

# Chương 7. Hồi quy trong sinh thái QXTV

## 7.1. Đặt vấn đề

### 7.1.1. Mục đích và ứng dụng hồi quy

- ✓ Trong sinh thái học (QXTV, Cảnh quan, Côn trùng, Bệnh cây...), chúng ta cần phải phân tích mối liên hệ giữa các loài với các yếu tố môi trường, dựa trên những quan sát loài và các biến môi trường ở một chuỗi lập địa nào đó.

## Chương 7. Hồi quy trong sinh thái QXTV

- ✓ Khi phân tích mối liên hệ giữa các loài với các yếu tố môi trường, loài được ghi nhận ở dạng độ phong phú (Abundences) hoặc chỉ đơn giản là sự có mặt loài.

## Chương 7. Hồi quy trong sinh thái QXTV

- ✓ Trong phân tích hồi quy, chúng ta chỉ có thể phân tích số liệu trên từng loài riêng biệt.
- ✓ Mỗi hồi quy nhắm vào một loài cụ thể và xem xét loài này có mối quan hệ với môi trường như thế nào.

## Chương 7. Hồi quy trong sinh thái QXTV

- Trong phân tích hồi quy:
- ✓ Độ phong phú hoặc sự có mặt của loài là biến phản hồi;
- ✓ Các biến môi trường là các biến giải thích.

## Chương 7. Hồi quy trong sinh thái QXTV

- ✓ Thuật ngữ “biến phản hồi” xuất phát từ ý tưởng cho rằng, các loài phản ứng lại hoặc phản hồi lại những tác động của môi trường (biến môi trường) theo cách thức nhân quả nào đó.

## Chương 7. Hồi quy trong sinh thái QXTV

- Từ phân tích hồi quy, nhà nghiên cứu có thể tìm ra nguyên nhân gây ra phản hồi của loài với môi trường hay không?
- Trả lời: Không

## Chương 7. Hồi quy trong sinh thái QXTV

- Vậy phân tích hồi quy chỉ trả lời được câu hỏi nào?
- **Đó là câu hỏi:** Loài có quan hệ với biến (yếu tố) môi trường trong chuỗi lập địa này hay không?
- ✓ **Khi mô hình tồn tại,** chúng ta nói “Loài có mối liên hệ với biến môi trường trong chuỗi lập địa này”.

## Chương 7. Hồi quy trong sinh thái QXTV

- Mục đích của phân tích hồi quy là gì?
- ✓ Mục đích của phân tích hồi quy là mô tả biến phản hồi như là một hàm số của một hoặc nhiều biến giải thích.
- ✓ Bằng cách phân tích hồi quy, chúng ta cố gắng tạo ra sai số nhỏ nhất.
- ✓ Giá trị dự đoán bằng hàm phản hồi là phản hồi kỳ vọng – đó là phản hồi với sai số trung bình nhỏ nhất.



## Chương 7. Hồi quy trong sinh thái QXTV

- **Trong sinh thái học**, phân tích hồi quy được sử dụng để giải quyết những vấn đề gì?
  - a. **Ước lượng các tham số sinh thái**. Ví dụ: Biên độ tối ưu sinh thái và biên độ sinh thái của loài.
  - b. **Đánh giá những biến môi trường nào** đóng góp chủ yếu vào phản hồi của loài và **biến môi trường nào xuất hiện không quan trọng**. Việc đánh giá được thực hiện thông qua kiểm định mức ý nghĩa.
  - c. **Dự đoán những phản hồi của loài** (*độ phong phú hoặc sự có mặt – vắng mặt*) ở chuỗi lập địa từ những giá trị thu thập của một hoặc nhiều biến môi trường.
  - d. **Dự đoán những giá trị của một hoặc nhiều biến môi trường** ở chuỗi lập địa từ những giá trị thu thập của một hoặc nhiều loài. Những dự đoán này được gọi là **chẩn đoán (Calibration)**.

## Chương 7. Hồi quy trong sinh thái QXTV

### 7.1.2. Mô hình phản hồi và kiểu biến phản hồi

- **Mô hình phản hồi** bao gồm hai thành phần: hệ thống và sai số.

## Chương 7. Hồi quy trong sinh thái QXTV

- ✓ **Thành phần hệ thống** mô tả cách thức mà biến phản hồi phụ thuộc vào biến giải thích. Thành phần hệ thống chỉ rõ bằng một hàm hồi quy.
- ✓ **Thành phần sai số** mô tả cách thức mà phản hồi quan sát sai lệch với phản hồi kỳ vọng. Thành phần sai số có thể diễn tả bằng phân bố của sai số.

## Chương 7. Hồi quy trong sinh thái QXTV

**Ví dụ:**

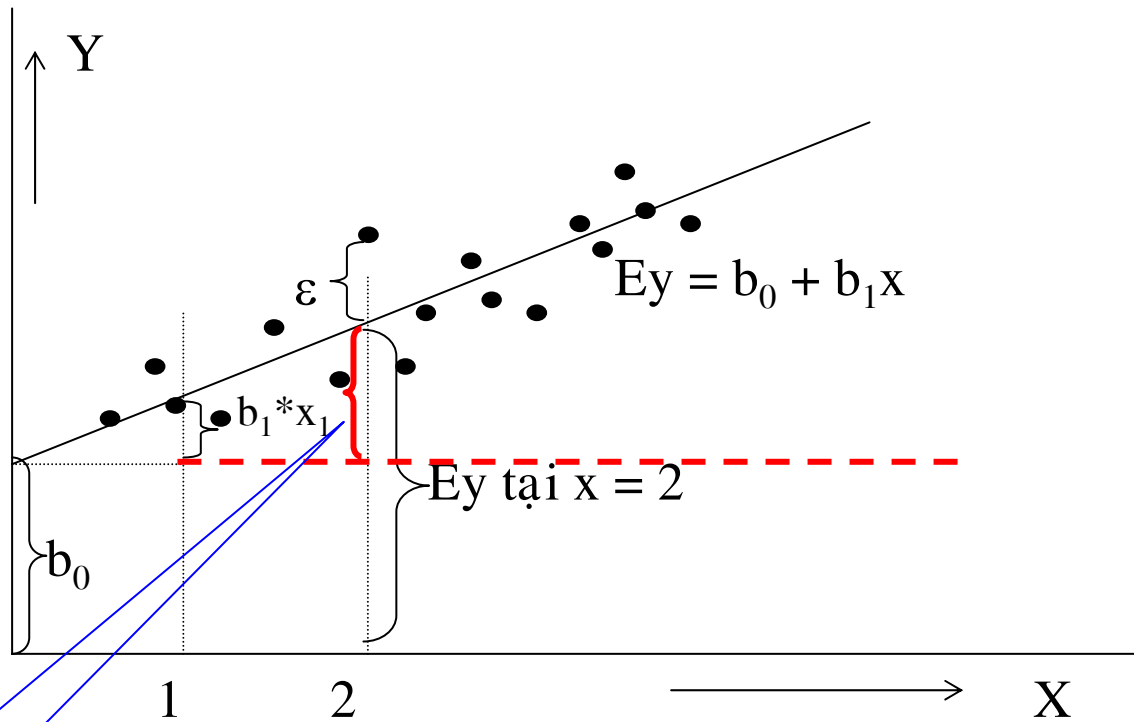
- ✓ **Khi làm phù hợp một đường thẳng với số liệu, mô hình phản hồi có dạng:**

$$y = b_0 + b_1x + \varepsilon \quad (7.1)$$

**trong đó:**

- ✓ **y = biến phản hồi**
- ✓ **x = biến giải thích**
- ✓  **$\varepsilon$  = sai số**
- ✓  **$b_0$  và  $b_1$  = hệ số ( $b_0$  = điểm chặn,  $b_1$  = độ dốc)**

# Chương 7. Hồi quy trong sinh thái QXTV



**Hình 7.1.** Mô hình phản hồi ở dạng tuyến tính

## Chương 7. Hồi quy trong sinh thái QXTV

- Mục đích phân tích hồi quy trong sinh thái học là gì ?
- ✓ Đó là xác định thành phần hệ thống và thành phần sai số của mô hình từ toàn bộ số liệu quan sát.

## Chương 7. Hồi quy trong sinh thái QXTV

- ✓ Ở dạng chung nhất, các tham số hồi quy ( $b_0$  và  $b_1$ ) được xác định theo nguyên lý “Min{Sum  $(y^* - y)^2$ ”.
- ✓ Phân bố của sai số  $\varepsilon$  được giả định là phân bố ngẫu nhiên chuẩn.

## Chương 7. Hồi quy trong sinh thái QXTV

- Độ phong phú của loài phân bố theo kiểu nào?
- ✓ Đó là phân bố lệch.
- ✓ Phân bố lệch của độ phong phú biểu hiện ở chỗ, những giá trị nhỏ thì nhiều, còn giá trị lớn thì ít.
- ✓ Dạng phân bố lệch của độ phong phú xuất hiện ngay cả ở những nơi có điều kiện môi trường khá thuận nhất.



## Chương 7. Hồi quy trong sinh thái QXTV

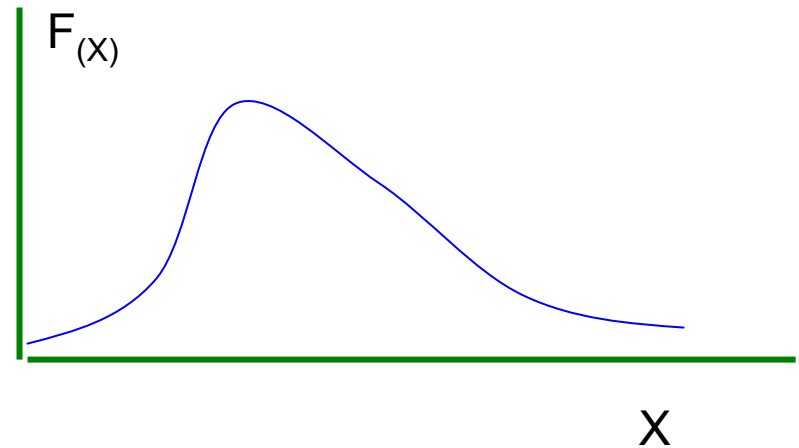
- ✓ Phân bố lệch của độ phong phú có thể mô tả bằng **phân bố log-normal**.

## Chương 7. Hồi quy trong sinh thái QXTV

- Phân bố log-normal có dạng:

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right)$$

$\mu$  và  $\sigma$  = tương ứng là kỳ vọng toán và độ lệch chuẩn trung bình của biến đã được biến đổi Y



## Chương 7. Hồi quy trong sinh thái QXTV

- Khi biến đổi độ phong phú bằng cách lấy logarit cơ số tự nhiên ( $\log_e$ ), thì phân bố của chúng sẽ trở thành phân bố chuẩn (Williamson, 1972).

## Chương 7. Hồi quy trong sinh thái QXTV

- ✓ Do đó, khi phân tích những giá trị độ phong phú bằng hồi quy bình phương nhỏ nhất, thì việc biến đổi độ phong phú bằng cách lấy logarit là cần thiết.
- Khi độ phong phú lấy giá trị zero, thì logarit của zero là không xác định.
- ✓ Do đó, đối với độ phong phú đo đạc ở dạng “có mặt loài = 1” và “vắng mặt loài = 0” thì kỹ thuật phân tích hồi quy thích hợp là **hồi quy logit**.

## Chương 7. Hồi quy trong sinh thái QXTV

- ✓ Hồi quy logit biểu thị xác suất bắt gặp loài là một hàm số của biến độc lập (biến giải thích, biến môi trường).

## Chương 7. Hồi quy trong sinh thái QXTV

### 7.1.3. Những kiểu biến giải thích và những kiểu đường cong phản hồi

- Biến giải thích có 3 kiểu:
  - ✓ Biến định danh (Nominal variable)
  - ✓ Biến có thứ bậc (Ordinal variable)
  - ✓ Biến định lượng (Quantitative variable)

## Chương 7. Hồi quy trong sinh thái QXTV

- ✓ **Kỹ thuật phân tích hồi quy** có thể thực hiện dễ dàng đối với những biến môi trường định lượng và định danh, **nhưng không thể giải quyết được đối với biến có thứ bậc.**

## Chương 7. Hồi quy trong sinh thái QXTV

- ✓ Đối với những biến có thứ bậc nhỏ, cách giải quyết là chuyển thành biến định danh.
- ✓ Đối với những biến có thứ bậc lớn, cách giải quyết là chuyển thành biến định lượng.
- ✓ Hồi quy đối với một biến giải thích định lượng bao gồm việc làm phù hợp số liệu với một đường cong.

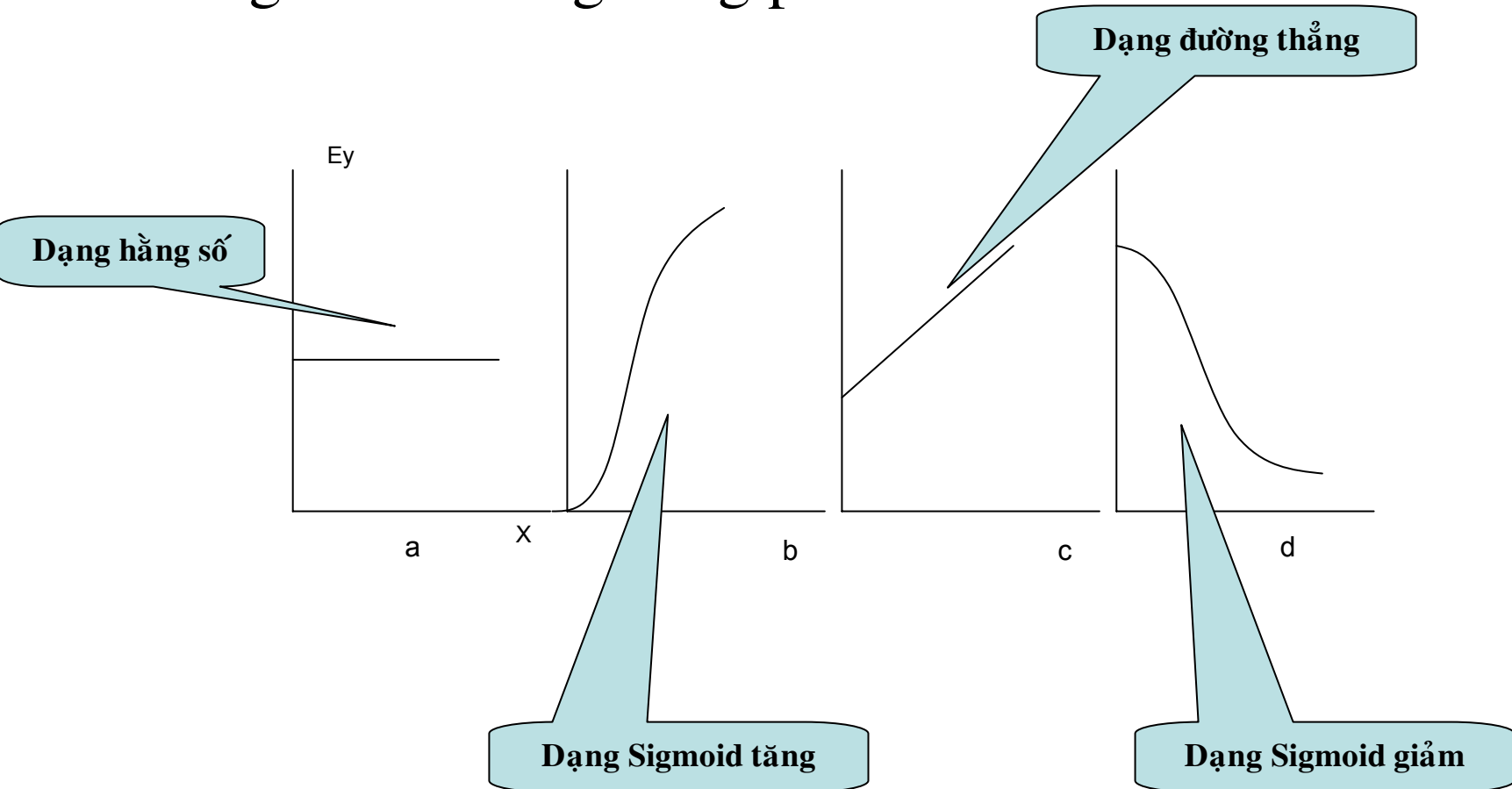


## Chương 7. Hồi quy trong sinh thái QXTV

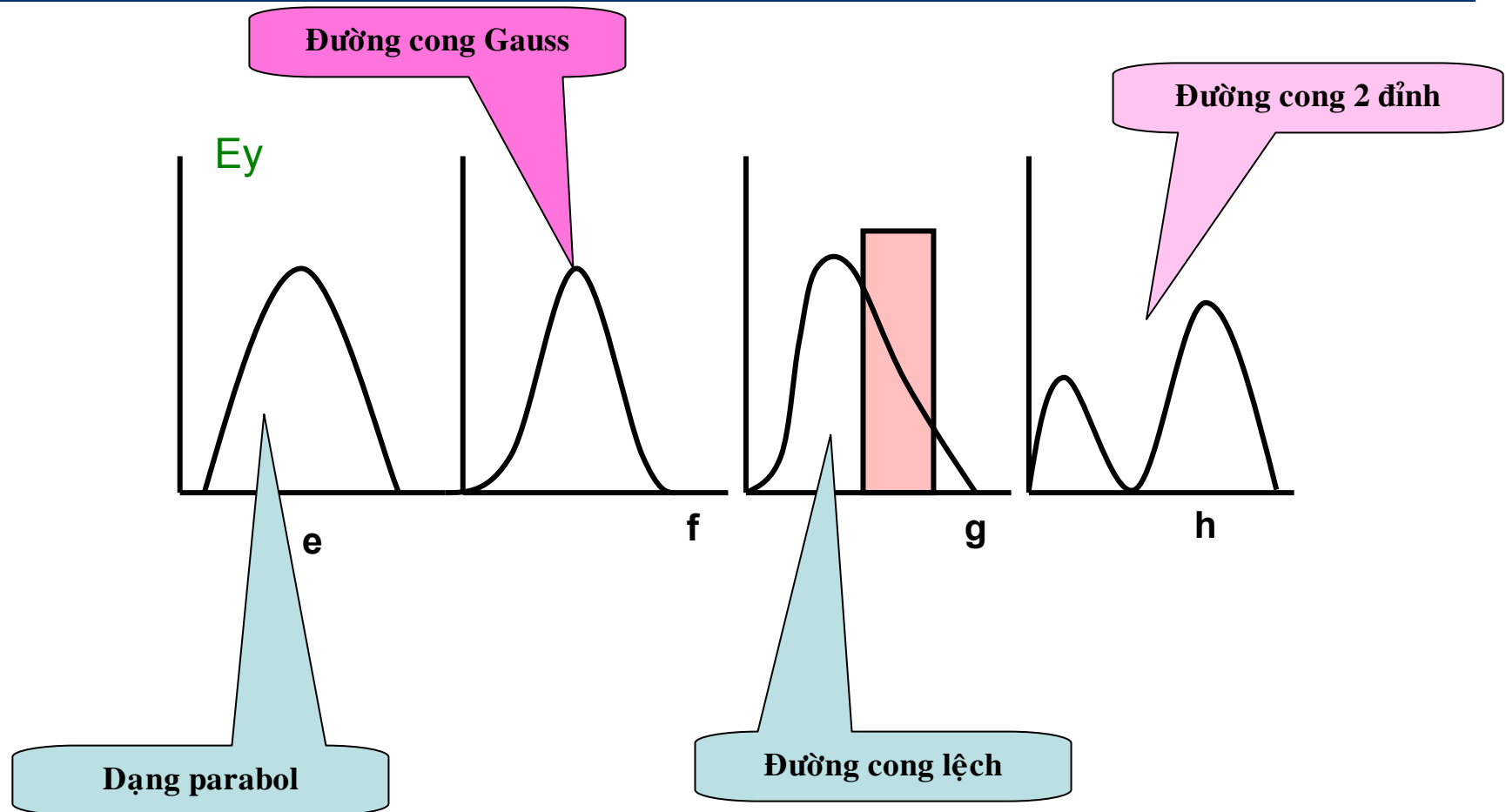
- ✓ Để mô tả quan hệ giữa các biến phản hồi với các biến giải thích, chúng ta cần chọn dạng đường cong nào là thích hợp với số liệu?
- ✓ Những cách thức nào cho phép tìm kiếm mô hình phản hồi thích hợp?
  - Sử dụng đồ thị phân tán.
  - Dựa theo lý thuyết.

