định tham số của hàm hồi quy
- Kiểm định hàm hồi quy
- Dự báo bằng hàm hồi quy
- Các dạng hàm hay được sử dụng

3.3. Dự báo bằng phương pháp hồi quy

Hồi quy đơn biến

- **Hàm hồi quy tổng thể:**
- Giả sử có 2 biến X và Y; Y phụ thuộc tuyến tính X
- Mô hình hồi quy tổng thể biểu diễn mối quan hệ tuyến tính giữa X và Y:

$$\mathbf{Y = \alpha + \beta X + \varepsilon}$$

α, β: Các hằng số-tham số của hàm hồi quy tổng thể

ε : Sai số ngẫu nhiên thể hiện ảnh hưởng của các yếu tố khác (không được nghiên cứu) đến Y

Sai số là biến ngẫu nhiên có phân phối chuẩn, có trung bình bằng 0 và phương sai bằng nhau và độc lập với nhau

- Thực tế không thể xác định chính xác các tham số **α, β** của hàm hồi quy tuyến tính của tổng thể mà chỉ có thể ước lượng các tham số đó từ các giá trị quan sát của mẫu.

3.3. Dự báo bằng phương pháp hồi quy

Hồi quy đơn biến

- **Hàm hồi quy mẫu:**
- Giả sử có **(x_1, y_1), (x_2,y_2),...,(x_n,y_n)** mẫu gồm **(n)** cặp quan sát thu thập ngẫu nhiên từ **X và Y**
- Mối quan hệ X,Y thực tế qua mẫu được biểu diễn: $\boldsymbol{y_i = a + bx_i + e_i}$
  - **a, b**: Các hằng số ước lượng của **α và β**
  - **e_i** : Các sai số của hàm hồi quy tuyến tính mẫu
- Hàm hồi quy tuyến tính mẫu: $\boldsymbol{\hat{y}_i = a + bx_i}$
- **Phương pháp tổng bình phương bé nhất:** "Tổng bình phương các chênh lệch giữa giá trị thực tế **y_i** và giá trị hàm lý thuyết $\hat{y}_i$ là nhỏ nhất"

$$\sum_1^n e_i^2 = \sum_1^n (y_i - \hat{y}_i)^2 \rightarrow \min$$

$$\sum_1^n (y_i - a - bx_i)^2 \rightarrow \min$$

3.3. Dự báo bằng phương pháp hồi quy

Hồi quy đơn biến

- **Phương pháp tổng bình phương bé nhất:** $\sum_{1}^{n} (y_i - a - bx_i)^2 \to \min$
- Xác định các tham số **a, b** của hàm hồi quy tuyến tính mẫu:
  - Lấy đạo hàm theo **a, b** và **cho bằng 0**
  - Giải hệ tìm **a và b**

$$\begin{cases} a.n + b\sum x_i = \sum y_i \\ a\sum x_i + b\sum x_i^2 = \sum x_i y_i \end{cases}$$

$$b = \frac{\overline{xy} - \bar{x}\,\bar{y}}{\overline{(x^2)} - \bar{x}^2} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \frac{\sum x_i y_i - n\bar{x}\,\bar{y}}{\sum x_i^2 - n\bar{x}^2}$$

$$a = \bar{y} - b\bar{x}$$

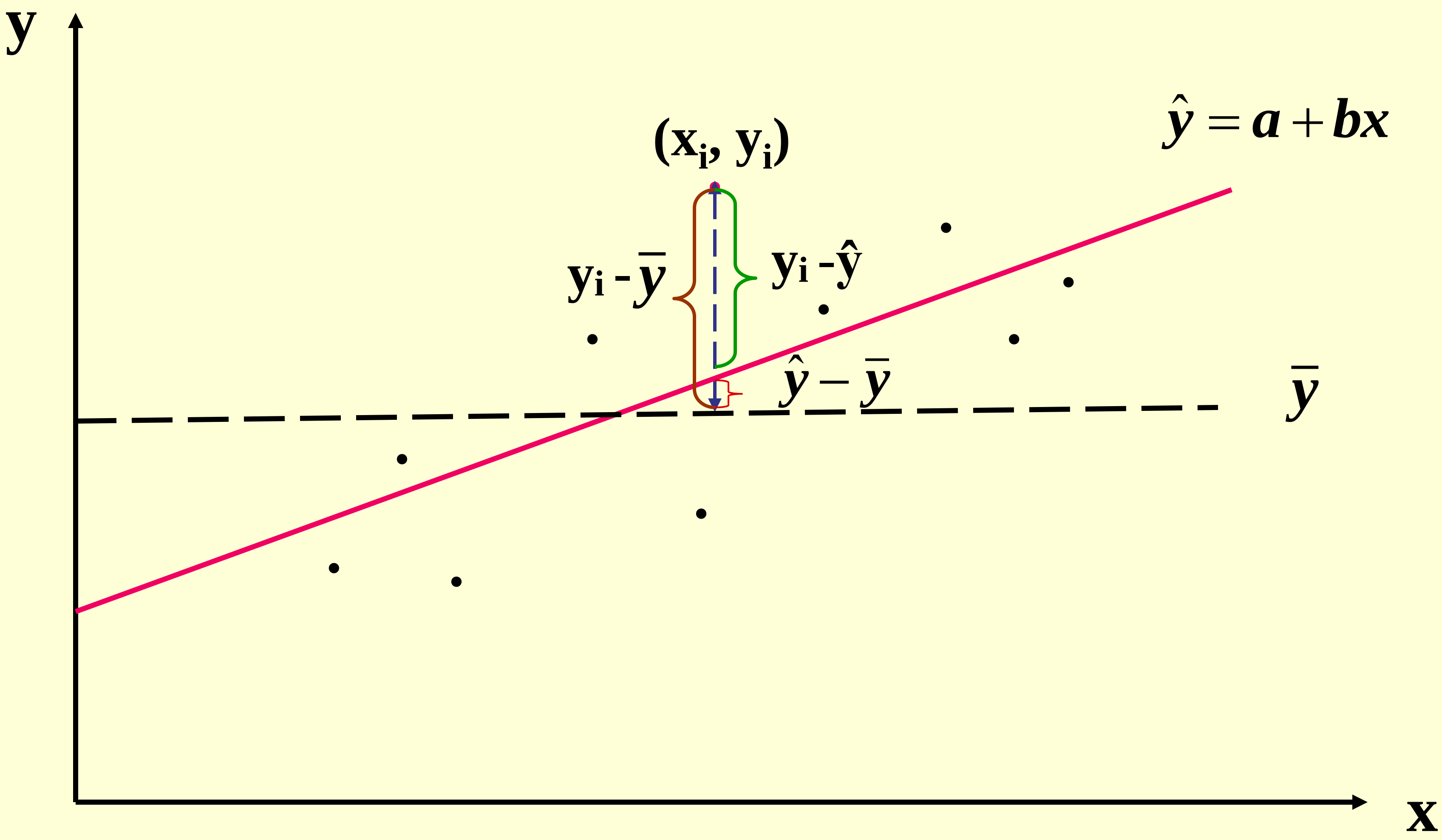
  - Với các ký hiệu:

$$\bar{x} = \frac{\sum x_i}{n} \qquad \bar{y} = \frac{\sum y_i}{n} \qquad \overline{xy} = \frac{\sum x_i y_i}{n} \qquad \overline{x^2} = \frac{\sum x_i^2}{n} \qquad \bar{x}^2 = \left(\frac{\sum x_i}{n}\right)^2$$

  - Hệ số **a,b** có thể xác định theo **công thức** trên hoặc **Excel/tool/analysis/regression hoặc phần mềm Eviews**

3.3. Dự báo bằng phương pháp hồi quy

Hồi quy đơn biến



3.3. Dự báo bằng phương pháp hồi quy

Kiểm định hồi quy đơn

- **Hệ số xác định:**
  - Bao nhiêu % biến thiên của Y có thể được giải thích bởi sự phụ thuộc tuyến tính vào X?

$$TSS = \sum (y_i - \bar{y})^2 = \sum (\hat{y} - \bar{y})^2 + \sum (y_i - \hat{y})^2 = SSR + SSE$$

$$R^2 = \frac{SSR}{TSS} = 1 - \frac{SSE}{TSS} = \frac{\sum (\hat{y} - \bar{y})^2}{\sum (y_i - \bar{y})^2} = r^2 = \left[\frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} \right]^2$$

- **TSS:** (Total Sum of Square) Thể hiện toàn bộ biến thiên của **(y)**
- **SSR:** (Sum of Square for Regression) Thể hiện phần biến thiên của **(y)** được giải thích bởi **(x)**
- **SSE:** (Sum of Square for Error) Thể hiện phần biến thiên của **(y)** do các nhân tố không được nghiên cứu đến
- **$0 \leq R^2 \leq 1$**
- **R^2 càng lớn mô hình hồi quy đã xác định càng thích hợp và có ý nghĩa trong việc giải thích sự biến thiên của Y qua biến độc lập X**

3.3. Dự báo bằng phương pháp hồi quy

Kiểm định hồi quy đơn

- **Kiểm định F:**
  - Sử dụng kiểm định **F** nhằm kiểm định giả thuyết về sự tồn tại của mối liên hệ tuyến tính giữa **X và Y**
  - Giả thuyết **H_0: b = 0**   Không có liên hệ tuyến tính

    H_1: b ≠ 0   Có liên hệ tuyến tính
  - Giá trị kiểm định: $F_{(1,n-2)} = \dfrac{MSE}{MSR} = \dfrac{SSR/1}{SSE/(n-2)}$
  - Quy tắc quyết định:
    - **$F_{(1,n-2)} > F_{\alpha}(1, n-2)$ Bác bỏ H_0**
    - **$F_{(1,n-2)} < F_{\alpha}(1, n-2)$ Không bác bỏ H_0, chấp nhận H_0**
    - **Nếu P-value của F thấp có thể loại bỏ H_0 (P-value ≤ α)**

3.3. Dự báo bằng phương pháp hồi quy

Kiểm định hồi quy đơn

- **Kiểm định t:**
  - Trong hồi quy tuyến tính đơn Kiểm định **F** và Kiểm định **t** tương đương nhau
  - Giả thuyết $\mathbf{H_0: \beta = 0}$ Không có liên hệ tuyến tính

    $\mathbf{H_1: \beta \neq 0}$ Có liên hệ tuyến tính
  - Giá trị kiểm định: $t_{n-2}^{b} = \dfrac{b}{Se_b} = \dfrac{b}{\sqrt{\dfrac{SSE/(n-2)}{\sum (x_i - \bar{x})^2}}} = \dfrac{b}{\sqrt{\dfrac{\left[\sum (y_i - \hat{y})^2\right]/(n-2)}{\sum (x_i^2 - n\bar{x}^2)}}}$
  - Quy tắc quyết định:
    - **Kiểm định 2 phía ở mức ý nghĩa α**
    - $|t^{b}_{(n-2)}| > t_{(n-2,\alpha/2)}$ **Bác bỏ H_0**

3.3. Dự báo bằng phương pháp hồi quy

Kiểm định hồi quy đơn

- **Kiểm định t giả thuyết đối với a:**
  - Giả thuyết **H_0: a = a_0**
  - **H_1: a ≠ a_0**
  - Giá trị kiểm định: $t_{n-2}^{a} = \dfrac{a - a_0}{Se_a}$

$$Se_a = \sqrt{\frac{\frac{SSE}{(n-2)}(\sum x_i^2)/n}{\sum (x_i - \bar{x})^2}} = \sqrt{\frac{\frac{\sum (y_i - \hat{y})^2}{n-2}(\sum x_i^2)/n}{\sum (x_i - \bar{x})^2}} = \sqrt{\frac{\sum (y_i - \hat{y})^2}{n-2}\left[\frac{(\sum x_i^2)/n}{\sum x_i^2 - n.\bar{x}^2}\right]}$$

  - Quy tắc quyết định:
    - **Kiểm định 2 phía ở mức ý nghĩa α**
    - $|t^a_{(n-2)}| > t_{(n-2,\alpha/2)}$ **Bác bỏ H_0**

3.3. Dự báo bằng phương pháp hồi quy

Kiểm định hồi quy đơn

- **Kiểm định t giả thuyết đối với b:**
  - Giá trị kiểm định: $t_{n-2}^{b} = \dfrac{b - b_0}{Se_b}$

$$Se_b = \sqrt{\frac{SSE/(n-2)}{\sum (x_i - \bar{x})^2}} = \sqrt{\frac{\sum (y_i - \hat{y})^2/(n-2)}{\sum x_i^2 - n\bar{x}^2}}$$

  - Quy tắc quyết định:

Loại giả thuyết	Giả thuyết H_0	Giả thuyết H_1	Miền bác bỏ
Hai phía	$b = b_0$	$b \neq b_0$	$\lvert t^b{}_{(n-2)} \rvert > t_{(n-2,\alpha/2)}$
Phía phải	$b \leq b_0$	H_1: $b > b_0$	$t^b{}_{(n-2)} > t_{(n-2,\alpha)}$
Phía trái	$b \geq b_0$	H_1: $b < b_0$	$t^b{}_{(n-2)} > -t_{(n-2,\alpha)}$

3.3. Dự báo bằng phương pháp hồi quy

Ước lượng các tham số hàm hồi quy

- **Ước lượng khoảng tin cậy của các tham số hàm hồi quy:**
  - **Ước lượng khoảng tin cậy của tham số (a) với độ tin cậy (1-α)**

    $a \pm t_{(n-2,\ \alpha/2)}Se_a$ **Se_a** Độ lệch chuẩn ước lượng của **(a)**

$$Se_a = \sqrt{\frac{\frac{SSE}{(n-2)}\sum x_i^2}{\sum (x_i-\bar{x})^2}} = \sqrt{\frac{\frac{\sum (y_i-\hat{y})^2}{n-2}\sum x_i^2}{\sum (x_i-\bar{x})^2}} = \sqrt{\frac{\sum (y_i-\hat{y})^2}{n-2}\left[\frac{\left(\sum x_i^2\right)/n}{\sum x_i^2 - n.\bar{x}^2}\right]}$$

  - **Ước lượng khoảng tin cậy của tham số (b) với độ tin cậy (1-α)**

    $b \pm t_{(n-2,\ \alpha/2)}Se_b$ **Se_b** Độ lệch chuẩn ước lượng của **(b)**

$$Se_b = \sqrt{\frac{SSE/(n-2)}{\sum x_i^2 - n.\bar{x}^2}} = \sqrt{\frac{\sum (y_i-\hat{y})^2/(n-2)}{\sum x_i^2 - n\bar{x}^2}}$$

3.3. Dự báo bằng phương pháp hồi quy

Dự báo bằng hàm hồi quy

- **Dự báo điểm:**

  $\hat{y}^{DBĐ}{}_0 = a + b.x_0$

  x_0 Giá trị biến ngoại sinh trong tương lai

- **Dự báo khoảng cho giá trị cá biệt Y với độ tin cậy (1-α)**

  $\hat{y}^{DBK}{}_0 = \hat{y}^{DBĐ}{}_0 \pm t_{\alpha/2,n-2} Se_{(y_0 - \hat{y}_0)}$

$$Se_{(y-\hat{y}_0)} = \sqrt{\frac{SSE}{n-2}\left[1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{\sum(x_i-\bar{x})^2}\right]} = \sqrt{\frac{\sum(y_i-\hat{y})^2}{n-2}}\sqrt{1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{\sum(x_i-\bar{x})^2}}$$

- **Dự báo khoảng cho giá trị trung bình của Y với độ tin cậy (1-α)**

  $E(Y/x = x_0) = \hat{y}^{DBĐ}{}_0 \pm t_{\alpha/2,n-2} Se_{\hat{y}_0}$

$$Se_{\hat{y}_0} = \sqrt{\frac{SSE}{n-2}\left[\frac{1}{n}+\frac{(x_0-\bar{x})^2}{\sum(x_i-\bar{x})^2}\right]} = \sqrt{\frac{\sum(y_i-\hat{y}_i)^2}{n-2}}\sqrt{\frac{1}{n}+\frac{(x_0-\bar{x})^2}{\sum(x_i-\bar{x})^2}}$$

3.3. Dự báo bằng phương pháp hồi quy

Ví dụ dự báo bằng phương pháp hồi quy

- Có số liệu về lượng hàng bán được và chi phí ở các đợt quảng cáo của một công ty như sau:

Lượng hàng bán được (10000SP)	Chi phí quảng cáo (10000$)
2.0	1
3.0	3
2.5	4
2.0	2
2.0	1
3.5	7

- Hãy xây dựng hàm hồi quy tuyến tính thể hiện mối quan hệ giữa chi phí quảng cáo và lượng hàng bán được của Công ty?
- Hãy tính và giải thích ý nghĩa của hệ số xác định?
- Đánh giá các khoảng tin cậy của tham số hàm dự báo với độ tin cậy 95%?
- Kiểm định giả thuyết mối quan hệ tuyến tính giữa lượng hàng bán được và chi phí quảng cáo của công ty với mức ý nghĩa 5%?
- Hãy dùng phương pháp hồi quy để dự báo lượng hàng bán được khi chi phí quảng cáo tăng gấp đôi so với mức lớn nhất đã chi trong quá khứ?

3.3. Dự báo bằng phương pháp hồi quy

Ví dụ dự báo bằng phương pháp hồi quy

- Giả thiết lượng hàng bán ra của công ty phụ thuộc tuyến tính vào chi phí quảng cáo. Y- Lượng hàng bán ra và X- Chi phí quảng cáo; Y = α + βX+ ε. Qua mẫu điều tra có hàm hồi quy mẫu: y = a+bx+e. Hàm hồi quy lý thuyết có dạng: ŷ = a+ bx. Các hệ số a và b được xác định theo công thức:

$$\bar{x} = \frac{\sum x_i}{n} \qquad \bar{y} = \frac{\sum y_i}{n}$$

$$b = \frac{\sum x_i y_i - n\bar{x}\,\bar{y}}{\sum x_i^2 - n\bar{x}^2}$$

$$a = \bar{y} - b\bar{x}$$

Lượng hàng bán (y)	Chi phí quảng cáo (x)	X^2	xy
2.0	1	1	2
3.0	3	9	9
2.5	4	16	10.0
2.0	2	4	4.0
2.0	1	1	2.0
3.5	7	49	24.5
$\sum y = 15.0$ $\bar{y}$ =2.5	$\sum x = 18.0$ $\bar{x}$ =3.0	$\sum x^2 = 80.0$	$\sum x.y = 51.5$