# Chương 7. Hồi quy trong sinh thái QXTV

## Những kiểu đường cong phản hồi



# Chương 7. Hồi quy trong sinh thái QXTV



## Chương 7. Hồi quy trong sinh thái QXTV

### Chương 7 giới thiệu 2 vấn đề chính:

1. Kỹ thuật hồi quy dùng trong phân tích số liệu dạng bất gặp loài (kí hiệu 1) & không bất gặp loài (kí hiệu 0).
2. Kỹ thuật hồi quy dùng trong phân tích số liệu độ phong phú định lượng (hồi quy bình phương nhỏ nhất).

## Chương 7. Hồi quy trong sinh thái QXTV

- ❖ Trong cả hai phần, trước hết giới thiệu những mô hình mà biến giải thích là biến định danh (định tính), kể đến là những mô hình với biến giải thích định lượng.

## Chương 7. Hồi quy trong sinh thái QXTV

- ❖ Đối với dữ liệu độ phong phú, chúng ta nhận được các đường cong:
  - ✓ dạng mũ
  - ✓ dạng Gauss

## Chương 7. Hồi quy trong sinh thái QXTV

- Đối với dữ liệu dạng bất gặp – không bất gặp loài, chúng ta nhận được các đường cong:
  - ✓ Sigmoid
  - ✓ Logit Gauss

## Chương 7. Hồi quy trong sinh thái QXTV

- Đường cong parabol cho phép ước lượng:
  - ✓ Giá trị chỉ thị (tối ưu) của loài.
  - ✓ Biên độ sinh thái (tính chống chịu) của loài.



## Chương 7. Hồi quy trong sinh thái QXTV

- Hồi quy đa biến có thể được vận dụng để nghiên cứu ảnh hưởng của nhiều biến môi trường (cả biến định lượng lẫn biến định tính) đến biến phản hồi của loài.

## Chương 7. Hồi quy trong sinh thái QXTV

### 7.2. Hồi quy đối với số liệu định lượng

#### 7.2.1. Thuật toán phân tích hồi quy

- Thuật toán phân tích hồi quy bao gồm những bước nào?
- Có 5 bước cơ bản:

## Chương 7. Hồi quy trong sinh thái QXTV

1. Ước lượng các tham số của mô hình bằng phương pháp bình phương nhỏ nhất.
2. Tính các thống kê mô tả (trung bình, sai tiêu chuẩn, khoảng tin cậy...).
3. Kiểm định sự tồn tại của mô hình bằng thống kê t hoặc F (phân tích ANOVA).
4. Ước lượng khoảng tin cậy 95% của các tham số (phản hồi kỳ vọng) và khoảng dự đoán 95% của những phản hồi mới.
5. Tính hệ số tương quan.

## Chương 7. Hồi quy trong sinh thái QXTV

### 7.2.2. Đường thẳng

- ✓ Mỗi quan hệ giữa độ phong phú của loài cây với độ ẩm môi trường đất (**biến giải thích tính theo phần trăm**) có thể ở dạng đường thẳng.

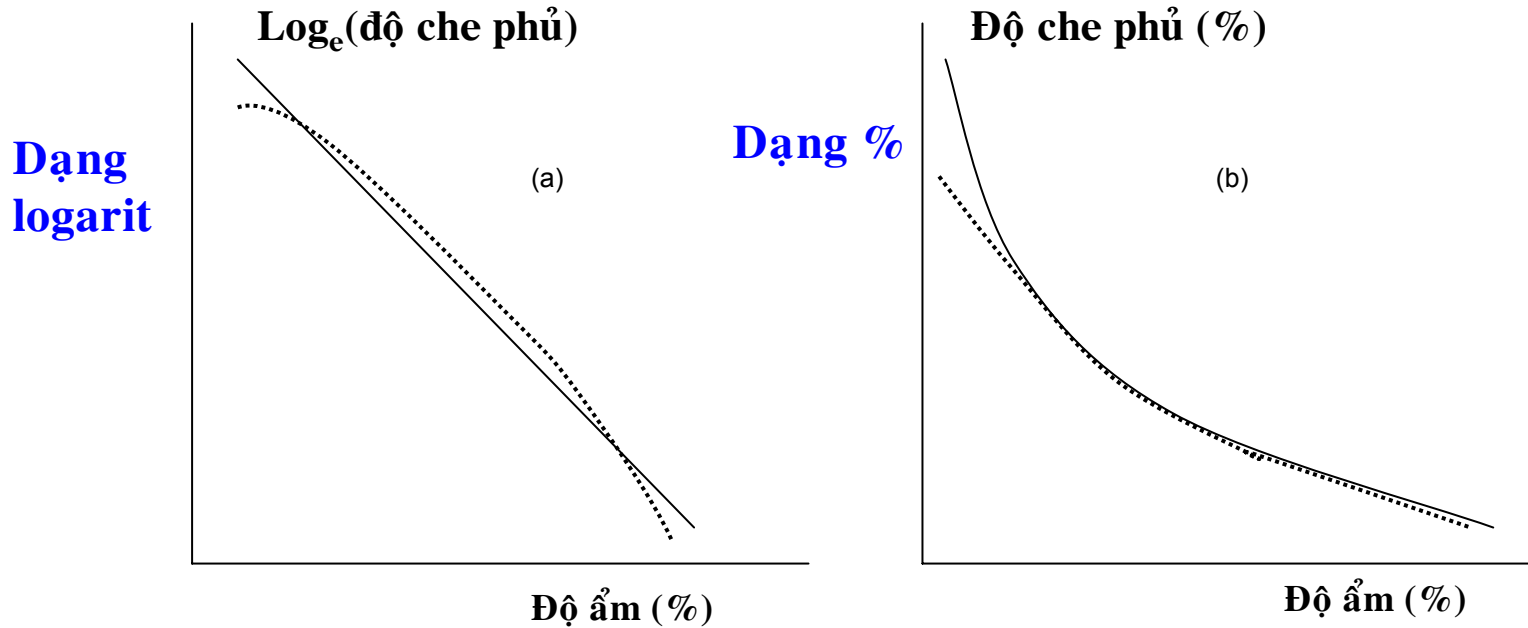
## Chương 7. Hồi quy trong sinh thái QXTV

- Mô hình phản hồi của loài có dạng:

$$E_y = b_0 + b_1x + \varepsilon \quad (7.2)$$

- ✓  $E_y$  = biến phản hồi kỳ vọng
- ✓  $x$  = biến giải thích
- ✓  $b_0$  và  $b_1$  = hệ số cần tìm.

# Chương 7. Hồi quy trong sinh thái QXTV



Hình 7.5. Quan hệ giữa độ phong phú của loài với độ ẩm đất

## Chương 7. Hồi quy trong sinh thái QXTV

- Xác định những tham số của mô hình
- ✓ Các tham số  $b_0$  và  $b_1$  của hàm hồi quy ước lượng được xác định bằng phương pháp bình phương nhỏ nhất.

## Chương 7. Hồi quy trong sinh thái QXTV

- ✓ Bản chất của phương pháp bình phương nhỏ nhất là làm nhỏ nhất tổng bình phương các sai lệch giữa giá trị của biến  $y_i$  và giá trị ước lượng của biến  $y_i^*$ .
- ✓ Tiêu chuẩn bình phương nhỏ:  $\min \text{Sum}(y_i - y_i^*)^2$ .
- ✓ Để  $\text{Sum}(y_i - y_i^*)^2$  nhỏ nhất, thì  $b_0$  và  $b_1$  được tính theo công thức:



## Chương 7. Hồi quy trong sinh thái QXTV

$$b_1 = \frac{\sum(x_i - x_{bq})(y_i - y_{bq})}{\sum(x_i - x_{bq})^2} \quad (7.3a)$$

$$b_0 = y_{bq} - b_1 x_{bq} \quad (7.3b)$$

$n$  = tổng số quan sát

$x_i$  = giá trị của biến độc lập tương ứng với quan sát  $i$

$y_i$  = giá trị của biến phụ thuộc tương ứng với quan sát  $i$

$x_{bq}$  = giá trị trung bình của biến độc lập

$y_{bq}$  = giá trị trung bình của biến phụ thuộc.

## Chương 7. Hồi quy trong sinh thái QXTV

### 7.2.3. Đường cong Parabol và đường cong Gauss

#### (1) Đường cong Parabol

$$E_y = b_0 + b_1x + b_2x^2 \quad (7.4)$$

Trong đó:

- ✓  $E_y$  = biến phản hồi kỳ vọng
- ✓  $x$  = biến giải thích
- ✓  $b_0, b_1$  và  $b_2$  = các hệ số cần tìm.

## Chương 7. Hồi quy trong sinh thái QXTV

### (2) Đường cong Gauss

- Khi làm phù hợp một parabol với độ phong phú đã chuyển sang dạng logarit, thực tế chúng ta làm phù hợp một đường cong phản hồi Gauss đối với số liệu độ phong phú gốc (tính theo %).

## Chương 7. Hồi quy trong sinh thái QXTV

- Đường cong phản hồi Gauss có dạng:

$$z = c * \exp[-0,5(x - u)^2/t^2] \quad (7.5)$$

- ✓  $z$  = giá trị độ phong phú ban đầu (%)
- ✓  $c$  = độ phong phú lớn nhất của loài
- ✓  $u$  = optimum (giá trị  $x$  cho độ phong phú lớn nhất)
- ✓  $t$  = tính chống chịu sinh thái (số đo biên độ sinh thái của loài).

## Chương 7. Hồi quy trong sinh thái QXTV

- ✓ Nếu lấy logarit cơ số tự nhiên hai vế của phương trình 7.4, chúng ta thu được:

$$\log_e(z) = \log_e(c) - 0,5(x - u)^2/t^2$$

$$\text{Hay } \log_e(z) = b_0 + b_1x + b_2x^2 \quad (7.6)$$

## Chương 7. Hồi quy trong sinh thái QXTV

- Khai triển các thành phần của **hàm 7.6** như sau:
  - ✓  $(x - u)^2 = x^2 - 2ux + u^2$
  - ✓  $b_0 = \log_e(c) - u^2/(2t^2)$
  - ✓  $b_1 = u/t^2;$
  - ✓  $b_2 = -1/(2t^2).$  (7.7)

## Chương 7. Hồi quy trong sinh thái QXTV

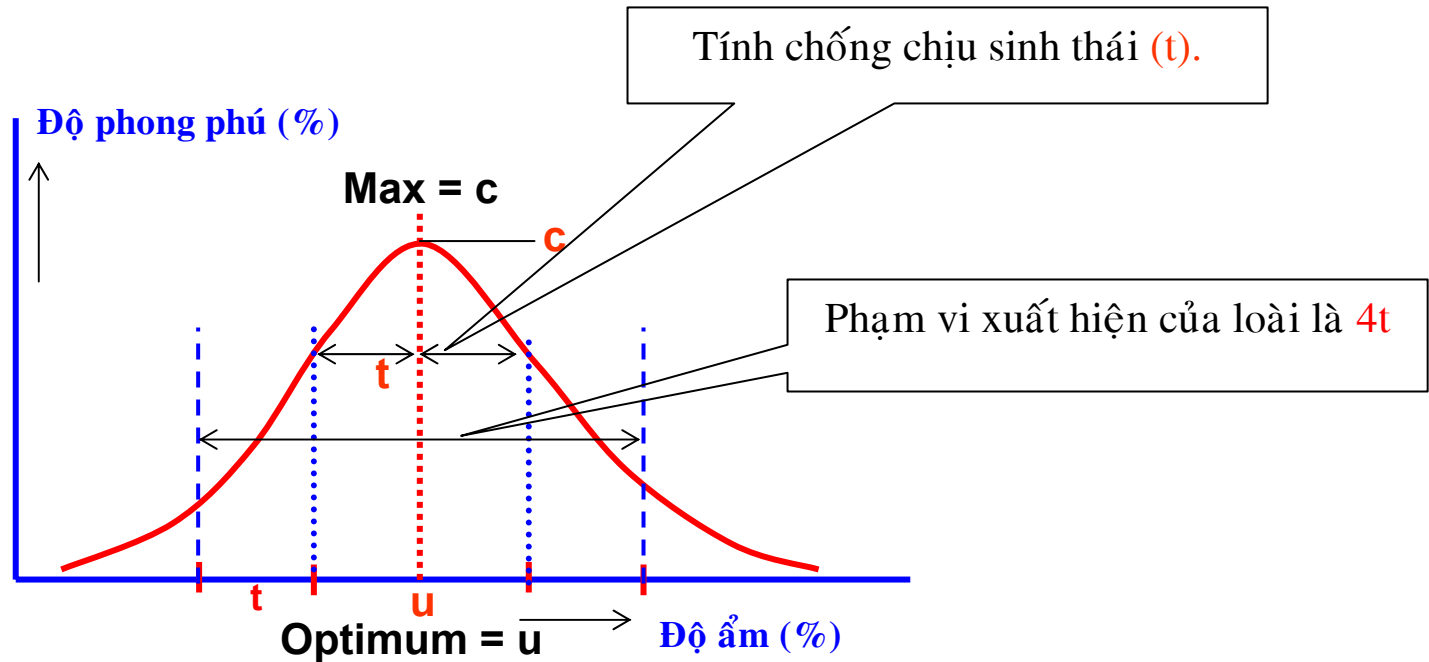
- Bằng việc làm phù hợp Parabol với  $\log_e$ (độ phong phú), chúng ta thu được các ước lượng bình phương nhỏ nhất  $b_0$ ,  $b_1$  và  $b_2$ .

## Chương 7. Hồi quy trong sinh thái QXTV

- Từ đó chúng ta cũng thu được các ước lượng sau đây:
  - ✓ tối ưu:  $u = -b_1/(2b_2)$  (7.8a)
  - ✓ tính chống chịu:  $t = 1/\sqrt{-2b_2}$  (7.8b)
  - ✓ lớn nhất:  $c = \exp(b_0 + b_1u + b_2u_2)$  (7.8c)
  - ✓ hay  $c = -(b_1^2 - 4b_2b_0)/4b^2$



# Chương 7. Hồi quy trong sinh thái QXTV



Hình 7.6. Đường cong phản hồi Gauss với các tham số sinh thái

## Chương 7. Hồi quy trong sinh thái QXTV

### ❖ Một số lưu ý:

1. Tối ưu sinh thái (**u**) là giá trị của biến giải thích đảm bảo cho loài xuất hiện nhiều nhất.
2. Tính chống chịu sinh thái (**t**) là số đo biên độ sinh thái của loài.
3. Độ phong phú lớn nhất (**c**) nhận được tương ứng với tối ưu sinh thái (**u**).

## Chương 7. Hồi quy trong sinh thái QXTV

4. Các hàm 7.8 có hệ số  $b_2 < 0$ .
5. Nếu ước lượng  $b_2 > 0$ , thì đường cong có một cực tiểu thay vì một cực đại.
6. Các tham số  $b_0$ ,  $b_1$  và  $b_2$  có thể xác định dễ dàng nhờ các phần mềm thống kê (Minitab, Statgraphics và SPSS...).

## Chương 7. Hồi quy trong sinh thái QXTV

- Ví dụ:
- Giữa tổng sinh khối khô (SKK, g/cây) của gỗ đở 6 tháng tuổi với hàm lượng phân tổng hợp NPK (X,%) tồn tại quan hệ rất chặt chẽ ( $R^2 = 99,2\%$ ) dưới dạng hàm bậc 2.

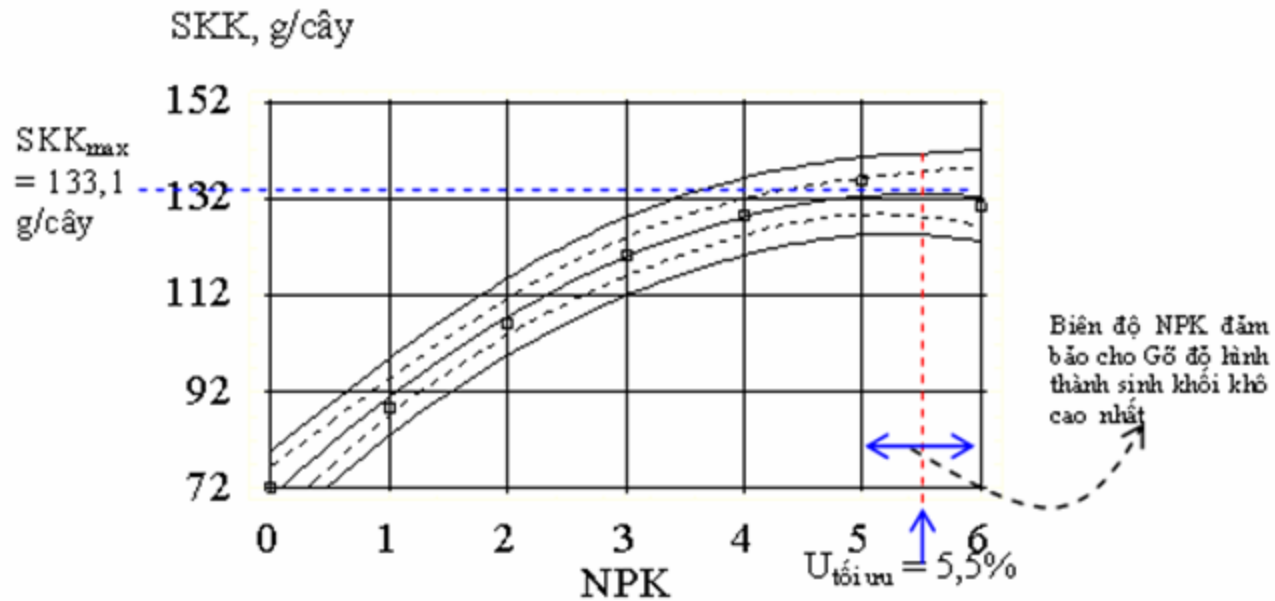
## Chương 7. Hồi quy trong sinh thái QXTV

- Phương trình mối quan hệ có dạng:

$$SKK = 70,4119 + 22,7107*X - 2,0583*X^2 \quad (**)$$

- ✓  $R^2 = 99,2\%$ ;
- ✓  $S = \pm 2,538$ ;
- ✓ Sai số tuyệt đối trung bình = 1,64

# Chương 7. Hồi quy trong sinh thái QXTV



Quan hệ giữa sinh khối khô (SKK, g/cây) của Gỗ độ 6 tháng tuổi với hàm lượng phân NPK, %

## Chương 7. Hồi quy trong sinh thái QXTV

- Khai triển hàm (\*\*) có thể xác định:
  - ✓  $U = 5,5\%$
  - ✓  $t = 0,5\%$
  - ✓  $U \pm t = 5\% - 6\%$
  - ✓  $U \pm 4t = [3,5 - 7,5]$
  - ✓  $C = 133,1 \text{ g/cây}$

## Chương 7. Hồi quy trong sinh thái QXTV

### 7.3. HỒI QUY ĐỐI VỚI DỮ LIỆU CÓ – KHÔNG CÓ: hồi quy logit

#### 7.3.1. Những biến giải thích định danh: kiểm định $\chi^2$

- ✓ **Bảng 7.1** chỉ ra số lượng đồng cỏ bắt gặp và không bắt gặp loài cây *Achillea ptarmica*.
- ✓ Đồng cỏ được chia thành 4 cấp (A, B, C, D) tùy theo biện pháp sử dụng.



## Chương 7. Hồi quy trong sinh thái QXTV

**Bảng 7.1.** Số lượng đồng cỏ bắt gặp loài cây *Achillea ptarmica*

	Đồng cỏ				
Loài cây	<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>	Tổng
Có	37	40	27	9	113
Không	109	356	402	558	1425
<b>Tổng</b>	<b>146</b>	<b>396</b>	<b>429</b>	<b>567</b>	<b>1538</b>
<b>Tần số</b>	<b>0,254</b>	<b>0,101</b>	<b>0,063</b>	<b>0,016</b>	<b>0,073</b>

## Chương 7. Hồi quy trong sinh thái QXTV

- Vấn đề đặt ra ở đây:
- Tần số xuất hiện loài cây *Achillea ptarmica* có phụ thuộc căn bản vào biện pháp sử dụng đất hay không?
- Hay biện pháp sử dụng đất có ảnh hưởng đến loài cây *Achillea ptarmica* hay không?

## Chương 7. Hồi quy trong sinh thái QXTV

- ✓ Khi biến phản hồi được ghi là 1 (có mặt) và 0 (vắng mặt), thì phản hồi kỳ vọng  $Ey$  là tần số kỳ vọng hay là xác suất xuất hiện loài cây *Achillea ptarmica*.
- ✓ Nếu xác suất xuất hiện loài cây *Achillea ptarmica* giống nhau ở cả 4 cấp, thì sự xuất hiện của loài không phụ thuộc vào biện pháp sử dụng đồng cỏ.

## Chương 7. Hồi quy trong sinh thái QXTV

- Vì thế, chúng ta đặt giả thuyết  $H_0^+$ :
- ✓ Xác suất xuất hiện loài cây *Achillea ptarmica* là giống nhau ở cả 4 kiểu sử dụng đồng cỏ.
- ✓ Hay 4 kiểu sử dụng đồng cỏ không ảnh hưởng đến sự xuất hiện loài cây *Achillea ptarmica*.
- ✓ **Đối thuyết  $H_0^-$** : Sự xuất hiện của loài phụ thuộc vào biện pháp sử dụng đồng cỏ.
- **Giả thuyết  $H_0$**  được kiểm định bằng thống kê  $\chi^2$ .

## Chương 7. Hồi quy trong sinh thái QXTV

- **Những lưu ý đối với kiểm định  $\chi^2$** 
  - a. Kiểm định  $\chi^2$  là kiểm định tính phù hợp.
  - b. Kiểm định  $\chi^2$  chỉ thích hợp với một tập dữ liệu lớn.
  - c. Kiểm định  $\chi^2$  rất nhạy cảm với hai đuôi của phân bố.
  - d. Kiểm định  $\chi^2$  chỉ được sử dụng khi 75% số ô của bảng  $R \times C$  có tần số lớn hơn 5.

## Chương 7. Hồi quy trong sinh thái QXTV

- Thủ tục kiểm định  $\chi^2$
- ✓ **Giả thuyết  $H_0$ :**
- ✓ (1) Sự xuất hiện của loài là độc lập với việc sử dụng đất.
- ✓ (2) Biện pháp sử dụng đất khác nhau có ảnh hưởng giống nhau đến sự xuất hiện của loài.

## Chương 7. Hồi quy trong sinh thái QXTV

### ✓ Quy tắc quyết định:

- Nếu  $\chi^2 > \chi^2_{(0,05 \text{ hay } 0,01)}$  hoặc  $P < 0,05$  (0,01) thì  $H_0^-$ .
- Nếu  $\chi^2 < \chi^2_{(0,05 \text{ hay } 0,01)}$  hoặc  $P > 0,05$  (0,01) thì  $H_0^+$ .

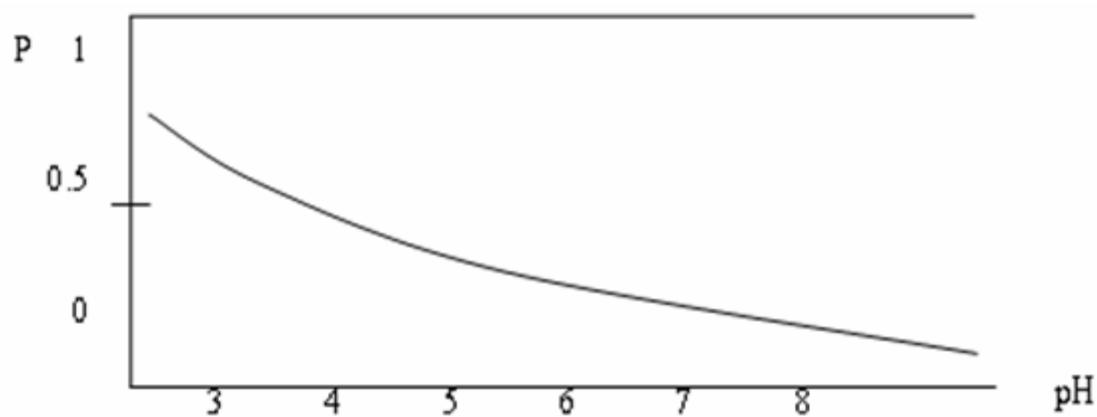
## Chương 7. Hồi quy trong sinh thái QXTV

### 7.3.2. Đường cong sigmoid

- Bây giờ chúng ta xét trường hợp sau đây:
- ✓ **Biến phản hồi ( $y$ )** là biến định danh không có thứ bậc hơn kém (dạng 1 và 0).
- ✓ **Biến giải thích ( $x$ )** là biến định lượng.
- ✓ Số liệu ở dạng này chỉ ra ở **hình 7.7**.



## Chương 7. Hồi quy trong sinh thái QXTV



Hình 7.7. Đường cong biểu diễn xác suất (p) xuất hiện loài trong quan hệ với pH môi trường

## Chương 7. Hồi quy trong sinh thái QXTV

- Khác với trường hợp ở mục 7.3.1, trong trường hợp này biến phản hồi kỳ vọng là xác suất xuất hiện loài ở một lập địa tương ứng với một biến môi trường định lượng nhất định.

## Chương 7. Hồi quy trong sinh thái QXTV

- Xác suất này được mô tả bằng một đường cong.
- Xác suất lấy giá trị giữa 0 và 1.

## Chương 7. Hồi quy trong sinh thái QXTV

- Số liệu trên đây có thể mô tả bằng hàm đường thẳng hay không?

$$E y = b_0 + b_1 x \quad (7.9)$$

## Chương 7. Hồi quy trong sinh thái QXTV

- Câu trả lời: **Không**
- Bởi vì hàm đường thẳng cũng có thể lấy giá trị âm.

## Chương 7. Hồi quy trong sinh thái QXTV

- Để  $Ey > 0$ , chúng ta có thể làm phù hợp số liệu với hàm số mũ?

$$Ey = \exp(b_0 + b_1x) \quad (7.10)$$

## Chương 7. Hồi quy trong sinh thái QXTV

- ✓ Hàm số mũ cũng không phù hợp, bởi vì  $Ey$  có thể nhận giá trị lớn hơn 1.

## Chương 7. Hồi quy trong sinh thái QXTV

- Do đó, để  $Ey$  nhận giá trị từ 0 – 1, chúng ta biến đổi hàm 7.10 thành đường cong sigmoid như sau:

$$Ey = p = \frac{\exp(b_0 + b_1x)}{1 + \exp(b_0 + b_1x)} \quad (7.11)$$

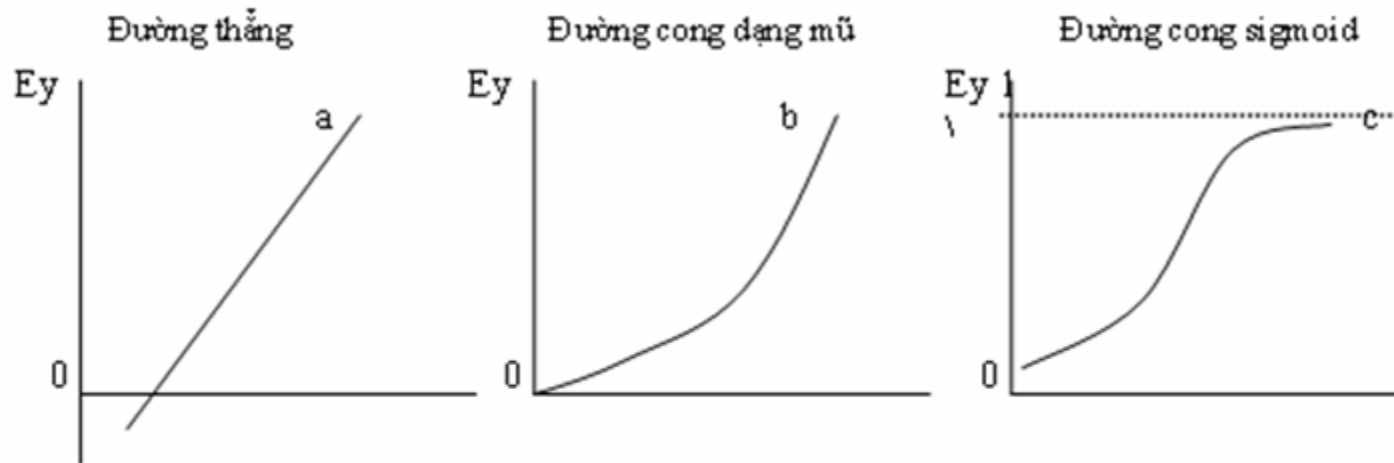


## Chương 7. Hồi quy trong sinh thái QXTV

- Các đường cong 7.9 - 7.11 được chỉ ra ở hình 7.8; trong đó hàm 7.11 là một đường cong sigmoid.
- $Ey = b_0 + b_1x$  (7.9)
- $Ey = \exp(b_0 + b_1x)$  (7.10)

$$Ey = p = \frac{\exp(b_0 + b_1x)}{1 + \exp(b_0 + b_1x)} \quad (7.11)$$

## Chương 7. Hồi quy trong sinh thái QXTV



**Hình 7.8.** Đường thẳng (a), đường cong hàm số mũ (b) và đường cong sigmoid (c) tương ứng với các **hàm 7.9 – 7.11.**

## Chương 7. Hồi quy trong sinh thái QXTV

- ✓ Đường cong sigmoid (hàm 7.11) thỏa điều kiện là “Tất cả giá trị nhận được đều nằm trong khoảng giữa 0 và 1”.
- ✓ Những đường cong sigmoid tăng đều hoặc giảm đều và có hai tham số  $b_0$  và  $b_1$ .
- ✓ Thành phần  $b_0 + b_1x$  biểu thị dự đoán đường thẳng.
- ✓ Đối với xác suất, chúng ta sử dụng kí hiệu  $p$  thay cho  $Ey$ .

## Chương 7. Hồi quy trong sinh thái QXTV

- ✓ Đối với hàm 7.11, khi ước lượng tham số từ số liệu, chúng ta không thể sử dụng hồi quy bình phương nhỏ nhất, mà phải sử dụng hồi quy logit.

## Chương 7. Hồi quy trong sinh thái QXTV

- ✓ Thuật ngữ **logit** xuất phát từ biến đổi logit, đó là biến đổi của  $p$ :

$$\log_e[p/(1-p)] = b_0 + b_1x \quad (7.12)$$

$$p = \frac{\exp(b_0 + b_1x)}{1 + \exp(b_0 + b_1x)} \quad (7.13)$$

## Chương 7. Hồi quy trong sinh thái QXTV

- **Hàm logit** còn được gọi là **hàm logistic**.
- Trong mô hình tuyến tính tổng quát, các tham số được ước lượng theo nguyên lý hợp lý tối đa.
- Ngày nay những tham số của hàm logistic được xác định dễ dàng bằng các phần mềm thống kê **SPSS và Statgraphics...**

## Chương 7. Hồi quy trong sinh thái QXTV

### 7.3.3. Đường cong logit Gauss

- ✓ Khi thay thế thành phần dự đoán tuyến tính trong **hàm 7.11** bằng hàm parabol, chúng ta thu được đường cong logit Gauss:

$$p = \frac{\exp(b_0 + b_1x + b_2x^2)}{1 + \exp(b_0 + b_1x + b_2x^2)}$$

$$\text{Hay } p = \frac{c \cdot \exp[-0.5(x - u)^2/t^2]}{1 + c \cdot \exp(-0.5(x - u)^2/t^2)} \quad (7.14)$$

## Chương 7. Hồi quy trong sinh thái QXTV

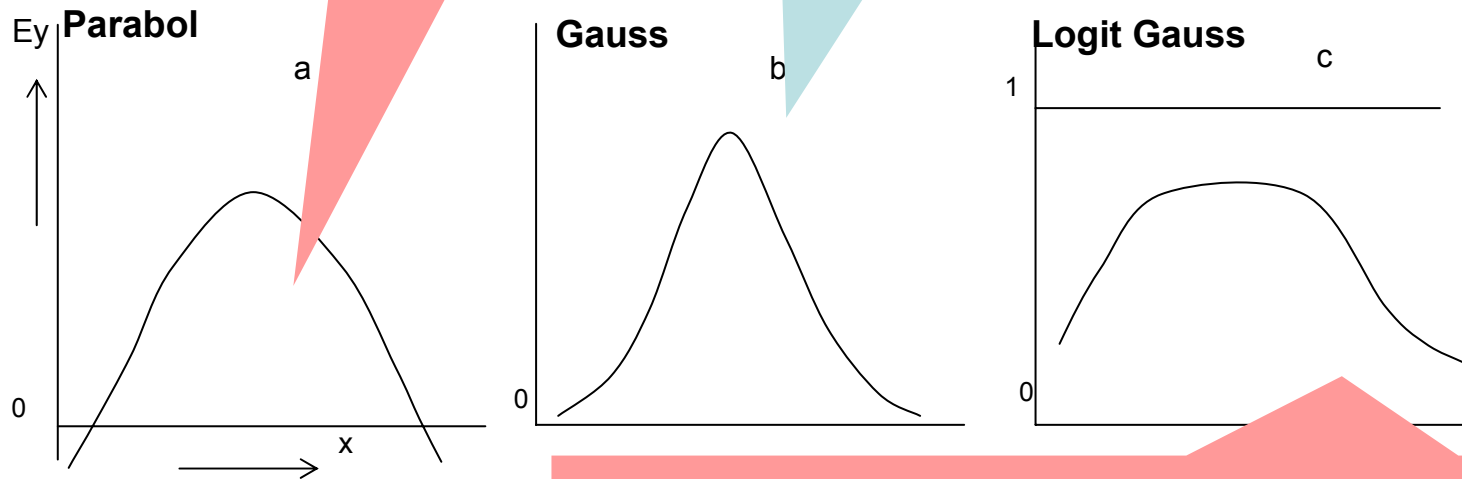
- Mọi liên hệ giữa parabol, đường cong Gauss và đường cong logit Gauss được chỉ ra ở hình 7.9.



# Chương 7. Hồi quy trong sinh thái QXTV

$$E_y = b_0 + b_1x + b_2x^2$$

$$\log_e(z) = b_0 + b_1x + b_2x^2$$



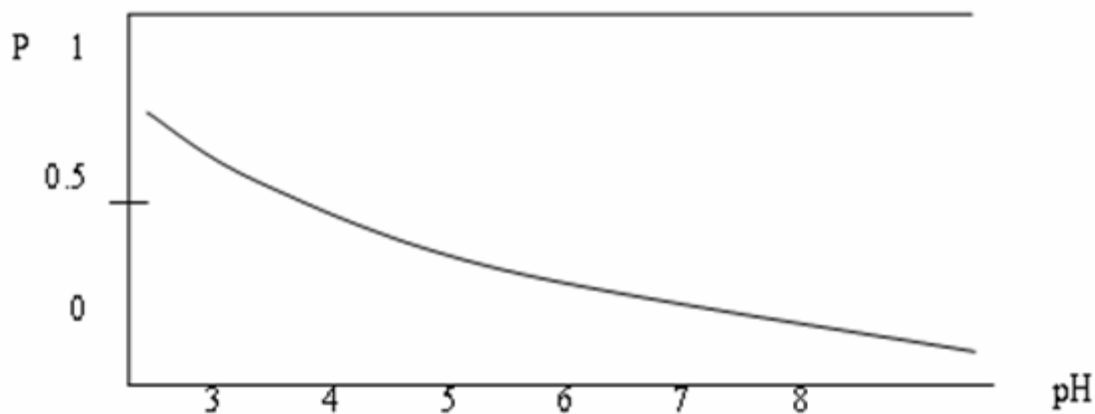
**Hình 7.9.** Đường parabol (a), đường cong Gauss (b) và đường cong Logit Gauss (c) tương ứng với các hàm 7.4, 7.5 và 7.14.

## Chương 7. Hồi quy trong sinh thái QXTV

- Đường cong logit Gauss có đỉnh bẹt hơn đường cong Gauss.
- Nhưng khi maximum của đường cong logit Gauss lấy giá trị nhỏ ( $< 0.5$ ), thì sự khác biệt giữa chúng không lớn.

## Chương 7. Hồi quy trong sinh thái QXTV

- Ví dụ 1: Mô tả quan hệ giữa độ phong phú của loài với pH đất ở hình 7.7.