3.3. Dự báo bằng phương pháp hồi quy

Ví dụ dự báo bằng phương pháp hồi quy

$$b = \frac{\sum x_i y_i - n\bar{x}\,\bar{y}}{\sum x_i^2 - n\bar{x}^2} = \frac{\sum 51.5 - 6*3*2.5}{\sum 80 - 6*3^2} = 0.25$$

$$a = \bar{y} - b\bar{x} = 2.5 - 0.25*3 = 1.75$$

$$\hat{y} = 1.75 + 0.25x$$

- Hệ số xác định $R^2 = \dfrac{SSR}{TSS} = \dfrac{\sum(\hat{y} - \bar{y})^2}{\sum(y_i - \bar{y})^2} = 0.8125$

$$r^2 = 0.901388^2 = 0.8125$$

- Như vậy sự biến thiên của lượng hàng bán được có liên quan tuyến tính khá chặt với chi phí quảng cáo và được giải thích 81% qua sự biến thiên của chi phí quảng cáo và 19% là các yếu tố không được xem xét đến.

3.3. Dự báo bằng phương pháp hồi quy

Ví dụ dự báo bằng phương pháp hồi quy

- **Ước lượng các tham số hàm hồi quy với độ tin cậy 95%**
  - **Ước lượng khoảng tin cậy của tham số (a) với độ tin cậy (1-α)**

    a ± t$_{(n-2,\ \alpha/2)}$Se$_a$ **Se$_a$** Sai số chuẩn ước lượng của **(a)**

$$Se_a = \sqrt{\frac{SSE}{n-2} \times \frac{\sum x_i^2}{\sum (x_i - \bar{x})^2}} = \sqrt{\frac{\sum (y_i - \hat{y})^2}{n-2} \times \frac{\sum x_i^2}{\sum (x_i - \bar{x})^2}} = \sqrt{\frac{\sum (y_i - \hat{y})^2}{n-2} \left[\frac{\left(\sum x_i^2\right)/n}{\sum x_i^2 - n.\bar{x}^2} \right]}$$

  - **Ước lượng khoảng tin cậy của tham số (b) với độ tin cậy (1-α)**

    b ± t$_{(n-2,\ \alpha/2)}$Se$_b$ **Se$_b$** Sai số chuẩn ước lượng của **(b)**

$$Se_b = \sqrt{\frac{SSE/(n-2)}{\sum x_i^2 - n.\bar{x}^2}} = \sqrt{\frac{\sum (y_i - \hat{y})^2 /(n-2)}{\sum x_i^2 - n\bar{x}^2}}$$

3.3. Dự báo bằng phương pháp hồi quy

Ví dụ dự báo bằng phương pháp hồi quy

Lượng hàng bán (y)	Chi phí quảng cáo (x)	x^2	$\hat{y}$	$y-\hat{y}$	$(y-\hat{y})^2$
2	1	1	2.00	0	0
3	3	9	2.50	0.5	0.25
2.5	4	16	2.75	-0.25	0.0625
2.0	2	4	2.25	-0.25	0.0625
2.0	1	1	2.00	0	0
3.5	7	49	3.50	0	0
$\sum y = 15.0$ $\bar{y} = 2.5$	**$\sum x = 18.0$ $\bar{x} = 3.0$**	**$\sum x^2 = 80$ $\overline{x^2} = 13.3$**			**0.375**

$Se_a = 0.219265$; $Se_b = 0.060048$

Khoảng tin cậy của a: 1.75 ± 2.776*0.219265 = 1.75 ± 0.608678

Khoảng tin cậy của b: 0.25 ± 2.776*0.060048 = 0.25 ± 0.166693

3.3. Dự báo bằng phương pháp hồi quy

Ví dụ dự báo bằng phương pháp hồi quy

- Kiểm định giả thuyết mối quan hệ tuyến tính giữa lượng hàng bán được và chi phí quảng cáo của công ty với mức ý nghĩa 5%.
- **Kiểm định t:**
  - Giả thuyết **H_0: β = 0** Không có liên hệ tuyến tính

    H_1: β ≠ 0 Có liên hệ tuyến tính
  - Giá trị kiểm định:

$$t_{(n-2)} = \frac{b}{Se_b} = \frac{b}{\sqrt{\frac{SSE/(n-2)}{\sum(x_i-\bar{x})^2}}} = \frac{b}{\sqrt{\frac{\left[\sum(y_i-\hat{y})^2\right]/(n-2)}{\sum(x_i^2-n\bar{x}^2)}}} = \frac{0.25}{0.060048} = 4.163336$$

  - **Quy tắc quyết định:**
    - **Kiểm định 2 phía ở mức ý nghĩa α**
    - **$|t_{(n-2)}| > t_{(n-2,\alpha/2)}$ Bác bỏ H_0**
    - **|4.163336|> 2.766 Bác bỏ H_0**, tức là có mối quan hệ tuyến tính giữa lượng bán được và chi phí quảng cáo của công ty với mức ý nghĩa 5%

3.3. Dự báo bằng phương pháp hồi quy

Ví dụ dự báo bằng phương pháp hồi quy

- Hãy dùng phương pháp hồi quy để dự báo lượng hàng bán được khi chi phí quảng cáo tăng gấp đôi so với mức lớn nhất đã chi trong quá khứ.
- **Dự báo điểm:**

$$\hat{y}^{DBĐ}_0 = a + b.x_0 = 1.75+0.25*14 = 5.25$$

- **Dự báo khoảng giá trị cá biệt với độ tin cậy (1-α)**

$$\hat{y}^{DBK}_0 = \hat{y}^{DBĐ}_0 \pm t_{\alpha/2,n-2} Se_{(y_0-\hat{y}_0)} = 5.25 \pm 2.776*0.738697 = 5.25 \pm 2.050623$$

$$Se_{(y-\hat{y}_0)} = \sqrt{\frac{SSE}{n-2}\left[1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{\sum(x_i-\bar{x})^2}\right]} = 0.306186\sqrt{1+\frac{1}{n}+\frac{(14-3)^2}{26}} = 0.738697$$

$$\sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{\sum(y_i-\hat{y}_i)^2}{n-2}} = \sqrt{\frac{0.375}{4}} = 0.306186$$