Hình 7.7. Đường cong biểu diễn xác suất (p) xuất hiện loài trong quan hệ với pH môi trường

## Chương 7. Hồi quy trong sinh thái QXTV

- **Bây giờ thử mô tả bài toán này bằng hai dạng:**
  - ✓ **Đường cong sigmoid**
  - ✓ **Đường cong logit Gauss**

## Chương 7. Hồi quy trong sinh thái QXTV

- ✓ **Bảng 7.2.** Đường cong sigmoid: Tham số ước lượng và bảng sai tiêu chuẩn đối với dữ liệu dạng bất gặp – không bắt gặp loài.

<b>Tham số</b>		<b>Ước lượng</b>	<b>Sai số</b>	<b>t</b>
<b>Hằng số</b>	<b>b<sub>0</sub></b>	<b>2.03</b>	<b>1.98</b>	<b>1.03</b>
<b>pH</b>	<b>b<sub>1</sub></b>	<b>-0.484</b>	<b>0.357</b>	<b>-1.36</b>
		<b>d.f</b>	<b>Sai lệch</b>	<b>Sai lệch trung bình</b>
<b>Sai lệch</b>		<b>33</b>	<b>43.02</b>	<b>1.304</b>

## Chương 7. Hồi quy trong sinh thái QXTV

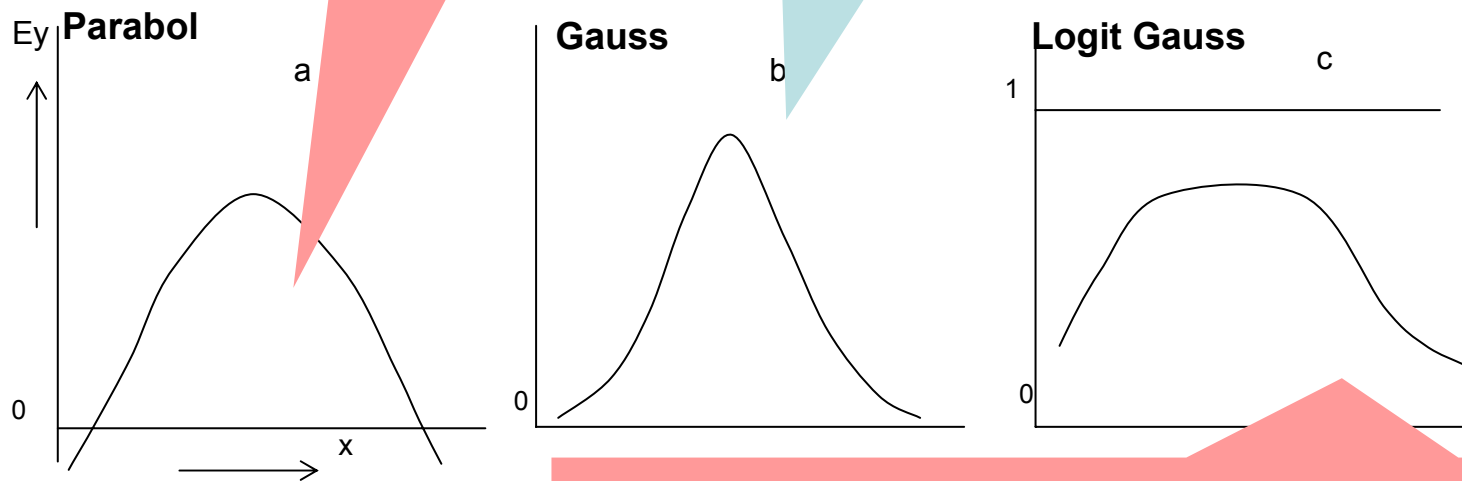
- Bảng 7.3. Đường cong logit Gauss:** Tham số ước lượng và bảng sai tiêu chuẩn đối với dữ liệu dạng bất gặp – không bất gặp loài.

<b>Tham số</b>		<b>Ước lượng</b>	<b>Sai số</b>	<b>t</b>
<b>Hằng số</b>	<b>b<sub>0</sub></b>	<b>-128.8</b>	<b>51.1</b>	<b>-2.52</b>
<b>pH</b>	<b>b<sub>1</sub></b>	<b>49.4</b>	<b>19.8</b>	<b>2.5</b>
<b>pH<sup>2</sup></b>	<b>b<sub>2</sub></b>	<b>4.68</b>	<b>1.9</b>	<b>-2.47</b>
		<b>d.f</b>	<b>Sai lệch</b>	<b>Sai lệch trung bình</b>
<b>Sai lệch</b>		<b>32</b>	<b>23.17</b>	<b>0.724</b>

# Chương 7. Hồi quy trong sinh thái QXTV

$$E_y = b_0 + b_1x + b_2x^2$$

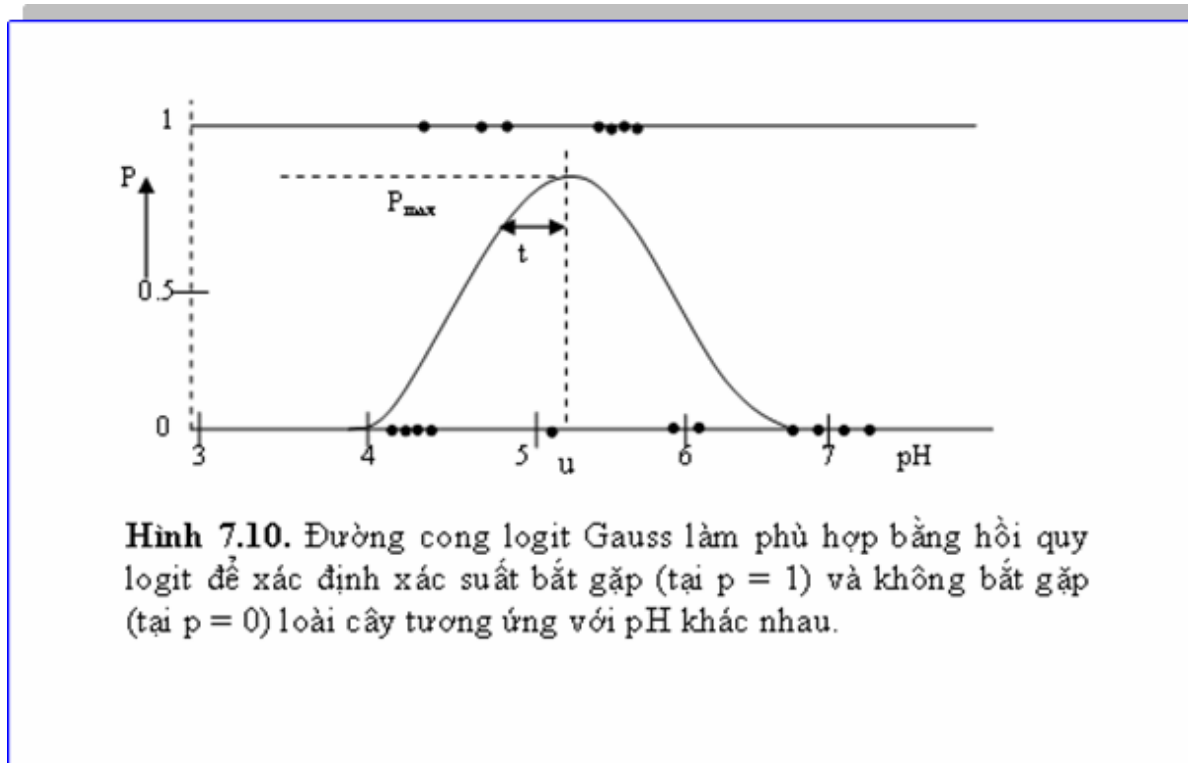
$$\log_e(z) = b_0 + b_1x + b_2x^2$$



**Hình 7.9.** Đường parabol (a), đường cong Gauss (b) và đường cong Logit Gauss (c) tương ứng với các hàm 7.4, 7.5 và 7.14.

## Chương 7. Hồi quy trong sinh thái QXTV

- Biểu diễn đường cong logit Gauss từ ví dụ 1 bằng đồ thị





## Chương 7. Hồi quy trong sinh thái QXTV

- Từ ví dụ 1, chúng ta thu được các tham số sinh thái sau đây:
  - ✓  $u = b_1/(2b_2) = 5,28$ . Khoảng tin cậy:  $[5,0 - 5,8]$
  - ✓  $t = 1/\sqrt{-2b_2} = 0,327$ .
  - ✓ Biên độ sinh thái với pH:  $u \pm t = [4,9 - 5,6]$
  - ✓  $P_{\max} = 0,858$ .

## Chương 7. Hồi quy trong sinh thái QXTV

- ✓ So với mô hình sigmoid (bảng 7.2), mô hình logit Gauss (bảng 7.3) thích hợp hơn

$$(43,02 - 23,17 = 19,85) \gg \chi^2_{0,05}(1) = 3,84$$

## Chương 7. Hồi quy trong sinh thái QXTV

### 7.4. HỒI QUY ĐỐI VỚI ĐỘ PHONG PHÚ NHẬN NHIỀU GIÁ TRỊ ZERO

- Độ phong phú của loài nhận nhiều giá trị zero luôn luôn có phân bố lệch.
- Vì thế, trước khi phân tích chúng bằng hồi quy bình phương nhỏ nhất, chúng ta cần phải biến đổi số liệu ban đầu.

## Chương 7. Hồi quy trong sinh thái QXTV

- Khi độ phong phú của loài nhận nhiều giá trị zero, chúng ta phải biến đổi số liệu như thế nào?

## Chương 7. Hồi quy trong sinh thái QXTV

### (1) Sử dụng phép tính logarit độ phong phú?

- Cách biến đổi này là không phù hợp, bởi vì phép tính logarit độ phong phú với những số zero thì không xác định.

## Chương 7. Hồi quy trong sinh thái QXTV

- (2) **Biến đổi bằng cách thêm một giá trị nhỏ vào độ phong phú quan sát?**
- Cách làm này cũng không phù hợp.
  - **Bởi vì**, nếu chúng ta thêm vào độ phong phú quan sát những giá trị khác nhau, **thì kết quả phân tích dữ liệu có nhiều số zero sẽ nhận những kết quả khác nhau.**
  - **Ngoài ra**, mô hình có thể nhận những giá trị độ phong phú âm.

## Chương 7. Hồi quy trong sinh thái QXTV

(3) Những cách biến đổi khác cũng không thực hiện được.

## Chương 7. Hồi quy trong sinh thái QXTV

- Một phân bố xác suất cho phép giải quyết đối với những giá trị zero là phân bố Poisson.
- Những quan sát xuất hiện từ phân bố Poisson có thể nhận những giá trị nguyên  $0, 1, 2, 3 \dots k$  và phương sai thì bằng số trung bình.



## Chương 7. Hồi quy trong sinh thái QXTV

- Số cây bắt gặp trong ô tiêu chuẩn đặt trong một quần xã thực vật chỉ lấy những giá trị nguyên.
- Khi giả định phân bố số cây trong ô tiêu chuẩn tuân theo phân bố Poisson, thì đường cong phản hồi không nhận giá trị âm, nhưng nó có thể lệch về phía giá trị 1.

## Chương 7. Hồi quy trong sinh thái QXTV

- **Vì thế**, khi số liệu độ phong phú nhận những giá trị zero, chúng ta có thể sử dụng cách biến đổi theo luật số mũ.

## Chương 7. Hồi quy trong sinh thái QXTV

- Hàm số mũ có dạng:

$$E_y = \exp(b_0 + b_1 x)$$

## Chương 7. Hồi quy trong sinh thái QXTV

- Đường cong theo luật số mũ có thể làm phù hợp số liệu bằng hồi quy log-tuyến tính.

## Chương 7. Hồi quy trong sinh thái QXTV

- Hồi quy log-tuyến tính có dạng:

$$\log_e(Ey) = b_0 + b_1x \quad (7.15)$$

## Chương 7. Hồi quy trong sinh thái QXTV

- Từ phương trình 7.15, nếu thay thế

$$\begin{aligned} & \triangleright \log_e(Ey) = b_0 + b_1x \\ & = \log_e(Ey) = b_0 + b_1x + b_2x^2 \end{aligned}$$

thì chúng ta lại nhận được đường cong logit Gauss với  $b_2 < 0$ .

## Chương 7. Hồi quy trong sinh thái QXTV

- Như vậy, khi sử dụng hồi quy log – tuyến tính, chúng ta đã tránh được vấn đề phải lấy logarit những số zero.
- Đồng thời từ những giá trị  $b_0$ ,  $b_1$  và  $b_2$ , chúng ta lại nhận được những thông tin về optimum ( $u$ ), tính chống chịu sinh thái ( $t$ ) và độ phong phú lớn nhất ( $p_{max}$ ).

## Chương 7. Hồi quy trong sinh thái QXTV

### 7.5. HỒI QUY ĐA BIẾN SỐ

- Phần trên chúng ta đã nghiên cứu phản hồi của loài với một biến giải thích (**môi trường**).
- **Nhưng trong nhiều trường hợp**, chúng ta cần phải nghiên cứu phản hồi của loài với nhiều hơn một biến giải thích.
- **Để đạt được mục đích này**, chúng ta cần đến phân tích hồi quy đa biến số.



## Chương 7. Hồi quy trong sinh thái QXTV

- Trong phân tích hồi quy đa biến số, biến phản hồi là một hàm số của hai hay nhiều biến giải thích (phân tích mặt phản hồi).

## Chương 7. Hồi quy trong sinh thái QXTV

- Phân tích riêng rẽ phản hồi đối với mỗi biến môi trường không thể thay thế được phân tích phản hồi đa biến số.
- Vì sao?

## Chương 7. Hồi quy trong sinh thái QXTV

- Nguyên nhân:
- ✓ Biến môi trường này có thể có quan hệ chặt chẽ với một biến môi trường khác.
- ✓ Nhiều biến có ảnh hưởng tương tác lẫn nhau, nghĩa là ảnh hưởng của biến này lại phụ thuộc vào biến khác.

## Chương 7. Hồi quy trong sinh thái QXTV

- Trong phân tích hồi quy đa biến số, số liệu phải tuân theo những giả định sau đây:
  - a. Từng biến giải thích phải có quan hệ với biến phản hồi.
  - b. Các biến giải thích phải độc lập với nhau hay không có quan hệ chặt chẽ với nhau.
  - c. Phương sai của các biến giải thích là bằng nhau.

## Chương 7. Hồi quy trong sinh thái QXTV

- Trong phần này chúng ta xem xét hai vấn đề:
  - ✓ Phân tích hồi quy đa biến số bằng nguyên lý bình phương nhỏ nhất.
  - ✓ Hồi quy logit đa biến số đối với những biến định lượng và cả biến định danh.

## Chương 7. Hồi quy trong sinh thái QXTV

### 7.5.1. Hồi quy bình phương nhỏ nhất đa biến

#### (1) Khi có hai biến giải thích

$$E_y = b_0 + b_1x_1 + b_2x_2 \quad (7.16)$$

✓ trong đó:

+  $x_1$  và  $x_2$  = hai biến giải thích

+  $b_0, b_1, b_2$  = hệ số hồi quy phải tìm

## Chương 7. Hồi quy trong sinh thái QXTV

(2) Khi mở rộng hồi quy bậc 2, chúng ta nhận được mặt bậc hai như sau:

$$E\mathbf{y} = \mathbf{b}_0 + \mathbf{b}_1\mathbf{x}_1 + \mathbf{b}_2\mathbf{x}_1^2 + \mathbf{b}_3\mathbf{x}_2 + \mathbf{b}_4\mathbf{x}_2^2 \quad (7.17)$$

✓ trong đó:

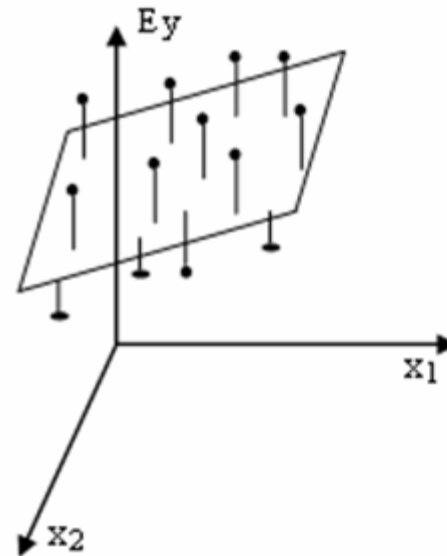
+  $\mathbf{x}_1$  và  $\mathbf{x}_2$  = hai biến giải thích

+  $\mathbf{b}_0, \mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3, \mathbf{b}_4$  = hệ số hồi quy phải tìm

## Chương 7. Hồi quy trong sinh thái QXTV

- Hồi quy đối với hai biến giải thích là một mặt phẳng (hình 7.11).

Hình 7.11. Hình chiếu 3 chiều của một mặt phẳng được làm phù hợp bằng hồi quy bình phương nhỏ nhất với 2 biến  $x_1$  và  $x_2$ .





## Chương 7. Hồi quy trong sinh thái QXTV

- Từ hàm 7.17, chúng ta có thể tính được những thông tin về tối ưu ( $u$ ) và sức chống chịu sinh thái ( $t$ ) của loài đối với biến  $x_1$  bằng việc thêm các tham số  $b_1$  và  $b_2$  vào phương trình 7.8(a,b).
  - tối ưu:  $u = -b_1/(2b_2)$  (7.8a)
  - tính chống chịu:  $t = 1/\sqrt{-2b_2}$  (7.8b)
  - lớn nhất:  $c = \exp(b_0 + b_1u + b_2u^2)$  (7.8c)

## Chương 7. Hồi quy trong sinh thái QXTV

- Tương tự, **đối với biến  $x_2$** :
  - ✓  $u = -b_3/(2b_4)$
  - ✓  $t = 1/\sqrt{-2b_4}$
  - ✓  $c = \exp(b_0 + b_3u + b_4u^2)$

## Chương 7. Hồi quy trong sinh thái QXTV

- Thủ tục phân tích hồi quy đa biến:
- ✓ Kiểm định những giả thuyết của phân tích hồi quy đa biến.
- ✓ Ước lượng các tham số  $b_0$ ,  $b_1$  và  $b_2$  bằng nguyên lý bình phương nhỏ nhất.

## Chương 7. Hồi quy trong sinh thái QXTV

- ✓ Kiểm định sự tồn tại của mô hình hay sự tồn tại của các hệ số hồi quy ( $b_0$ ,  $b_1$  và  $b_2$ ) bằng thống kê t.
- ✓ Tính sai số chuẩn ước lượng.
- ✓ Tính hệ số xác định.
- Ngày nay tất cả những thủ tục phân tích hồi quy đa biến được giải quyết dễ dàng bằng các phần mềm thống kê Minitab, SPSS và Statgraphics...

## Chương 7. Hồi quy trong sinh thái QXTV

- Khi  $Ey$  trong mô hình 7.17 ( $Ey = b_0 + b_1x_1 + b_2x_1^2 + b_3x_2 + b_4x_2^2$ ) là logarit độ phong phú ( $\log_e[p/(1-p)]$ ), chúng ta làm phù hợp mặt phản hồi Gauss hai biến bằng hồi quy đa biến đối với độ phong phú quan sát; trong đó hệ số  $b_2$  và  $b_4$  lấy giá trị âm.