3.3. Dự báo bằng phương pháp hồi quy

Ví dụ dự báo bằng phương pháp hồi quy

- **Dự báo khoảng giá trị trung bình với độ tin cậy (1-α)**

$E(Y/x = x_0) = \hat{y}^{DBĐ}{}_0 \pm t_{\alpha/2,n-2} Se_{\hat{y}_0} = 5.25 \pm 2.776*0.672252 = 5.25 \pm 1.866171$

$$Se_{\hat{y}_0} = \sqrt{\frac{SSE}{n-2}\left[\frac{1}{n} + \frac{(x_0 - \overline{x})^2}{\sum (x_i - \overline{x})^2}\right]} = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n-2}} \sqrt{\frac{1}{n} + \frac{(x_0 - \overline{x})^2}{\sum (x_i - \overline{x})^2}} = 0.672252$$

SUMMARY OUTPUT

Regression Statistics		
Multiple R	**0.901387819**	**r**
R Square	**0.8125**	**R^2**
Adjusted R Square	**0.765625**	**R^2điều chỉnh = 1-[SSE/(n-k-1)]/[TSS/(n-1)]**
Standard Error	**0.306186218**	$\sqrt{SSE/(n-2)}$
Observations	**6**	**n**

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	**1**	**1.625**	**1.625**	**17.33333**	**0.014107073**
Residual	**4**	**0.375**	**0.09375**		
Total	**5**	**2**			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	**1.75**	**0.219264505**	**7.981227976**	**0.001336**	**1.141222878**	**2.358777122**
X Variable 1	**0.25**	**0.060048058**	**4.163331999**	**0.014107**	**0.083279519**	**0.416720481**

MSE = SSR/1

Significance F: p value

Coefficients: Hệ số hồi quy (a, b)

Intercept: Điểm cắt tung độ (a)

Standard Error: Sai số chuẩn của ước lượng các hệ số (Se_a, Se_b)

X Variable 1: Hệ số góc (b)

Lower 95%: a-tn-2,α/2*Se_a; b-$t_{n-2,\alpha/2}$*Se_b

Upper 95%: a+$t_{n-2,\alpha/2}$*Se_a; b+$t_{n-2,\alpha/2}$*Se_b

Included observations: 6

Y=C(1)+C(2)*X

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	1.750000	0.219265	7.981228	0.0013
C(2)	0.250000	0.060048	4.163332	0.0141

R-squared	0.812500	Mean dependent var	2.500000
Adjusted R-squared	0.765625	S.D. dependent var	0.632456
S.E. of regression	0.306186	Akaike info criterion	0.731955
Sum squared resid	0.375000	Schwarz criterion	0.662542
Log likelihood	-0.195865	Durbin-Watson stat	2.333333

3.3. Dự báo bằng phương pháp hồi quy

Hồi qui đơn biến

- **Các dạng hàm hay được sử dụng trong hồi quy đơn biến**
  - $y = a + bx$
  - $y = a + b_1x + b_2x^2$

$$\begin{cases} na + b_1\sum x + b_2\sum x^2 = \sum y \\ a\sum x + b_1\sum x^2 + b_2\sum x^3 = \sum xy \\ a\sum x^2 + b_1\sum x^3 + b_2\sum x^4 = \sum x^2y \end{cases}$$

  - $y = a + b_1x + b_2x^2 + b_3x^3$

$$\begin{cases} na + b_1\sum x + b_2\sum x^2 + b_3\sum x^3 = \sum y \\ a\sum x + b_1\sum x^2 + b_2\sum x^3 + b_3\sum x^4 = \sum xy \\ a\sum x^2 + b_1\sum x^3 + b_2\sum x^4 + b_3\sum x^5 = \sum x^2y \\ a\sum x^3 + b_1\sum x^4 + b_2\sum x^5 + b_3\sum x^6 = \sum x^3y \end{cases}$$

  - $y = a + b/x$ Đặt $1/x = X \rightarrow$ **$y = a + bX$**
  - $y = a + b^x$ Log hóa 2 vế **$\lg y = \lg a + x\lg b \rightarrow Y = A + Bx$**
  - $y = ax^b$ Log hóa 2 vế **$\lg y = \lg a + b\lg x \rightarrow Y = A + bX$**
  - $y = a + b\lg x$ Đặt $\lg x = X \rightarrow$ **$y = a + bX$**

3.3. Dự báo bằng phương pháp hồi quy

Hồi quy đa biến - Hồi quy bội

- Mối quan hệ giữa biến phụ thuộc (Y) với nhiều biến độc lập (X_1, X_2,...X_k), $Y = f(X_1, X_2,...X_k)$
- Hàm sản xuất Cobb-Douglas: $\mathbf{Q = aK^{\alpha}L^{\beta}}$
- Hàm sản xuất KLEM: $\mathbf{Q = aK^{\alpha}L^{\beta}E^{\gamma}M^{\eta}}$
- Mối quan hệ giữa nhu cầu với giá sản phẩm, giá sản phẩm thay thế, thu nhập...: **Q = f (Px, Py, I...)**
- Hàm f (X_1, X_2, X_3...X_k) có thể ở dạng tuyến tính hoặc phi tuyến
- Phương pháp xác định các tham số của hàm hồi quy bội tương tự hồi quy đơn biến – Bình phương cực tiểu (Bảng tính, phần mềm Excel, Eviews...)
- Các bước tiến hành dự báo bằng phương pháp hồi quy bội tương tự như phương pháp dự báo bằng hồi quy đơn biến

3.3. Mô hình hồi qui đa biến (HQ bội)

Giả thiết của mô hình hồi qui đa biến

- **Các giả thiết cho mô hình hồi qui bội:**

1. $E(u_i) = 0$ — Kỳ vọng của các yếu tố ngẫu nhiên u_i bằng 0

2. $Var(u_i) = \sigma^2$ — Phương sai bằng nhau với mọi u_i

3. $Cov(u_i, u_j) = 0$ — Không có sự tương quan giữa các u_i

4. $Cov(u_i, x_i) = 0$ — U và X không tương quan với nhau

5. u_i — $U \sim N(0, \sigma^2)$ Phân phối chuẩn

6. Giữa các $X_1, X_2, ..X_k$ không có quan hệ tuyến tính

Hay không tồn tại $\lambda_i \equiv 0$: $\lambda_1 X_{1i} + \lambda_2 X_{2i} + \lambda_3 X_{3i} + ... + \lambda_k X_{ki} = 0$