## Chương 7. Hồi quy trong sinh thái QXTV

### 7.5.3. Hồi quy logit đa biến: mặt logit và mặt logit Gauss

#### (1) Hồi quy logit đa biến

- Hồi quy logit thu được từ hồi quy bình phương nhỏ nhất bằng việc thay thế  $Ey = \log_e[p/(1-p)]$ .

## Chương 7. Hồi quy trong sinh thái QXTV

### (2) Mặt phẳng logit:

$$\log_e[p/(1-p)] = b_0 + b_1x_1 + b_2x_2 \quad (7.19)$$

Hay

$$p = \frac{\exp(b_0 + b_1x_1 + b_2x_2)}{1 + \exp(b_0 + b_1x_1 + b_2x_2)}$$

## Chương 7. Hồi quy trong sinh thái QXTV

### (3) Mặt logit Gauss 2 biến số

$$\log_e[p/(1-p)] = b_0 + b_1x_1 + b_2x_1^2 + b_3x_2 + b_4x_2^2$$

$$p = \frac{\exp(b_0 + b_1x_1 + b_2x_1^2 + b_3x_2 + b_4x_2^2)}{[1 + \exp(b_0 + b_1x_1 + b_2x_1^2 + b_3x_2 + b_4x_2^2)]}$$

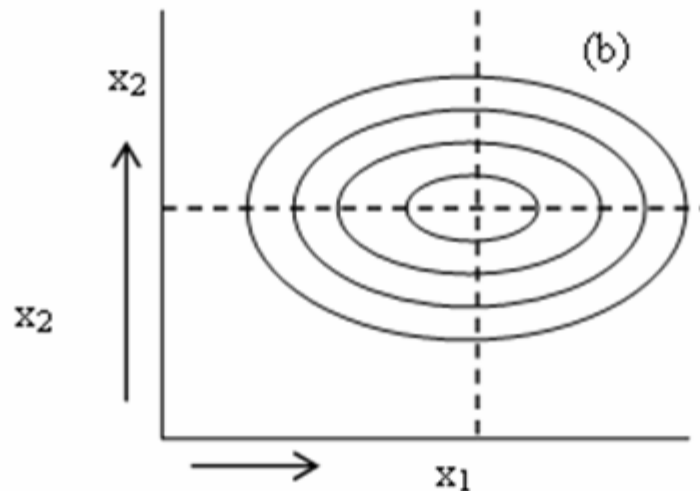
trong đó hai tham số  **$b_2$**  và  **$b_4$**  nhận giá trị âm.



## Chương 7. Hồi quy trong sinh thái QXTV

- ✓ Mặt logit Gauss 2 biến số có dạng êlip (đường viền quanh cho biết xác suất bằng nhau) với các trục chính song song với trục  $x_1$  và  $x_2$  (hình 7.14b).

## Chương 7. Hồi quy trong sinh thái QXTV



**Hình 7.14b.** Xác suất xuất hiện loài (P) được vẽ trong mặt phẳng của hai biến giải thích  $x_1$  và  $x_2$ . Một trục ellip song song với  $x_1$ , trục kia song song với  $x_2$ .

## Chương 7. Hồi quy trong sinh thái QXTV

- ✓ Các tham số của mô hình bậc 2 được ước lượng theo nguyên lý GLM (*General linear model*).
- ✓ Những phân tích khác cũng được thực hiện như hàm 7.17, ngoại trừ kiểm định F phải thay bằng kiểm định sai lệch.

## Chương 7. Hồi quy trong sinh thái QXTV

### 7.5.4. Sự tương tác giữa các biến giải thích

- ✓ Hai biến giải thích có ảnh hưởng tương tác với nhau, nếu ảnh hưởng của biến này phụ thuộc vào giá trị của biến khác.
- ✓ Chúng ta có thể kiểm định ảnh hưởng tương tác bằng cách mở rộng phương trình hồi quy với tích số ( $x_1 * x_2$ ).

## Chương 7. Hồi quy trong sinh thái QXTV

- ✓ Bằng cách mở rộng hàm 2 biến số (hàm 7.16), chúng ta có:

$$E y = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_1 * x_2$$

$$\text{Hay } E y = (b_0 + b_2 x_2) + (b_1 + b_3 x_2) x_1 \quad (7.20)$$

## Chương 7. Hồi quy trong sinh thái QXTV

- ✓ Biểu thức (7.20) cho biết mối liên hệ giữa  $Ey$  với  $x_1$  trong mô hình này vẫn là một dạng tuyến tính, nhưng điểm chặn ( $b_0 + b_2x_2$ ) và độ dốc ( $b_1 + b_3x_2$ ) và cả ảnh hưởng của  $x_1$  đều phụ thuộc  $x_2$ .
- ✓ Ngược lại, ảnh hưởng của  $x_2$  chỉ phụ thuộc vào giá trị  $x_1$ .

## Chương 7. Hồi quy trong sinh thái QXTV

- ✓ Tham số  $b_1$ ,  $b_2$  và  $b_3$  trong hàm (7.20) có thể ước lượng bằng thủ tục hồi quy đa biến; trong đó đặt  $x_3 = x_1 * x_2$ .
- ✓ Tương tác có thể kiểm định bằng thống kê t.

## Chương 7. Hồi quy trong sinh thái QXTV

- Bằng việc mở rộng mô hình Gauss của **hàm 7.17** với một tích số, chúng ta thu được trong trường hợp logit:

$$\log_e[p/(1-p)] = b_0 + b_1x_1 + b_2x_1^2 + b_3x_2 + b_4x_2^2 + b_5x_1x_2 \quad (7.21)$$

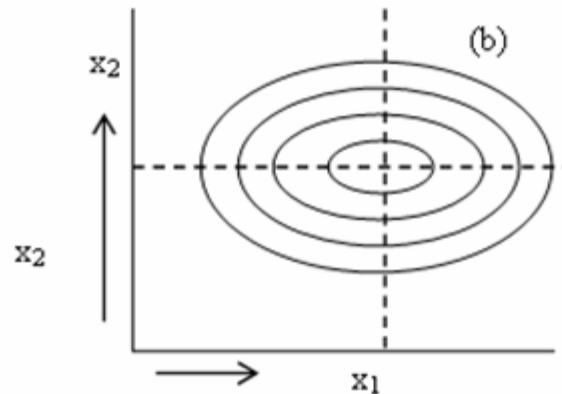
trong đó:

- ✓  $x_1$  và  $x_2$  = hai biến giải thích
- ✓  $b_0, b_1, b_2, b_3, b_4$  = các hệ số hồi quy



## Chương 7. Hồi quy trong sinh thái QXTV

- ✓ Nếu  $b_2 + b_4 < 0$  và  $4b_2b_4 - b_5^2 > 0$ , thì phương trình 7.21 là một mặt phẳng với đường viền ellipsoid (hình 7.14b).



Hình 7.14b. Xác suất xuất hiện loài (P) được vẽ trong mặt phẳng của hai biến giải thích  $x_1$  và  $x_2$ . Một trục ellip song song với  $x_1$ , trục kia song song với  $x_2$ .

## Chương 7. Hồi quy trong sinh thái QXTV

- ✓ Nếu một trong những điều kiện này không thoả mãn, thì phương trình 7.21 mô tả một mặt với một tối thiểu hoặc một điểm lõm (Carroll, 1972).

## Chương 7. Hồi quy trong sinh thái QXTV

- ✓ Khi bề mặt là phẳng, tối ưu  $(u_1, u_2)$  có thể được tính từ hệ số của hàm 7.21 như sau:

$$u_1 = (b_5 b_3 - 2b_1 b_4)/d \quad (7.22a)$$

$$u_2 = (b_5 b_1 - 2b_3 b_2)/d \quad (7.22b)$$

$$d = 4b_2 b_4 - b_5^2 \quad (7.22c)$$

## Chương 7. Hồi quy trong sinh thái QXTV

- ✓ Tối ưu đối với  $X_1$  tương ứng với  $X_2$  nhất định được tính theo công thức:

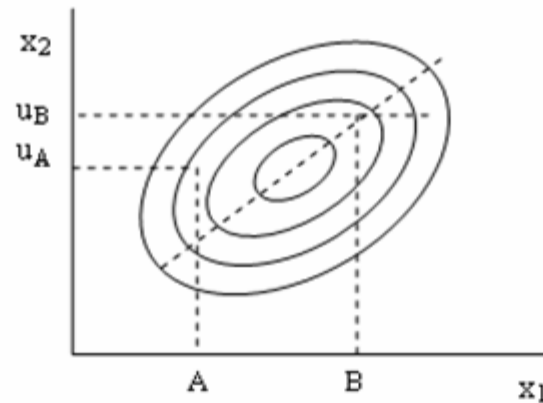
$$u_1 = -(b_1 + b_5x_2)/(2b_2)$$

## Chương 7. Hồi quy trong sinh thái QXTV

- ✓ Nếu  $b_5 \neq 0$ , thì tối ưu đối với  $X_1$  phụ thuộc vào  $X_2$ , và hai biến tương tác với nhau.
- ✓ Điều kiện này được kiểm định bằng thống kê t.

## Chương 7. Hồi quy trong sinh thái QXTV

- ✓ Đồ thị mô tả sự tương tác giữa  $x_1$  với  $x_2$  như sau (hình 7.15):



**Hình 7.15.** Tương tác trong mô hình logit Gauss  
Đường viền ellip biểu thị xác suất xuất hiện ( $p$ ) đối với các biến giải thích  $x_1$  và  $x_2$ . Các trục chính của ellip không song song với hai trục  $x_1$  và  $x_2$ .  $u_A$  và  $u_B$  là optimum đối với  $x_2$  tương ứng với mức  $x_1 = A$  và  $x_1 = B$ .

## Chương 7. Hồi quy trong sinh thái QXTV

### 7.5.5. Các biến giải thích định danh

- ✓ Hồi quy đa biến cũng được sử dụng để nghiên cứu ảnh hưởng của các biến giải thích (môi trường) định danh (định tính) hoặc cả hai biến môi trường định danh và định lượng.

## Chương 7. **Hội quy trong sinh thái QXTV**

- Giả sử biến định danh “**kiểu đất**” có ba cấp:
  - ✓ Đất sét
  - ✓ Đất than bùn
  - ✓ Đất cát



## Chương 7. Hồi quy trong sinh thái QXTV

- Giả sử đất sét là biến kiểm tra.
- Bởi vì loại đất bao gồm ba mức (sét, than bùn và cát); do đó chúng ta cần có  $3 - 1 = 2$  biến giả để chỉ các loại đất đem so sánh (đất than bùn và cát).

## Chương 7. Hồi quy trong sinh thái QXTV

- Mỗi biến giả được định nghĩa như sau:
  - ✓  $x_2 = 0$  nếu loại đất là sét và cát
  - ✓  $x_2 = 1$  nếu loại đất là than bùn
  - ✓  $x_3 = 1$  nếu loại đất là cát
  - ✓  $x_3 = 0$  nếu loại đất là sét và than bùn

## Chương 7. Hồi quy trong sinh thái QXTV

- Theo định nghĩa như trên, chúng ta có các giá trị  $x_2$  và  $x_3$  như sau:

•	Loại đất	$X_2$	$X_3$
✓	sét	0	0
✓	than bùn	1	0
✓	cát	0	1

## Chương 7. Hồi quy trong sinh thái QXTV

- Cách mã hóa này cho phép so sánh từng cặp biến phản hồi trên đất than bùn và đất cát với biến phản hồi trên đất sét.

## Chương 7. Hồi quy trong sinh thái QXTV

- ✓ Hàm hồi quy biểu thị quan hệ giữa biến phản hồi ( $Ey$ ) với 2 biến giả được viết dưới dạng sau đây:

$$E(y) = b_1 + b_2x_2 + b_3x_3 \quad (*)$$

## Chương 7. Hồi quy trong sinh thái QXTV

- Để giải thích các tham số  $b_1$ ,  $b_2$  và  $b_3$ , chúng ta xem xét ba biến động của 3 hàm hồi quy (\*) như sau:

## Chương 7. Hồi quy trong sinh thái QXTV

### ✓ Đối với đất sét:

$$E(y/\text{đất sét}) = b_1 + b_2(0) + b_3(0) = b_1$$

- ✓ Giá trị  $b_1$  là kỳ vọng của biến phản hồi trên loại đất sét.

## Chương 7. Hồi quy trong sinh thái QXTV

### ✓ Đối với đất than bùn:

$$E(y/\text{đất than bùn}) = b_1 + b_2(1) + b_3(0) = b_1 + b_2$$

- ✓ Giá trị  $b_1 + b_2$  là kỳ vọng của biến phản hồi trên loại đất than bùn.
- ✓ Tham số  $b_2$  là sự khác biệt trung bình của biến phản hồi trên loại đất sét và đất than bùn.



## Chương 7. Hồi quy trong sinh thái QXTV

✓ **Đối với đất cát:**

$$E(y/\text{đất cát}) = b_1 + b_2(0) + b_3(1) = b_1 + b_3$$

- ✓ Giá trị  $b_1 + b_3$  là kỳ vọng của biến phản hồi trên loại đất cát.
- ✓ Tham số  $b_3$  là sự khác biệt trung bình của biến phản hồi trên loại đất sét và đất cát.

## Chương 7. Hồi quy trong sinh thái QXTV

- Như vậy, so với biến phản hồi trung bình trên đất sét ( $b_1$ ), biến phản hồi trung bình trên đất than bùn và đất cát khác nhau tương ứng là  $b_2$  và  $b_3$ .

## Chương 7. Hồi quy trong sinh thái QXTV

- Lưu ý rằng, các hệ số  $b_1$ ,  $b_2$  và  $b_3$  cũng được ước lượng theo nguyên lý bình phương nhỏ nhất.

## Chương 7. Hồi quy trong sinh thái QXTV

- Ví dụ:
- ✓ Nghiên cứu ảnh hưởng của tính dẫn điện ( $EC, mSm^{-1}$ ) của ba loại đất (sét, than bùn và cát) đến độ phong phú của loài *Equisetum fluviatile* bằng hồi quy logit.

## Chương 7. Hồi quy trong sinh thái QXTV

- Ở đây độ phong phú của loài được đo bằng 2 dấu hiệu:
  - ✓ Bắt gặp loài = 1
  - ✓ Không bắt gặp loài = 0.

## Chương 7. Hồi quy trong sinh thái QXTV

- Mô hình hồi quy logit có dạng:

$$\log_e(p/(1-p)) =$$

$$= b_0 + b_1x_1 + b_2x_1^2 + b_3x_2 + b_4x_3 \quad (7.23)$$

## Chương 7. Hồi quy trong sinh thái QXTV

$$p = \frac{\exp(b_0 + b_1x_1 + b_2x_1^2 + b_3x_2 + b_4x_3)}{[1 + \exp(b_0 + b_1x_1 + b_2x_1^2 + b_3x_2 + b_4x_3)]}$$

- **trong đó**
- ✓  **$x_1$  = logarit của độ dẫn điện**
- ✓  **$x_2$  và  $x_3$  = hai biến giả**
- ✓  **$b_3$  và  $b_4$  = ảnh hưởng của loại đất**

## Chương 7. Hồi quy trong sinh thái QXTV

- ✓ Hệ số  $b_3$  là sự khác biệt giữa logit lớn nhất của đường cong đối với đất than bùn và đất sét.
- ✓ Hệ số  $b_4$  là sự khác biệt giữa logit lớn nhất của đường cong đối với đất cát và đất sét.

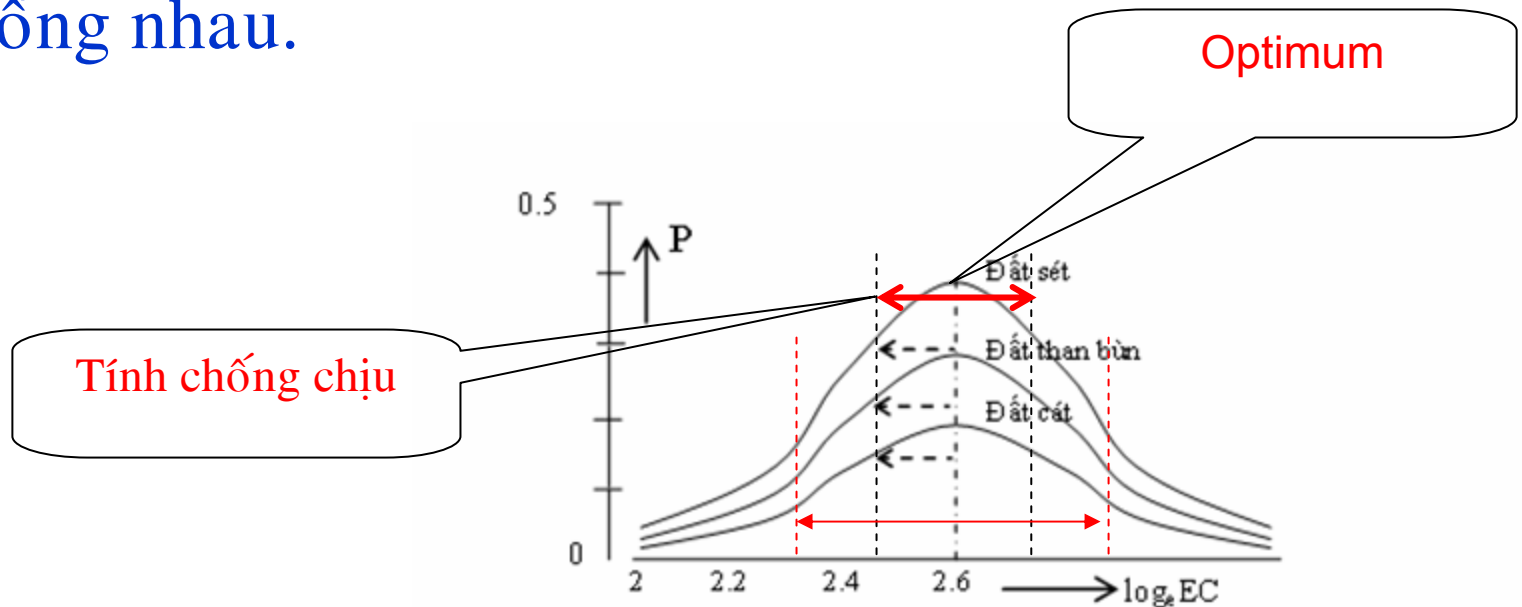


## Chương 7. Hồi quy trong sinh thái QXTV

- Từ hàm 7.23, chúng ta có thể tính được tối ưu sinh thái và tính chống chịu sinh thái của loài.
- Ví dụ:
- ✓ Bài tập 3.6. [Phan tích quan xa\Bai tap PT QXSVat\Baitap 3.5 va 3.6\Bai tap 3.6. Hoi quy logit da bien.xls](#)

## Chương 7. Hồi quy trong sinh thái QXTV

- ✓ Hình 7.16 mô tả ba đường cong với 3 cực đại khác nhau, nhưng tối ưu sinh thái và tính chống chịu là giống nhau.



**Hình 7.16.** Đường cong phản hồi độ phong phú của loài *Equisetum fluviatile* được làm phù hợp bằng hồi quy logit đa biến với logarit độ dẫn điện của 3 loại đất ( $\log_e EC$ ).

## Chương 7. Hồi quy trong sinh thái QXTV

- Sự khác biệt giữa các cực đại của ba đường cong này được kiểm định bằng cách so sánh sai lệch của mô hình bao gồm biến  $x_1$  và  $x_1^2$  với sai lệch của hàm 7.23.
- Nếu kiểu đất không có ảnh hưởng đến sự phân bố của loài, thì sự khác biệt là  $\chi^2$  với 2 độ tự do.

## Chương 7. Hồi quy trong sinh thái QXTV

- Để tính tối ưu sinh thái và tính chống chịu sinh thái ở **hàm 7.23**, chúng ta sử dụng **hàm 7.8**.

# Chương 7. Hồi quy trong sinh thái QXTV

## 7.6. CHỌN MÔ HÌNH VÀ PHÉP CHUẨN ĐOÁN HỒI QUY

- Có nhiều nguyên nhân làm sai lệch kết quả phân tích hồi quy.
- **Chẳng hạn:**
  - ✓ Chọn lựa không chính xác kiểu đường cong phản hồi và phân bố sai số.
  - ✓ Sự xuất hiện những số ngoại lai làm ảnh hưởng đến hồi quy.

## Chương 7. Hồi quy trong sinh thái QXTV

- Do đó, phát hiện chính xác những sai lầm ấy là mục đích của chuẩn đoán hồi quy (Belsley *et al.*, 1980; Cook và Weisberg, 1982).

## Chương 7. Hồi quy trong sinh thái QXTV

- Cơ sở chọn lựa kiểu đường cong phản hồi:
- ✓ Quan sát hình dạng của đường cong bằng cách xây dựng đồ thị phân tán.
- ✓ Dựa trên cơ sở lý thuyết sinh thái và sinh lý cây – con.
- ✓ Những kinh nghiệm của các nhà nghiên cứu...

## Chương 7. Hồi quy trong sinh thái QXTV

- ✓ Nếu chưa hiểu rõ hình thái đường cong phản hồi, chúng ta có thể phân chia nhỏ biến môi trường định lượng thành nhiều cấp và tìm tần số xuất hiện cho mỗi cấp.



## Chương 7. Hồi quy trong sinh thái QXTV

- Những điều cần lưu ý khi chọn mô hình đường cong phản hồi:
  - (1) Để tránh việc chọn sai mô hình phản hồi, nhà nghiên cứu cần phải lưu ý đến tính đối xứng của đường cong phản hồi.
- ✓ Khi đường cong phản hồi là bất đối xứng, thì việc biến đổi biến giải thích bằng cách lấy logarit là cần thiết.

## Chương 7. Hồi quy trong sinh thái QXTV

- ✓ Muốn biết cách biến đổi số liệu có phù hợp hay không, chúng ta thực hiện so sánh sai lệch trước và sau khi chuyển đổi số liệu.

## Chương 7. Hồi quy trong sinh thái QXTV

- ✓ Ghi nhận lại sai lệch từ đường cong phản hồi giả định có thể giúp xác định quan hệ của loài với môi trường và cho phép bắt đầu một quá trình tìm kiếm một mô hình khác.

## Chương 7. Hồi quy trong sinh thái QXTV

- (2) Khi nghiên cứu độ phong phú của loài ở dạng có mặt hay vắng mặt, chúng ta cần chú ý phát hiện đúng những điểm số liệu ngoại lai.
- Bởi vì, những số ngoại lai có thể làm sai lệch đường cong phản hồi.

## Chương 7. Hồi quy trong sinh thái QXTV

- (3) Chúng ta cũng cần chú ý đến việc phân cấp các biến môi trường.
- Việc phân cấp không hợp lý các biến môi trường sẽ làm thay đổi hình thái đường cong phản hồi.

## Chương 7. Hồi quy trong sinh thái QXTV

- **Nói chung**, để phát hiện những sai lầm trong phân tích hồi quy, chúng ta có thể thực hiện những thủ tục sau đây:
  - ✓ (1) Xây dựng đồ thị sai lệch đối với biến độc lập  $x$ .
  - ✓ (2) Xây dựng đồ thị sai lệch đối với biến dự đoán  $y$ .

## Chương 7. Hồi quy trong sinh thái QXTV

- ✓ (3) Xây dựng đồ thị các sai lệch chuẩn.
- ✓ (4) Xây dựng đồ thị xác suất chuẩn.
- ✓ (5) Kiểm tra số ngoại lai theo số đo khoảng cách của Cook và Mahalanobis.

# Chương 7. Hồi quy trong sinh thái QXTV

## 7.7. PHƯƠNG PHÁP TRUNG BÌNH TRỌNG SỐ

- Phần này chúng ta xem xét cách xác định giá trị chỉ thị của loài bằng phương pháp trung bình trọng số.



## Chương 7. Hồi quy trong sinh thái QXTV

- Trong thuật ngữ đường cong phản hồi, chúng ta cần định nghĩa hai giá trị chỉ thị của loài.
- ✓ Một là tối ưu sinh thái.
- ✓ Hai là trung tâm đường cong phản hồi của loài.
- Nếu đường cong phản hồi có dạng đối xứng, thì hai giá trị này là giống nhau.

## Chương 7. Hồi quy trong sinh thái QXTV

- Trong mục 7.2.3 và 7.5.2, chúng ta đã thấy tối ưu sinh thái có thể được ước lượng bằng hai cách:
  - ✓ (1) Làm phù hợp số liệu với một đường cong.
  - ✓ (2) Làm phù hợp số liệu với một mặt phản hồi.

## Chương 7. Hồi quy trong sinh thái QXTV

- Lưu ý rằng, tối ưu và tính chống chịu sinh thái của loài chỉ được xác định bằng phương pháp hồi quy khi biến phản hồi tuân theo một đường cong nhất định.

## Chương 7. Hồi quy trong sinh thái QXTV

- Nhưng trước đây các nhà sinh thái định lượng đã xác định những giá trị chỉ thị của loài bằng phương pháp trung bình trọng số (Ellenberg, 1948; 1979).

## Chương 7. Hồi quy trong sinh thái QXTV

- Khi phân bố của loài theo yếu tố môi trường có dạng đường cong một đỉnh, thì loài thường xuất hiện nhiều nhất ở khu vực gần với tối ưu của đường cong.

## Chương 7. Hồi quy trong sinh thái QXTV

- **Vì thế**, khi ước lượng giá trị chỉ thị của loài, chúng ta có thể lấy trung bình những giá trị của biến môi trường ở những nơi mà loài xuất hiện.

## Chương 7. Hồi quy trong sinh thái QXTV

- Đối với số liệu độ phong phú của loài, trung bình trọng số được tính theo công thức:

$$u^+ = \frac{(y_1x_1 + y_2x_2 + \dots + y_nx_n)}{(y_1 + y_2 + \dots + y_n)} \quad (7.24)$$

## Chương 7. Hồi quy trong sinh thái QXTV

- **Trong đó:**
  - ✓  $u^+$  = trung bình trọng số
  - ✓  $y_1, y_2 \dots y_n$  = độ phong phú của loài
  - ✓  $x_1, x_2 \dots x_n$  = giá trị của biến môi trường ở mẫu 1, 2...n.



## Chương 7. Hồi quy trong sinh thái QXTV

- Từ công thức 7.24 cho thấy:
- ✓ Trung bình trọng số không chú ý đến sự vắng mặt của loài.
- ✓ Trung bình trọng số phụ thuộc vào phân bố của biến môi trường.

## Chương 7. Hồi quy trong sinh thái QXTV

- Theo Braak và Looman (1986), phương pháp trung bình trọng số có hiệu lực tương tự như phương pháp hồi quy logit Gauss khi:
  - ✓ (1) loài hiếm gặp;
  - ✓ (2) loài có vùng phân bố hẹp;
  - ✓ (3) phân bố của biến môi trường tương đối đồng nhất trên những nơi loài xuất hiện trong môi trường này.

## Chương 7. Hồi quy trong sinh thái QXTV

- Một số nhà nghiên cứu cho rằng, trung bình trọng số cũng có thể cho kết quả sai lệch.

## Chương 7. Hồi quy trong sinh thái QXTV

- Đối với số liệu định lượng, ước lượng tối ưu sinh thái bằng phương pháp trung bình trọng số chỉ cho kết quả giống với phương pháp hồi quy logit Gauss khi:
  - ✓ Độ phong phú của loài phân bố theo luật Poisson.
  - ✓ Lập địa đặc trưng cho toàn bộ biên độ sinh thái của loài là thuần nhất.

# Chương 7. Hồi quy trong sinh thái QXTV

## 7.9. MÔ HÌNH LOGISTIC

### 7.9.1. Giới thiệu hồi quy logistic

- Phân tích hồi quy logistic là kỹ thuật thông dụng nhất để mô hình hóa những biến phụ thuộc nhị phân (nhận hai giá trị hoặc 0 hoặc 1).

## Chương 7. Hồi quy trong sinh thái QXTV

- Hồi quy logistic mô tả mối quan hệ giữa biến phụ thuộc nhị phân ( $Y$ ) với một số biến dự đoán ( $X_1, X_2, \dots, X_k$ ).

## Chương 7. Hồi quy trong sinh thái QXTV

- Mô hình logistic có dạng như sau:

$$E(Y) = \frac{1}{1 + \exp[-(b_0 + \sum_{j=1}^k b_j X_j)]} \quad (7.29)$$

## Chương 7. Hồi quy trong sinh thái QXTV

- Đối với các biến ngẫu nhiên  $Y$  nhận hai giá trị 1 và 0,  $E(Y)$  tương đồng với xác suất  $p(Y = 1)$ .

$$p(Y = 1) = \frac{1}{1 + \exp[-(b_0 + \sum_{j=1}^k b_j X_j)]} \quad (7.30)$$



## Chương 7. Hồi quy trong sinh thái QXTV

### 7.9.2. Ước lượng tỷ lệ khác nhau bằng hồi quy logistic

- Trong mô hình 7.30, hệ số  $b_j$  mô tả mối quan hệ giữa  $Y$  với các biến dự đoán ( $X_j$ ).
- Mô hình logistic xác định mối quan hệ bao gồm một tham số được gọi là tỷ lệ sai khác (odds ratio).

## Chương 7. Hồi quy trong sinh thái QXTV

- **Chẳng hạn:**
- ✓ Trong lâm nghiệp **tỷ lệ sai khác đo đạc** ảnh hưởng của các yếu tố môi trường đến sự xuất hiện loài cây nào đó.

## Chương 7. Hồi quy trong sinh thái QXTV

- Bằng cách đo đạc ảnh hưởng, chúng ta ngụ ý một số đo so sánh hai hay nhiều nhóm trong việc dự đoán biến phụ thuộc.

## Chương 7. Hồi quy trong sinh thái QXTV

- Để mô tả tỷ lệ sai khác, trước hết chúng ta cần định nghĩa sự sai khác (Odds).
- ✓ Sự sai khác là tỷ số giữa xác suất mà một số sự kiện xuất hiện chia cho xác suất không xuất hiện những sự kiện ấy.

## Chương 7. Hồi quy trong sinh thái QXTV

- Như vậy, sự sai khác (Odds) đối với một số sự kiện D được tính theo công thức:

$$\text{Odds}(D) = \frac{p(D)}{1 - p(D)}$$

## Chương 7. Hồi quy trong sinh thái QXTV

- Ví dụ:
  - ✓ Khi  $p(D) = 0,25$ , thì  $\text{odds}(D) = 0,25/(1-0,25) = 1/3$ .
  - ✓ Điều đó có nghĩa là xác suất xuất hiện sự kiện D bằng  $1/3$  so với xác suất không xuất hiện sự kiện D.

## Chương 7. Hồi quy trong sinh thái QXTV

- Tỷ lệ sai khác (odds ratio = OR) là tỷ lệ của hai sự sai khác.
- Tỷ lệ của hai sự sai khác ( $OR_{A-B}$ ) được tính theo công thức:

$$OR_{(A-B)} = \text{odds}(D_A) / \text{Odds}(D_B) = \frac{p(D_A)}{1 - p(D_A)} / \frac{p(D_B)}{1 - p(D_B)}$$

## Chương 7. Hồi quy trong sinh thái QXTV

- Ở đây A và B là hai nhóm (đối tượng nghiên cứu) đem so sánh.



## Chương 7. Hồi quy trong sinh thái QXTV

- Ví dụ:
- ✓ Khi so sánh nguy cơ phát triển bệnh vàng lá giữa hai nhóm cây được bón phân chuồng hoai (PC) và không được bón phân chuồng hoai (NPC), thì  $D_{PC}$  mô tả những cây được bón phân chuồng hoai phát triển bệnh vàng lá, còn  $D_{NPC}$  mô tả số cây không được bón phân chuồng hoai phát triển bệnh vàng lá.

## Chương 7. Hồi quy trong sinh thái QXTV

- Ví dụ:
- ✓ Nếu  $p_{(DPC)} = 0,10$  và  $p_{(DNPC)} = 0,25$ , thì OR so sánh tỷ lệ bệnh vàng lá ở những cây được bón phân chuồng hoai với những cây không được bón phân chuồng hoai.

## Chương 7. Hồi quy trong sinh thái QXTV

- ✓ Từ số liệu trên ta có:

$$OR_{(NPC-PC)} = (0,25/1 - 0,25)/(0,10/1 - 0,10) = 3$$

- ✓ Điều đó cho thấy, so với những cây được bón phân chuồng hoai, những cây không được bón phân chuồng hoai có nguy cơ mắc bệnh vàng lá cao hơn 3 lần.

## Chương 7. Hồi quy trong sinh thái QXTV

- ✓ Nếu  $OR_{(NPC-PC)} = 1$ , thì hai nhóm có nguy cơ mắc bệnh vàng lá ngang nhau.
- Kết quả nghiên cứu đã chứng tỏ điều gì?
- ✓ Bệnh vàng lá không có quan hệ với việc bón phân chuồng.

## Chương 7. Hồi quy trong sinh thái QXTV

- **Làm phù hợp hàm logistic**
- ✓ Để làm phù hợp hàm logistic, người ta sử dụng dạng logit của mô hình.
- ✓ Logit là sự biến đổi xác suất  $p(Y = 1)$ ; đó là logarit cơ số neper của odds(D), nghĩa là  $\ln[\text{Odds}(D)]$ .

# Chương 7. Hồi quy trong sinh thái QXTV

- Nói khác đi

$$\text{logit}[p(Y=1)] = \ln[\text{odds}(Y=1)] = \ln\left[\frac{p(Y=1)}{1 - p(Y=1)}\right] \quad (7.31)$$

## Chương 7. Hồi quy trong sinh thái QXTV

- Từ hàm 7.30 và 7.31

$$p(Y = 1) = \frac{1}{1 + \exp[-(b_0 + \sum_{j=1}^k b_j X_j)]} \quad (7.30)$$

$$\text{logit}[p(Y=1)] = \ln[\text{odds}(Y=1)] = \ln\left[\frac{p(Y=1)}{1 - p(Y=1)}\right] \quad (7.31)$$

## Chương 7. Hồi quy trong sinh thái QXTV

sau khi biến đổi chúng ta có:

$$\text{logit}[p(Y=1)] = b_0 + \sum_{j=1}^k b_j X_j \quad (7.32)$$

$$\text{Hay } \ln\left[\frac{p(Y=1)}{1 - p(Y=1)}\right] = b_0 + \sum_{j=1}^k b_j X_j$$



## Chương 7. Hồi quy trong sinh thái QXTV

- **Hàm 7.32** là dạng logit của mô hình hồi quy logistic.
- Dạng logit là một hàm tuyến tính.

## Chương 7. Hồi quy trong sinh thái QXTV

- Giả sử chỉ xem xét ảnh hưởng của việc bón phân chuồng ( $X_1$ ) đến sự phát triển bệnh vàng lá, chúng ta có thể viết phương trình logit như sau:

$$\ln[\text{odds}_{(\text{bón phân chuồng})}] = b_0 + b_1 X_1$$

$$\text{Hay } \ln(p/1-p) = b_0 + b_1 X_1$$

## Chương 7. Hồi quy trong sinh thái QXTV

- Để thu được biểu thức đối với tỷ lệ sai khác từ mô hình logistic, chúng ta phải so sánh sự sai khác giữa hai nhóm cây.
- Ở ví dụ trên có hai nhóm so sánh:
  - ✓ Nhóm cây được bón phân chuồng hoai ( $X_1 = 1$ );
  - ✓ Nhóm cây không được bón phân chuồng hoai ( $X_1 = 0$ ).

## Chương 7. Hồi quy trong sinh thái QXTV

- Từ đó chúng ta có thể viết
  - ✓  $\ln[\text{odds}_{(\text{bón phân chuồng})}] = b_0 + b_1 * 1 = b_0 + b_1$
  - ✓  $\ln[\text{odds}_{(\text{không bón phân chuồng})}] = b_0 + b_1 * 0 = b_0$

## Chương 7. Hồi quy trong sinh thái QXTV

- Tỷ lệ sai khác về bệnh vàng lá

$$\mathbf{OR}_{(PC-NPC)} = \mathbf{odds}_{(PC)} / \mathbf{odds}_{(NPC)} = \frac{e^{(b_0 + b_1)}}{e^{b_0}} = e^{b_1}$$

## Chương 7. Hồi quy trong sinh thái QXTV

- **Như vậy**, nếu chỉ sử dụng một biến dự đoán nhận hai giá trị 0 và 1, thì việc so sánh tỷ lệ sai khác của hai nhóm được thực hiện bằng cách lấy số mũ cơ số neper của hệ số của biến dự đoán trong mô hình.
- **Nói chung**, khi so sánh tỷ lệ sai khác, chúng ta có thể định nghĩa hai nhóm.

## Chương 7. Hồi quy trong sinh thái QXTV

- Hai nhóm so sánh với nhau dựa trên một tập hợp các biến dự đoán  $X_1, X_2, \dots, X_k$ .
- ✓ Nhóm A:  $X_A = (X_{A1}, X_{A2}, \dots, X_{Ak})$ .
- ✓ Nhóm B:  $X_B = (X_{B1}, X_{B2}, \dots, X_{Bk})$ .