Nếu $X_1, X_2, ..X_k$ có quan hệ tuyến tính - có hiện tượng **đa cộng tuyến**

X_1	x_{11}	x_{12}	x_{13}	x_{1n}
X_2	x_{21}	x_{22}	x_{23}	x_{2n}
...	...	...	...	...
X_k	x_{k1}	x_{k2}	x_{k3}	x_{kn}
Y	y_1	y_2	y_3	y_n

3.3. Mô hình hồi qui đa biến-Hồi quy bội

Giới thiệu mô hình hồi quy tuyến tính bội

- Hàm hồi qui tuyến tính bội tổng thể có dạng

  $\mathbf{Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \ldots \beta_k X_k + U}$

  α: Hệ số tự do (hệ số chặn)

  β_j: Hệ số hồi qui riêng

  U: Sai số ngẫu nhiên

- Hàm hồi quy tuyến tính bội mẫu có dạng:

  $\mathbf{y = a + b_1 x_1 + b_2 x_2 + \ldots b_k x_k + e}$

  $\mathbf{y_i = a + b_1 x_{1i} + b_2 x_{2i} + \ldots b_k x_{ki} + e_i}$

  Xác định các tham số của hàm hồi quy bội sử dụng phương pháp bình phương nhỏ nhất (OLS)

- Hàm hồi quy tuyến tính lý thuyết có dạng:

  $\mathbf{\hat{y} = a + b_1 x_1 + b_2 x_2 + \ldots b_k x_k}$

  $\mathbf{\hat{y}_i = a + b_1 x_{1i} + b_2 x_{2i} + \ldots b_k x_{ki}}$

  Phần dư: $\mathbf{e_i = y_i - \hat{y}_i \rightarrow \sum e_i^2 = \min}$

3.3. Mô hình hồi qui đa biến (HQ bội)

Ước lượng các tham số của mô hình hồi qui đa biến

- $\sum e_i^2 = \min \leftrightarrow \sum[y_i-(a + b_1x_{1i} + b_2x_{2i} + \ldots b_kx_{ki})]^2 = \min$
- Lấy đạo hàm theo (a) và (b_i) và cho bằng 0:

$$\begin{cases} a.n + b_1\sum x_{1i} + b_2\sum x_{2i} + \ldots + b_k\sum x_{ki} = \sum y_i \\ a\sum x_{1i} + b_1\sum x_{1i}^2 + b_2\sum x_{1i}x_{2i} + \ldots + b_k\sum x_{1i}x_{ki} = \sum x_{1i}y_i \\ a\sum x_{2i} + b_1\sum x_{2i}x_{1i} + b_2\sum x_{2i}^2 + \ldots + b_k\sum x_{2i}x_{ki} = \sum x_{2i}y_i \\ a\sum x_{3i} + b_1\sum x_{3i}x_{1i} + b_2\sum x_{3i}x_{2i} + \ldots + b_k\sum x_{3i}x_{ki} = \sum x_{3i}y_i \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a\sum x_{ki} + b_1\sum x_{ki}x_{1i} + b_2\sum x_{ki}x_{2i} + \ldots + b_k\sum x_{ki}^2 = \sum x_{ki}y_i \end{cases}$$

- Giải hệ phương trình tìm (a) và các (b_i) là các hệ số của hàm hồi quy bội
- Các hệ số có thể tìm được bằng phương pháp ma trận (Correlation Matrix) hoặc các phần mềm Excel, Eviews...

3.3. Mô hình hồi qui đa biến (HQ bội)

Kiểm định mô hình hồi quy đa biến

- **Hệ số xác định: $0 \le R^2 \le 1$**
  - Hệ số xác định R^2 đo lường phần biến thiên của Y có thể được giải thích bởi các biến độc lập X.
  - Đại lượng thể hiện sự thích hợp của mô hình hồi quy bội đối với dữ liệu
  - R^2 càng lớn thì mô hình hồi quy bội đã xây dựng được xem là càng thích hợp và càng có ý nghĩa trong việc giải thích sự biến thiên của Y

$$\boldsymbol{R}^2 = \frac{\boldsymbol{SSR}}{\boldsymbol{TSS}} = 1 - \frac{\boldsymbol{SSE}}{\boldsymbol{TSS}} = 1 - \frac{\sum (y_i - \hat{y})^2}{\sum (y_i - \bar{y})^2}$$

- Hệ số tương quan bội $\boldsymbol{R} = \sqrt{\boldsymbol{R}^2}$

3.3. Mô hình hồi qui đa biến (HQ bội)

Kiểm định mô hình hồi quy đa biến

- **Hệ số xác định điều chỉnh:** $\overline{R}^2$
  - Đo lường mức độ thích hợp của mô hình hồi quy bội
  - Khi tăng thêm số biến độc lập X vào mô hình, R^2 tăng. Cần xác định xem có nên đưa thêm một biến độc lập (giải thích) X_j nào đó vào mô hình hay không cần sử dụng Hệ số xác định điều chỉnh

$$\overline{R}^2 = 1 - \frac{SSE/[n-(k+1)]}{TSS/(n-1)}$$

k+1 : Số tham số trong mô hình kể cả hệ số cố định

  - Nếu Hệ số xác định điều chỉnh tăng lên, việc đưa thêm biến X_j vào mô hình làm tăng ý nghĩa mô hình, cần thiết để biến X_j trong mô hình
  - Đánh giá tầm quan trọng tương đối của biến độc lập cần xem độ tăng của R^2 khi một biến được đưa thêm vào mô hình
  - Mức tăng **R^2 thay đổi = R^2 – $R^2_{(j)}$** Với **$R^2_{(j)}$** là bình phương hệ số tương quan bội khi chưa có biến X_j
  - Mức độ thay đổi **R^2** lớn do 1 biến đưa thêm vào cho thấy biến này cung cấp những thông tin về biến phụ thuộc mà các biến khác không có

3.3. Mô hình hồi qui đa biến (HQ bội)

Kiểm định mô hình hồi quy đa biến

- **Kiểm định F:**
  - Dùng để kiểm định giả thuyết về sự tồn tại mối liên hệ tuyến tính giữa biến phụ thuộc Y với bất kỳ biến độc lập X_j nào đó
  - Giả thuyết **H_0: $b_1 = b_2 = ...= b_k = 0$**