## Chương 7. Hồi quy trong sinh thái QXTV

- Ví dụ:
- Khi sử dụng ba biến dự đoán ( $k = 3$ ), trong đó:
  - ✓  $X_1 =$  bón phân chuồng (1 = bón phân chuồng, 0 = không bón phân chuồng);
  - ✓  $X_2 =$  tuổi cây;
  - ✓  $X_3 =$  dòng (1 = dòng miền Bắc, 0 = dòng miền Nam)



## Chương 7. Hồi quy trong sinh thái QXTV

thì chúng ta xây dựng mô hình logit như sau:

$$\text{logit}(p/1-p) = b_0 + b_1X_1 + b_2X_2 + b_3X_3$$

## Chương 7. Hồi quy trong sinh thái QXTV

- Tỷ lệ sai khác giữa hai nhóm cây (bón phân chuồng và không bón phân chuồng) được tính theo công thức:

$$✓ OR_{(X_A-X_B)} = \text{odds}_{(X_A)} / \text{odds}_{(X_B)} = \exp(b_0 + \sum_{j=1}^k b_j X_A) / \exp(b_0 + \sum_{j=1}^k b_j X_B)$$

$$✓ \text{Hay } OR_{(X_A-X_B)} = \exp[\sum_{j=1}^k (X_A - X_B) b_j] \quad (7.33)$$

## Chương 7. Hồi quy trong sinh thái QXTV

- Ví dụ:
- So sánh bệnh vàng lá ở nhóm cây được bón phân chuồng (1) và nhóm cây không được bón phân chuồng (0), cây 5 tháng tuổi và dòng miền Bắc (1), chúng ta viết như sau:
  - ✓  $X_A = (1, 5, 1)$
  - ✓  $X_B = (0, 5, 1)$

## Chương 7. Hồi quy trong sinh thái QXTV

- Từ công thức (7.33), chúng ta tính tỷ lệ sai khác giữa hai nhóm cây như sau:

✓ 
$$OR_{(XA-XB)} = \exp(1 - 0)b_1 + (5 - 5)b_2 + (1-1)b_3$$

Hay 
$$OR_{(XA-XB)} = \exp(b_1)$$

## Chương 7. Hồi quy trong sinh thái QXTV

- Nếu so sánh  $X_A = (0, 4, 1)$  với  $X_B = (1, 2, 1)$ , chúng ta tính tỷ lệ sai khác giữa hai nhóm cây như sau:

✓ 
$$OR_{(X_A-X_B)} = \exp(0 - 1)b_1 + (4 - 2)b_2 + (1-1)b_3$$

Hay 
$$OR_{(X_A-X_B)} = \exp(-b_1) + 2b_2$$

## Chương 7. Hồi quy trong sinh thái QXTV

- Nếu cần xem xét ảnh hưởng tương tác của các biến trong mô hình logit, chúng ta có thể viết mô hình như sau:

$$\text{logit}(p/1-p) = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_1*X_2 + b_5X_1*X_3$$

## Chương 7. Hồi quy trong sinh thái QXTV

- Lưu ý rằng, giả sử chúng ta cần xem xét tỷ lệ sai khác về bệnh vàng lá của hai nhóm cây; biến xem xét bao gồm:
  - ✓  $\text{pH} = X_1$
  - ✓ Hai biến kiểm tra: tuổi ( $X_2$ ) và dòng ( $X_3$ ).

## Chương 7. Hồi quy trong sinh thái QXTV

- Chúng ta viết mô hình logit như sau:

$$\text{logit}(p/1-p) = b_0 + b_1X_1 + b_2X_2 + b_3X_3$$



## Chương 7. Hồi quy trong sinh thái QXTV

- Mô hình tương tác có dạng:

✓ 
$$\text{logit}(p/1-p) = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_1 * X_2 + b_5X_1 * X_3$$

## Chương 7. Hồi quy trong sinh thái QXTV

- ✓ Khi so sánh tình trạng bệnh của nhóm cây sống ở môi trường có độ pH = 6 và 4, từ công thức (7.33) chúng ta viết OR như sau:

$$OR_{(pH=6 - pH=4 | \text{tuổi, dòng})} = \exp(6 - 4)b_1 - (\text{tuổi} - \text{tuổi})b_2 + (\text{dòng} - \text{dòng})b_3 = \exp(2b_1)$$

## Chương 7. Hồi quy trong sinh thái QXTV

- Nói chung, nếu chỉ định so sánh pH là  $pH_0$  và  $pH_1$ , thì tỷ lệ sai khác (OR) được viết như sau:

$$OR_{(pH_1-pH_0 | \text{tuổi, dòng})} = \exp[(pH_1-pH_0)b_1]$$

## Chương 7. Hồi quy trong sinh thái QXTV

- Khoảng tin cậy của  $OR = e^{b_1}$  được tính theo công thức:

$$\text{Exp}(b_1^* \pm Z_{1-\alpha/2} S_{b_1}) \quad (7.34)$$

trong đó

- ✓  $b_1^*$  là ước lượng của hệ số  $b_1$
- ✓  $S_{b_1}$  là sai số của  $b_1^*$
- ✓  $1 - \alpha$  là mức tin cậy.

# Chương 7. Hồi quy trong sinh thái QXTV

## 7.10. PHÂN BỐ POISSON

### 7.10.1. Giới thiệu chung

- Hồi quy Poisson giả thiết rằng, biến phản hồi  $Y$  có phân bố Poisson.

## Chương 7. Hồi quy trong sinh thái QXTV

- Phân bố xác suất Poisson với tham số  $\mu$  được tính theo công thức:

$$P(Y; \mu) = \frac{\mu^y e^{-\mu}}{Y!}, \text{ với } Y = 1, 2, \dots, \infty \quad (7.35)$$

## Chương 7. Hồi quy trong sinh thái QXTV

- Phân bố Poisson được sử dụng nhằm mục đích gì?
- ✓ Mô tả sự xuất hiện của những sự kiện hiếm.

## Chương 7. Hồi quy trong sinh thái QXTV

- **Chẳng hạn:**
- ✓ Mô tả số cây tái sinh của một loài nào đó xuất hiện trong ô dạng bản.
- ✓ Mô tả số cây đạt đến một lớp chiều cao sau k năm.
- ✓ Mô tả số cây của một loài nào đó bị bệnh rỗng ruột ở một lớp tuổi hay lớp D nào đó...



## Chương 7. Hồi quy trong sinh thái QXTV

### 7.10.2. Ví dụ về phân bố Poisson

- Để minh họa tính ứng dụng của phân tích hồi quy Poisson, chúng ta xem xét số liệu so sánh số cá thể của loài sao đen thuộc những cấp kính khác nhau bị bệnh rỗng ruột ở vùng A và vùng B (bảng 7.6).

# Chương 7. Hồi quy trong sinh thái QXTV

**Bảng 7.6.** So sánh số cây sao đen bị bệnh rỗng ruột ở vùng A và B

TT	Cấp D, cm	Vùng A			Vùng B			RRi
		$Y_{iA}$	$(l_{iA})$	$\lambda_{iA}$	$Y_{iB}$	$(l_{iB})$	$\lambda_{iB}$	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1	10-20	2	5.500	0,00036	5	2.500	0,00200	5,5
2	21-30	4	6.000	0,00067	8	4.500	0,00178	2,7
3	31-40	5	8.500	0,00059	12	12.000	0,00100	1,7
4	41-50	8	10.000	0,00080	20	6.850	0,00292	3,6
5	51-60	10	9.500	0,00105	25	5.800	0,00431	4,1
6	61-70	11	5.500	0,00200	28	4.500	0,00622	3,1
7	71-80	12	4.550	0,00264	22	3.000	0,00733	2,8
8	> 81	6	10.000	0,00060	15	8.500	0,00176	2,9

(\*) Nhóm A được sử dụng làm đối tượng so sánh

## Chương 7. Hồi quy trong sinh thái QXTV

- Ở ví dụ này, biến phụ thuộc  $Y$  là số cây sao đen bị bệnh rỗng ruột (cột 3 và 6).
- Vì có tám cấp kính và 2 địa phương được xem xét, chúng ta kí hiệu:

## Chương 7. Hồi quy trong sinh thái QXTV

- ✓  $Y_{ij}$  = số cây sao đen bị bệnh ở cấp kính  $i$  thuộc địa phương  $j$ ;
- ✓  $I_{ij}$  = kích thước quần thể sao đen được nghiên cứu bệnh rỗng ruột ở cấp kính  $i$  thuộc vùng  $j$ ;
- ✓  $i = 1, 2, \dots, 8$ ;
- ✓  $j = 0$  (vùng A);
- ✓  $j = 1$  (vùng B).

## Chương 7. Hồi quy trong sinh thái QXTV

- **Mục đích của nghiên cứu này** là xác định sự rủi ro đối với bệnh rỗng ruột ở những cây sao đen thuộc những cấp kính khác nhau mọc ở vùng A có lớn hơn vùng B hay không.
- **Thuật ngữ rủi ro** biểu thị xác suất phát triển bệnh rỗng ruột ở cây sao đen.

## Chương 7. Hồi quy trong sinh thái QXTV

- Gọi  $\lambda_{ij}$  là tỷ lệ rủi ro thực tế của quần thể sao đen ở cấp kính  $i$  thuộc vùng  $j$ .
- Giá trị  $\lambda_{ij} = Y_{ij}/I_{ij}$  (cột 5 và 8 bảng 7.6)

## Chương 7. Hồi quy trong sinh thái QXTV

- Để so sánh sự khác nhau về tỷ lệ sao đen bị bệnh rỗng ruột ở vùng B so với vùng A, chúng ta sử dụng công thức:

$$RR_i = \frac{\lambda_{i1}}{\lambda_{i0}} \quad (7.36)$$

## Chương 7. Hồi quy trong sinh thái QXTV

- Từ công thức 7.36 cho thấy:
- ✓ Nếu  $RR_i = 1$ , thì tình trạng sao đen bị rỗng ruột ở hai vùng là bằng nhau.
- ✓ Nếu  $RR_i > 1$ , thì tỷ lệ cây sao đen bị bệnh rỗng ruột ở vùng B cao hơn vùng A.
- ✓ Nếu  $RR_i < 1$ , thì tỷ lệ cây sao đen bị bệnh rỗng ruột ở vùng B nhỏ hơn vùng A.



**Bảng 7.6.** So sánh số cây sao đen bị bệnh rỗng ruột ở vùng A và B

TT	Cấp D, cm	Vùng A			Vùng B			RR <sub>i</sub>
		Y <sub>iA</sub>	(I <sub>iA</sub> )	λ <sub>iA</sub>	Y <sub>iB</sub>	(I <sub>iB</sub> )	λ <sub>iB</sub>	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1	10-20	2	5.500	0,00036	5	2.500	0,00200	5,5
2	21-30	4	6.000	0,00067	8	4.500	0,00178	2,7
3	31-40	5	8.500	0,00059	12	12.000	0,00100	1,7
4	41-50	8	10.000	0,00080	20	6.850	0,00292	3,6
5	51-60	10	9.500	0,00105	25	5.800	0,00431	4,1
6	61-70	11	5.500	0,00200	28	4.500	0,00622	3,1
7	71-80	12	4.550	0,00264	22	3.000	0,00733	2,8
8	> 81	6	10.000	0,00060	15	8.500	0,00176	2,9

(\*) Nhóm A được sử dụng làm đối tượng so sánh

- Số liệu (RR<sub>i</sub>) ở cột 9 đã chứng tỏ số cây sao đen ở vùng B bị bệnh rỗng ruột cao hơn so với vùng A.

## Chương 7. Hồi quy trong sinh thái QXTV

- Mục tiêu sử dụng phân tích hồi quy Poisson ở ví dụ này là gì?
- ✓ Làm phù hợp hồi quy Poisson với số liệu.
- ✓ Sử dụng mô hình hồi quy Poisson để ước lượng tỷ lệ cây sao đen bị bệnh rỗng ruột cho tất cả các cấp kính ở hai địa phương A và B.
- ✓ Xác định sự sai khác về tỷ lệ cây sao đen bị bệnh rỗng ruột cho tất cả các cấp kính ở hai địa phương A và B.

## Chương 7. Hồi quy trong sinh thái QXTV

- Lưu ý rằng  $Y_{ij}$  là biến số nhị thức có phân bố ngẫu nhiên với trung bình  $\mu_{ij} = l_{ij} * \lambda_{ij}$ .

## Chương 7. Hồi quy trong sinh thái QXTV

- Để phát triển mô hình hồi quy Poisson, chúng ta gọi  $E(Y_{ij})$  là số kỳ vọng của những cây bị bệnh rỗng ruột ở cấp kính  $i$  thuộc vùng  $j$ .
- Ở đây có hai biến quan tâm là cấp kính và địa phương.

## Chương 7. Hồi quy trong sinh thái QXTV

- Vì cấp kính được phân thành 8 lớp, nên chúng ta sử dụng  $8 - 1 = 7$  biến giả để chỉ định cho biến cấp kính.
- Biến địa phương là 2, chúng ta sẽ sử dụng một biến giả.

## Chương 7. Hồi quy trong sinh thái QXTV

- Như vậy, mô hình biểu thị số cây sao đen bị bệnh rỗng ruột ở cấp kính  $i$  được viết như sau:

$$E(Y_{ij}) = \mu_{ij} = l_{ij} * \lambda_{ij}; \text{ với } i = 1, 2, \dots, 8; j = 0, 1.$$

$$\ln(\lambda_{ij}) = \alpha + \sum_{k=1}^7 \alpha_k U_k + bE \quad (7.37)$$

$$\text{Hay } \lambda_{ij} = \exp(\alpha + \sum_{k=1}^7 \alpha_k U_k + bE)$$

## Chương 7. Hồi quy trong sinh thái QXTV

- trong đó:
  - ✓  $U_k = 1$  nếu  $k = i$
  - ✓  $U_k = 0$  nếu  $k \neq i$ , với  $k = 1, 2, \dots, 7$
  - ✓  $E = 1$  nếu  $j = 1$  (vùng B)
  - ✓  $E = 0$  nếu  $j = 0$  (vùng A)

## Chương 7. Hồi quy trong sinh thái QXTV

- Khi sử dụng mô hình 7.37, chúng ta có thể viết sự rủi ro  $\lambda_{ij}$  ở dạng tham số  $\alpha_j$  và  $b$  để thu được:



# Chương 7. Hồi quy trong sinh thái QXTV

$$\checkmark \ln(\lambda_{i0}) = \alpha + \alpha_i$$

$$\checkmark \ln(\lambda_{i1}) = \alpha + \alpha_i + b, \text{ với } i = 1, 2, \dots, 7 \quad (7.38a)$$

-----

$$\checkmark \text{ Khi } i = 8, \text{ thì } \ln(\lambda_{80}) = \alpha$$

$$\checkmark \ln(\lambda_{81}) = \alpha + b \quad (7.38b)$$

## Chương 7. Hồi quy trong sinh thái QXTV

✓ Khi  $k = i = 8$  thì  $U_k = 0$

✓ Do đó,  $\ln(*\lambda_{i1}) - \ln(*\lambda_{i0}) = (\alpha + \alpha_i + b) - (\alpha + \alpha_i) = b$

✓  $\ln(*\lambda_{81}) - \ln(*\lambda_{80}) = \alpha + b - \alpha = b$  (7.38b)

## Chương 7. Hồi quy trong sinh thái QXTV

Mặt khác,  $RR_i = \frac{\lambda_{i1}}{\lambda_{i0}} = \exp[\ln(\frac{\lambda_{i1}}{\lambda_{i0}})]$

$$\exp[\ln(\frac{\lambda_{i1}}{\lambda_{i0}})] = \exp[\ln(\lambda_{i1}) - \ln(\lambda_{i0})] = \exp[b] = e^b$$

với  $i = 1, 2, \dots, 8$ .

## Chương 7. Hồi quy trong sinh thái QXTV

- **Như vậy**, khi sử dụng mô hình Poisson, chúng ta có thể ước lượng tỷ lệ cây sao đen bị bệnh rỗng ruột cho bất kỳ cấp kính nào bằng cách:
  - ✓ **Trước hết**, làm phù hợp mô hình và ước lượng hệ số  $b$  của biến  $E$ .
  - ✓ **Sau đó**, tính  $e^b$ .

## Chương 7. Hồi quy trong sinh thái QXTV

- Vì tỷ lệ cây bị bệnh rỗng ruột ( $e^b$ ) là độc lập với  $i$ , nên  $e^b$  là ước lượng tỷ lệ cây bị rỗng ruột của tất cả các cấp kính (hay toàn bộ đối tượng nghiên cứu) ở hai vùng A và B.

## Chương 7. Hồi quy trong sinh thái QXTV

- Lưu ý
- ✓ Ví dụ trên đây minh họa kiểu mô hình được sử dụng để mô tả phân tích hồi quy Poisson.
- ✓ Trong thực tế, chúng ta có thể sử dụng không chỉ hai biến số (cấp kính và địa phương), mà còn nhiều hơn hai biến số ( $k$  biến) dự đoán  $(X_1, X_2, \dots, X_k)$  để nghiên cứu.

# Chương 7. Hồi quy trong sinh thái QXTV

## 7.10.3. Hồi quy Poisson

- Để phân tích hồi quy Poisson, chúng ta gọi  $Y$  là biến phụ thuộc.
- Ví dụ số cây bị bệnh rỗng ruột thu được từ một nhóm cây được mô tả bằng tập biến dự đoán  $X_1, X_2, \dots, X_k$ .

# Chương 7. Hồi quy trong sinh thái QXTV

- Đối với nhóm  $i$  ( $i = 1, 2, \dots, n$ ), chúng ta gọi:
  - $Y_i$  = số cây bị bệnh rỗng ruột;
  - $l_i$  = tổng số cây của mỗi cấp kính được nghiên cứu bệnh rỗng ruột;
  - $X_i = (X_{i1}, X_{i2}, \dots, X_{ik})$  = tập hợp các giá trị của  $X_1, X_2, \dots, X_k$  của cấp kính  $i$ ;
  - $b = (b_0, b_1, \dots, b_k)$  = tập hợp những tham số chưa biết;
  - $\lambda(X_i, b)$  = hàm số đặc trưng đối với  $X_i$  và  $b$ ; nghĩa là

$$\lambda(X_i, b) = \exp(b_0 + \sum_{j=1}^k b_j X_{ij}).$$



## Chương 7. Hồi quy trong sinh thái QXTV

- Hàm  $\lambda(x_i, b)$  mô tả tỷ lệ cây bị bệnh rỗng ruột ở nhóm  $i$ .
- Khi đó số cây bị bệnh rỗng ruột trong nhóm  $i$  là

$$E(Y_i) = \mu_i = I_i \lambda(x_i, b) \quad (7.39)$$

- ✓  $i = 1, 2, \dots, n$ ;
- ✓  $Y_i =$  biến ngẫu nhiên có phân bố Poisson.
- ✓ Điều kiện  $\lambda(x_i, b) > 0$ .