    H_1: Có ít nhất một b_j nào đó khác 0 (Không phải tất cả b_j =0)
  - **Chấp nhận H_0:** Không tồn tại mối liên hệ tuyến tính giữa **Y** và bất kỳ một biến **X_j** nào
  - **Bác bỏ H_0:** Có mối liên hệ tuyến tính giữa **Y** với **ít nhất 1** trong các biến **X_j**
  - **Giá trị kiểm định:**

$$F_{k,n-(k+1)} = \frac{MSR}{MSE} = \frac{SSR/k}{SSE/[n-(k+1)]} = \frac{\sum(\hat{y}-\bar{y})^2/k}{\sum(y_i-\hat{y})^2/[n-(k+1)]}$$

$$F_{k,n-(k+1)} = \frac{R^2/k}{(1-R^2)/[n-(k+1)]}$$

- **Tiêu chuẩn kiểm định: $F_{k,n-(k+1)} > F_{\alpha, k,n-(k+1)}$** **Bác bỏ H_0**

  $F_{k,n-(k+1)} < F_{\alpha, k,n-(k+1)}$ **Chấp nhận H_0**

3.3. Mô hình hồi qui đa biến (HQ bội)

Dự báo bằng phương pháp hồi quy bội

- **Khoảng tin cậy của (a):**

  $a \pm t_{n-(k+1),\ \alpha/2}.Se_a$ **k**: Số biến độc lập của mô hình

- **Khoảng tin cậy của (b_i):**

  $b \pm t_{n-(k+1),\ \alpha/2}.Se_{bi}$ $Se_a = \sqrt{Var(a)}$ $Se_b = \sqrt{Var(b_i)}$

- Các sai số chuẩn của **(a)** và **(b_i)** có thể được tính từ phương pháp ma trận hoặc phần mềm Excel, Eviews...

- Phương sai của **(a)** và **(b_i)** là các thành phần nằm trên đường chéo ma trận: **$\{\sum(y_i-\hat{y})^2/[n-(k+1)]\}*(X'X)^{-1}$**

$$X = \begin{bmatrix} 1 & x_{11} & x_{21} & \cdots & x_{k1} \\ 1 & x_{12} & x_{22} & \cdots & x_{k2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{kn} \end{bmatrix} \quad X' = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ x_{k1} & x_{k2} & \cdots & x_{kn} \end{bmatrix} \quad X'X = \begin{bmatrix} n & \sum x_{1i} & \sum x_{2i} & \cdots & \sum x_{ki} \\ \sum x_{1i} & \sum x_{1i}^2 & \sum x_{1i}x_{2i} & \cdots & \sum x_{1i}x_{ki} \\ \sum x_{2i} & \sum x_{2i}x_{1i} & \sum x_{2i}^2 & \cdots & \sum x_{2i}x_{ki} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \sum x_{ki} & \sum x_{ki}x_{1i} & \sum x_{ki}x_{2i} & \cdots & \sum x_{ki}^2 \end{bmatrix}$$

3.3. Mô hình hồi qui đa biến (HQ bội)

Dự báo bằng phương pháp hồi quy bội

- **Dự báo điểm**

$\hat{y}^{DBĐ}_0 = a + b_1x_{10} + b_2x_{20} + \ldots + b_kx_{k0}$

- **Dự báo khoảng giá trị trung bình của Y với độ tin cậy (1-α)**

$E(Y_0) = \hat{y}^{DBĐ}_0 \pm t_{n-(k+1),\, \alpha/2} Se_{\hat{y}}$

$$Se_{\hat{y}} = \sqrt{\frac{\sum (y_i - \hat{y})^2}{n-(k+1)}} \sqrt{x_0'(X'X)^{-1}x_0}$$

- **Dự báo khoảng giá trị của Y với độ tin cậy (1-α)**

$\hat{y}^{DBK}_0 = \hat{y}^{DBĐ}_0 \pm t_{n-(k+1),\, \alpha/2} Se_{(y-\hat{y})}$

$$Se_{(y-\hat{y})} = \sqrt{\frac{\sum (y_i - \hat{y}_0)^2}{n-(k+1)}} \sqrt{1 + x_0'(X'X)^{-1}x_0}$$

$$x_0 = \begin{bmatrix} 1 \\ x_{10} \\ x_{20} \\ \vdots \\ x_{k0} \end{bmatrix} \quad x_0' = \begin{bmatrix} 1 & x_{10} & x_{20} & \cdots & x_{k0} \end{bmatrix} \quad X = \begin{bmatrix} 1 & x_{11} & x_{21} & \cdots & x_{k1} \\ 1 & x_{12} & x_{22} & \cdots & x_{k2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{kn} \end{bmatrix} \quad X' = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ x_{k1} & x_{k2} & \cdots & x_{kn} \end{bmatrix}$$

3.3. Mô hình hồi qui đa biến (HQ bội)

Các dạng hàm hồi quy bội phi tuyến

- **Dạng hàm Cobb-Douglas: $Y = aK^{\alpha}L^{\beta}$**
- **Dạng hàm KLEM : $Y = aK^{\alpha}L^{\beta}E^{\gamma}M^{\eta}$**
- **Dạng hàm mũ : $Y = e^{a+b_1x_1+b_2x_2+\ldots b_kx_k}$**
- **Dạng hàm hypecbol : $Y = a + b_1/x_1 + b_2/x_2 + \ldots b_k/x_k$**
- **Dạng hàm logarit : $\ln Y = a + b_1\ln x_1 + b_2\ln x_2 + \ldots . b_k\ln x_k$**
- **Dạng parabol : $Y = a + b_1X + b_2X^2 \leftrightarrow Y = a + b_1X + b_2V$**

3.3. Mô hình hồi qui đa biến (HQ bội)

Ví dụ dự báo bằng phương pháp hồi quy bội

Năm	TSL NNghiệp (10^6NT$)	Ngày lao động (10^6ngày)	Vốn ĐT (10^6NT$)
1	16607.7	275.5	17803.7
2	17511.3	274.4	18096.8
3	20171.2	269.7	18271.8
4	20932.9	267.0	19167.3
5	20406.0	267.8	19647.6
6	20831.6	275.0	20803.5
7	24806.3	283.0	22076.6
8	26465.8	300.7	23445.2
9	27403.0	307.5	24939.0
10	28628.7	303.7	26713.7
11	29904.5	304.7	29957.8
12	27508.2	298.6	31585.9
13	29035.5	295.5	33474.5
14	29281.5	299.0	34821.8
15	31535.8	288.1	41794.3

- $Y = aK^{\alpha}L^{\beta}$

lnY= -3.3384 +1.4988lnL+0.4899lnK

Se (2.4495) (0.5398) (0.1020)

t (-1.3629) (2.7765) (4.8005)

R^2 = 0.8890 df = 12

$\bar{R}^2_{đchỉnh}$ = 0.8705

- Nếu giữ nguyên vốn đầu tư, khi tăng 1% lao động đầu vào sẽ tăng trung bình 1.5% tổng sản lượng ngành nông nghiệp.
- Nếu giữ nguyên lao động, tăng 1% vốn đầu tư sẽ làm tăng gần 0.5% tổng sản lượng ngành nông nghiệp.
- Ngành nông nghiệp quốc gia này thu nhập tăng theo quy mô.

3.3. Mô hình hồi qui đa biến (HQ bội)

Ví dụ dự báo bằng phương pháp hồi quy bội

Y	X1	X2	X3	X4
142000.0	2310.0	2.0	2.0	20.0
144000.0	2333.0	2.0	2.0	12.0
151000.0	2356.0	3.0	1.5	33.0
150000.0	2379.0	3.0	2.0	43.0
139000.0	2402.0	2.0	3.0	53.0
169000.0	2425.0	4.0	2.0	23.0
126000.0	2448.0	2.0	1.5	99.0
142900.0	2471.0	2.0	2.0	34.0
163000.0	2494.0	3.0	3.0	23.0
169000.0	2517.0	4.0	4.0	55.0
149000.0	2540.0	2.0	3.0	22.0

SUMMARY OUTPUT

Regression Statistics	
Multiple R	**0.99837**
R Square	**0.99675**
Adjusted R Square	**0.99458**
Standard Error	**970.57846**
Observations	**11**

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	**4**	**1732393319**	**4.33E+08**	**459.7537**	**1.37231E-07**
Residual	**6**	**5652135**	**942022.6**		
Total	**10**	**1738045455**			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	**52317.8305**	**12237.3616**	**4.2753**	**0.0052**	**22374.0635**	**82261.5975**
X1	**27.6414**	**5.4294**	**5.0911**	**0.0022**	**14.3562**	**40.9266**
X2	**12529.7682**	**400.0668**	**31.3192**	**0.0000**	**11550.8392**	**13508.6972**
X3	**2553.2107**	**530.6692**	**4.8113**	**0.0030**	**1254.7091**	**3851.7122**
X4	**-234.2372**	**13.2680**	**-17.6543**	**0.0000**	**-266.7028**	**-201.7715**

3.3. Mô hình hồi qui đa biến (HQ bội)

Ví dụ dự báo bằng phương pháp hồi quy bội

Dependent Variable: Y
Method: Least Squares
Date: 11/26/09 Time: 22:28
Sample: 2002 2012
Included observations: 11

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	52317.83	12237.36	4.275254	0.0052
X1	27.64139	5.429374	5.091082	0.0022
X2	12529.77	400.0668	31.31919	0.0000
X3	2553.211	530.6692	4.811304	0.0030
X4	-234.2372	13.26801	-17.65428	0.0000

R-squared	0.996748	Mean dependent var	149536.4
Adjusted R-squared	0.994580	S.D. dependent var	13183.50
S.E. of regression	970.5785	Akaike info criterion	16.89662
Sum squared resid	5652135.	Schwarz criterion	17.07748
Log likelihood	-87.93139	F-statistic	459.7537
Durbin-Watson stat	1.796800	Prob(F-statistic)	0.000000

3.3. Mô hình hồi qui đa biến (HQ bội)

Ví dụ dự báo bằng phương pháp hồi quy bội

Năm	Dân số	GDP	Xuất khẩu	Nhập khẩu	Tiêu dùng CP
1980	64.77	125571.00	2566.00	1946.00	6671.00
1981	66.02	131968.00	2752.00	2404.00	9186.00
1982	67.24	139634.00	2338.00	2087.00	12081.00
1983	68.45	151782.00	2541.00	2581.00	23711.00
1984	69.64	164043.00	3924.00	2985.00	37010.00
1985	70.82	178534.00	5826.00	4054.00	44655.00
1986	72.00	195567.00	8155.00	5449.00	54589.00
1987	73.16	213833.00	11144.00	7256.00	62889.00
1988	74.31	231264.00	11592.00	9185.00	70749.00
1989	75.46	244596.00	11500.00	9360.00	73419.00
1990	76.60	256272.00	11742.00	11541.00	84817.00
1991	77.64	273666.00	15637.00	14483.00	103151.00
1992	78.69	292535.00	16218.00	15029.00	119430.00
1993	79.73	313247.00	19746.00	16706.00	133877.00
1994	80.90	336243.00	25256.00	20149.00	178541.00
1995	82.03	362435.00	31969.00	26485.00	221792.00
1996	83.11	392989.00	36761.00	32447.00	237410.00

3.3. Mô hình hồi qui đa biến (HQ bội)

Ví dụ dự báo bằng phương pháp hồi quy bội

Dependent Variable: GDP

Method: Least Squares

Date: 11/27/09 Time: 11:21

Sample: 1980 1996

Included observations: 17

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-463952.7	35077.46	-13.22652	0.0000
DSO	8879.637	517.4797	17.15939	0.0000
NK	2.347581	0.935706	2.508889	0.0275
XK	1.413069	1.116749	1.265341	0.2298
TDCP	-0.030349	0.141608	-0.214315	0.8339

R-squared	0.998429	Mean dependent var	235539.9
Adjusted R-squared	0.997906	S.D. dependent var	83579.44
S.E. of regression	3824.853	Akaike info criterion	19.57636
Sum squared resid	1.76E+08	Schwarz criterion	19.82142
Log likelihood	-161.3990	F-statistic	1906.983
Durbin-Watson stat	1.162508	Prob(F-statistic)	0.000000