## Chương 7. Hồi quy trong sinh thái QXTV

❖ Khi  $Y_i$  có phân bố Poisson với trung bình  $\mu_i$ , thì

✓ 
$$P(Y_i; \mu_i) = \frac{\mu_i^{y_i} e^{-\mu_i}}{Y_i!}, \quad i = 1, 2, \dots, n. \quad (7.40)$$

❖ Từ 7.39

$$✓ E(Y_i) = \mu_i = l_i \lambda(x_i, \mathbf{b}) \quad (7.39)$$

và 7.40

$$✓ P(Y_i; \mu_i) = \frac{\mu_i^{y_i} e^{-\mu_i}}{Y_i!}, \quad i = 1, 2, \dots, n. \quad (7.40)$$

chúng ta có:

$$✓ P(Y_i; \mathbf{b}) = \frac{l_i \lambda(X_i, \mathbf{b})^{y_i} e^{-l_i \lambda(X_i, \mathbf{b})}}{Y_i!} \quad (7.41)$$

## Chương 7. Hồi quy trong sinh thái QXTV

- ✓ Trong công thức 7.41,  $Y_i = 0, 1, \dots, \infty$  và  $i = 1, 2, \dots, n$ .
- ✓ Ví dụ ở bảng 7.6  $Y_i$  chỉ số cây bị bệnh rỗng ruột ở những cấp kính và địa phương khác nhau.

## Chương 7. Hồi quy trong sinh thái QXTV

- Dưới đây chúng ta thử xây dựng một số mô hình biểu diễn ảnh hưởng của cấp kính ( $i = 1, 2, \dots, n$ ) và địa phương ( $j = 0$  và  $j = 1$ ) đến tỷ lệ bệnh rỗng ruột ở loài sao đen.

## Chương 7. Hồi quy trong sinh thái QXTV

- **Mô hình thứ 1:** Mô tả tỷ lệ bệnh rỗng ruột ở loài sao đen phụ thuộc vào cấp kính (U) và địa phương (E).
- ✓ Số cây bị bệnh rỗng ruột:  $E(Y_{ij}) = \mu_{ij} = I_{ij}^* \lambda_{ij}$

$$\text{với } \ln(\lambda_{ij}) = \alpha + \sum_{k=1}^7 \alpha_k U_k + bE \quad (7.42)$$

## Chương 7. Hồi quy trong sinh thái QXTV

- Trong công thức 7.42
- ✓  $U_k$  = biến giả nhận hai giá trị 0 và 1 để biểu thị nhóm cấp kính;
- ✓  $E$  = biến giả nhận giá trị 0 và 1 để mô tả địa phương (vùng A = 0, vùng B = 1).

## Chương 7. Hồi quy trong sinh thái QXTV

- Các tham số của mô hình 7.42 được xác định bằng nguyên lý hợp lý tối đa.
- Tính phù hợp của mô hình hồi quy Poisson cũng được kiểm định bằng thống kê Chi-square.



## Chương 7. Hồi quy trong sinh thái QXTV

- Từ mô hình Poisson, chúng ta có thể tính được tỷ lệ cây bị bệnh rỗng ruột chung ở tất cả các cấp đường kính theo công thức:

$$RR_i = \frac{\lambda_{i1}}{\lambda_{i0}} = e^b. \quad (7.43)$$

✓ Khoảng tin cậy 95% của mẫu lớn đối với  $e^b$  được tính theo công thức:

$$\text{Exp}[b^{\wedge} \pm S_{e(Sb^{\wedge})}] \quad (7.44)$$

trong đó

- ✓  $b^{\wedge}$  = giá trị ước lượng của hệ số  $b$ ;
- ✓  $S_{e(Sb^{\wedge})}$  = sai số chuẩn của hệ số ước lượng  $b^{\wedge}$ .

## Chương 7. Hồi quy trong sinh thái QXTV

- Việc kiểm định giả thuyết  $H_0: b = 0$ , đối thuyết  $H_A: b \neq 0$  có thể được thực hiện dựa trên thống kê Wald:

$$Z = \frac{\hat{b} - 0}{S_{e(\hat{Sb})}} \quad (7.45)$$

ở đây  $Z$  có phân bố  $N(0,1)$  dưới giả thuyết  $H_0: b = 0$ .

## Chương 7. Hồi quy trong sinh thái QXTV

- Từ ví dụ ở bảng 7.6, bằng thuật toán thống kê chúng ta có thể xác định được mô hình Poisson biểu thị tỷ lệ cây sao đen bị bệnh rỗng ruột tùy thuộc vào cấp đường kính và địa phương (bảng 7.7 – 7.9).

# Chương 7. Hồi quy trong sinh thái QXTV

**Bảng 7.7.** Mô hình hồi quy Poisson ước lượng

Tham số	Ước lượng	Sai số chuẩn	RRi <sup>(*)</sup>
Hằng số	-7,46934	0,24666	
Vùng	1,14918	0,15785	3,16
Cấp kính: 1	-0,08693	0,43698	0,92
Cấp kính: 2	0,04080	0,36189	1,04
Cấp kính: 3	-0,44180	0,32655	0,64
Cấp kính: 4	0,44010	0,28876	1,55
Cấp kính: 5	0,79180	0,27625	2,21
Cấp kính: 6	1,24452	0,27067	3,47
Cấp kính: 7	1,44770	0,27767	4,25

(\*) Nhóm A được sử dụng làm đối tượng so sánh

# Chương 7. Hồi quy trong sinh thái QXTV

**Bảng 7.8.** Phân tích biến động<sup>1</sup>

Nguồn	Biến động	df	P
Mô hình	126,871	8	0,0000
Sai lệch	2,5459	7	0,9236
Tổng	129,417	15	

# Chương 7. Hồi quy trong sinh thái QXTV

**Bảng 7.9.** Kiểm định Likelihood Ratio

Yếu tố	$\chi^2$	df	P
Vùng	58,69	1	0,0000
Cấp kính	75,72	7	0,0000

## Chương 7. Hồi quy trong sinh thái QXTV

- Kết quả ở bảng 7.7 cho thấy mô hình Poisson biểu thị tỷ lệ cây sao đen bị bệnh rỗng ruột ở các cấp kính như sau:

✓

$$\lambda_{ij} = \exp(-7,46934 + 1,14918 \cdot \text{vùng} - 0,08693 \cdot D=1 + 0,0408 \cdot D=2 - 0,4418 \cdot D=3 + 0,4401 \cdot D=4 + 0,7918 \cdot D=5 + 1,2445 \cdot D=6 + 1,4477 \cdot D=7)$$

(mô hình 1)

## Chương 7. Hồi quy trong sinh thái QXTV

- Khi sử dụng mô hình (1) để dự đoán số cây sao đen bị bệnh rỗng ruột ( $Y_{it}$ ) ở 8 cấp kính thuộc hai vùng A và B, thì kết quả nhận được như bảng 7.10.



# Chương 7. Hồi quy trong sinh thái QXTV

**Bảng 7.10.** Kết quả dự đoán số cây sao đen bị bệnh rỗng ruột ở vùng A và B

TT	Cấp D, cm	Vùng A			Vùng B			RRi <sup>(*)</sup>
		Y <sub>iA</sub>	(I <sub>iA</sub> )	$\lambda_{iA}$	Y <sub>iB</sub>	(I <sub>iB</sub> )	$\lambda_{iB}$	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1	10-20	2,9	5.500	0,00052	4,1	2.500	0,00165	3,2
2	21-30	3,5	6.000	0,00055	8,4	4.500	0,00187	3,4
3	31-40	3,1	8.500	0,00037	13,9	12.000	0,00116	3,2
4	41-50	8,9	10.000	0,00089	19,1	6.850	0,00280	3,2
5	51-60	12,0	9.500	0,00126	23,0	5.800	0,00397	3,2
6	61-70	10,9	5.500	0,00198	28,1	4.500	0,00625	3,2
7	71-80	11,0	4.550	0,00243	23,0	3.000	0,00766	3,2
8	> 81	5,7	10.000	0,00057	15,3	8.500	0,00180	3,2

(\*) Nhóm A được sử dụng làm đối tượng so sánh

## Chương 7. Hồi quy trong sinh thái QXTV

- Lưu ý rằng
- ✓ Các giá trị  $\lambda_{ij}$  ở cột 5 và 8 của bảng 7.10 được tính từ công thức 7.38(a,b).
- $\ln(\lambda_{i0}) = \alpha + \alpha_i$  và  $\ln(\lambda_{i1}) = \alpha + \alpha_i + b$ , với  $i = 1, 2, \dots, 7$ . (7.38a)
- Khi  $i = 8$ , chúng ta có  $\ln(\lambda_{80}) = \alpha$  và  $\ln(\lambda_{81}) = \alpha + b$  (7.38b)

## Chương 7. Hồi quy trong sinh thái QXTV

- **Kết quả tính toán cho thấy**, tỷ lệ bệnh rỗng ruột ở các cấp kính như sau:

+ Đối với vùng A

$$\lambda_{1A} = \exp(-7,46934 - 0,08693*1) = \exp(-7,55627)$$

$$\lambda_{2A} = \exp(-7,46934 + 0,0408*1) = \exp(-7,51014)$$

$$\lambda_{3A} = \exp(-7,46934 - 0,4418*1) = \exp(-7,91114)$$

$$\lambda_{4A} = \exp(-7,46934 + 0,4401*1) = \exp(-7,02924)$$

$$\lambda_{5A} = \exp(-7,46934 + 0,7918*1) = \exp(-6,67754)$$

$$\lambda_{6A} = \exp(-7,46934 + 1,2445*1) = \exp(-6,22484)$$

$$\lambda_{7A} = \exp(-7,46934 + 1,4477*1) = \exp(-6,02164)$$

$$\lambda_{8A} = \exp(-7,46934)$$

# Chương 7. Hồi quy trong sinh thái QXTV

+ Đối với vùng B

$$\lambda_{1B} = \exp(-7,46934 + 1,14918*1 - 0,08693*1)$$

$$\lambda_{2B} = \exp(-7,46934 + 1,14918*1 + 0,0408*1)$$

$$\lambda_{3B} = \exp(-7,46934 + 1,14918*1 - 0,4418*1)$$

$$\lambda_{4B} = \exp(-7,46934 + 1,14918*1 + 0,4401*1)$$

$$\lambda_{5B} = \exp(-7,46934 + 1,14918*1 + 0,7918*1)$$

$$\lambda_{6B} = \exp(-7,46934 + 1,14918*1 + 1,2445*1)$$

$$\lambda_{7B} = \exp(-7,46934 + 1,14918*1 + 1,4477*1)$$

$$\lambda_{8B} = \exp(-7,46934 + 1,14918*1)$$

## Chương 7. Hồi quy trong sinh thái QXTV

- Ở ví dụ trên đây hệ số ước lượng biểu thị ảnh hưởng của vùng sinh thái đến tỷ lệ cây sao đen bị bệnh rỗng ruột là  $b^{\wedge} = 1,14918$ , với sai số  $Se(s_{b^{\wedge}}) = 0,15785$ .
- Giả thuyết  $H_0: b = 0$  bị bác bỏ, bởi vì  $Z = 7,28$  với  $P < 0,01$ .

## Chương 7. Hồi quy trong sinh thái QXTV

- Do đó, ước lượng tỷ lệ sai khác về bệnh rỗng ruột ở sao đen giữa hai vùng  $RR_i = e^{b^{\wedge}} = e^{1,14918} = 3,16$ .
- Khoảng tin cậy 95% của  $e^{b^{\wedge}} = \exp[b^{\wedge} \pm 1,96 * Se(s_{b^{\wedge}})] = \exp(1,14918 \pm 1,96 * 0,15785) = (2,32; 4.30)$ .

## Chương 7. Hồi quy trong sinh thái QXTV

- Bây giờ chúng ta xem xét hai câu hỏi khác như sau:
  1. Tỷ lệ cây sao đen bị bệnh rỗng ruột ở những cấp kính khác nhau của hai vùng có khác nhau hay không?
  2. Nếu tỷ lệ cây sao đen bị bệnh rỗng ruột ở những cấp kính khác nhau của hai vùng là không khác nhau, thì chúng ta có cần phải xem xét ảnh hưởng của cấp kính hay không?

## Chương 7. Hồi quy trong sinh thái QXTV

- Để trả lời câu hỏi 1, chúng ta xây dựng mô hình 2 bằng cách đưa thêm thành phần tương tác giữa cấp kính với địa phương ( $U_k * E$ ) vào mô hình 1.
- Mô hình 2 có dạng:

$$\ln(\lambda_{ij}) = \alpha + \sum_{k=1}^7 \alpha_k U_k + bE + \sum_{k=1}^7 \delta_k (U_k * E) \quad (7.46)$$



## Chương 7. Hồi quy trong sinh thái QXTV

- Ở mô hình 2, ý nghĩa của tương tác giữa cấp kính với địa phương được kiểm định thông qua giả thuyết  $H_0: \delta_1 = \delta_2 = \dots \delta_7 = 0$ .
- Kiểm định này được thực hiện bằng thống kê Chi-Square với độ tự do bằng số cấp kính ở hai vùng trừ đi số tham số trong mô hình. Trong trường hợp của ví dụ ở bảng 7.6, độ tự do  $= 16 - 9 = 7$ .

## Chương 7. Hồi quy trong sinh thái QXTV

- Để trả lời câu hỏi 2, chúng ta xây dựng mô hình 3 bằng cách bỏ thành phần cấp đường kính ở mô hình 1.

## Chương 7. Hồi quy trong sinh thái QXTV

- Mô hình 3 có dạng:

$$\ln(\lambda_{ij}) = \alpha + bE \quad (7.47)$$

## Chương 7. Hồi quy trong sinh thái QXTV

- Nếu bỏ yếu tố cấp đường kính, chúng ta cần phải xem xét hệ số  $b^{\wedge}$  hoặc  $e^{b^{\wedge}}$  thay đổi có ý nghĩa hay không.
- Nếu xử lý mô hình 3 theo số liệu của bảng 7.6, chúng ta sẽ nhận được kết quả sau đây:

## Chương 7. Hồi quy trong sinh thái QXTV

$$\ln(\lambda_{ij}) = -6,9341 + 1,0678 * \text{vùng}$$

$$\text{Hay } \lambda_{ij} = \exp(-6,9341 + 1,0678 * \text{vùng})$$

# Chương 7. Hồi quy trong sinh thái QXTV

## ▪ Khi xử lý với mô hình 1:

- ✓ Hệ số  $b^{\wedge} = 1,14918$
- ✓  $Se(s_{b^{\wedge}}) = 0,15785$
- ✓  $RR_i = e^{b^{\wedge}} \exp(1,14918) = 3,16$
- ✓ Khoảng tin cậy 95% của  $e^{b^{\wedge}} = (2,32; 4.30)$

## ▪ Khi xử lý với mô hình 3:

- ✓ Hệ số  $b^{\wedge} = 1,0678$
- ✓  $S_e(s_{b^{\wedge}}) = 0,1570$
- ✓  $RR_i = e^{b^{\wedge}} = e^{1,0678} = 2,91$
- ✓ Khoảng tin cậy 95% của  $e^{b^{\wedge}} = (2,14; 3,95)$

## Chương 7. Hồi quy trong sinh thái QXTV

- Như vậy, so với mô hình 1, hệ số  $b^2$  ở mô hình 3 thay đổi đáng kể.
- Điều đó chứng tỏ rằng, để dự đoán tỷ lệ bệnh rỗng ruột ở sao đen chúng ta không thể bỏ qua yếu tố cấp đường kính.
- Nhận định này có phù hợp với lý thuyết hay không?

## Chương 7. Hồi quy trong sinh thái QXTV

- **Bây giờ chúng ta xem xét mô hình 4** chỉ bao gồm ảnh hưởng tuyến tính của yếu tố đường kính, **không có tương tác giữa đường kính với vùng (E).**



## Chương 7. Hồi quy trong sinh thái QXTV

- Mô hình 4 có dạng:

$$\ln \lambda_{ij}^{\wedge} = \alpha + \theta \cdot \ln D_i + bE \quad (7.48)$$

## Chương 7. Hồi quy trong sinh thái QXTV

- ✓ Trong đó  $D_i$  được biến đổi như sau:

$$D_i = \frac{[\text{Giá trị giữa của cấp đường kính thứ } i] - 10}{35}$$

với  $i = 1, 2, \dots, 8$ .

# Chương 7. Hồi quy trong sinh thái QXTV

Từ số liệu của bảng 7.6, sau khi tính toán chúng ta có (bảng 7.11):

**Bảng 7.11.** Giá trị  $\ln\lambda_{i0}$  và  $\ln\lambda_{i1}$  được biến đổi từ số liệu ở bảng 7.6

Cấp D, cm	Giữa cấp D	$D_i$	$\ln D_i$	$\ln\lambda_{i0}$	$\ln\lambda_{i1}$
(1)	(2)	(3)	(4)	(5)	(6)
10-20	15	0,14286	-1,9459	-7,9194	-6,2146
21-30	25	0,42857	-0,8473	-7,3132	-6,3324
31-40	35	0,71429	-0,3365	-7,4384	-6,9078
41-50	45	1,00000	0,0000	-7,1309	-5,8363
51-60	55	1,28571	0,2513	-6,8565	-5,4467
61-70	65	1,57143	0,4520	-6,2146	-5,0796
71-80	75	1,85714	0,6190	-5,938	-4,9153
> 81	85	2,14286	0,7621	-7,4186	-6,3398

# Chương 7. Hồi quy trong sinh thái QXTV

- Mô hình 4 cho biết:

- ✓  $\lambda_{i0} = e^{\alpha} D_i^{\theta}$
- ✓  $\lambda_{i1} = e^{\alpha} D_i^{\theta} e^b$
- ✓  $RR_i = \frac{\lambda_{i1}}{\lambda_{i0}} = e^b.$

# Chương 7. Hồi quy trong sinh thái QXTV

Khi  $i = 4$  ( $D_4 = 1$ ) và  $j = 0$ , chúng ta có:

$$\lambda_{40} = e^{\alpha} D_4^{\theta} = e^{\alpha} (1)^{\theta} = e^{\alpha}$$

$$\ln \lambda_{40} = \alpha$$

## Chương 7. Hồi quy trong sinh thái QXTV

- Do đó  $\alpha$  (điểm chặn trong mô hình 4) là logarit cơ số neper của tỷ lệ bệnh rỗng ruột thuộc nhóm tuổi 41-50.

## Chương 7. Hồi quy trong sinh thái QXTV

- Từ số liệu ở bảng 7.6 và 7.11, chúng ta nhận được kết quả như sau (bảng 7.12-7.14):

# Chương 7. Hồi quy trong sinh thái QXTV

**Bảng 7.12.** Mô hình hồi quy Poisson ước lượng

Tham số	Ước lượng	Sai số chuẩn	RRi
Hằng số	-6,97112	0,13327	
Vùng	1,06251	0,15700	2,89361
$\ln T_i$	0,50885	0,12541	1,66337

**Bảng 7.14.** Kiểm định Likelihood Ratio

Yếu tố	$\chi^2$	df	P
Vùng	50,632	1	0,0000
Cấp kính	18,866	1	0,0000

**Bảng 7.13.** Phân tích biến động

Nguồn	Biến động	df	P
Mô hình	70,0166	2	0,0000
Sai lệch	59,4001	13	0,0000
<b>Tổng</b>		<b>15</b>	



## Chương 7. Hồi quy trong sinh thái QXTV

- Mô hình hồi quy ước lượng:

$$Y_{ij} = \exp(-6,97112 + 1,06251 * \text{vùng} + 0,50885 * \ln Ti)$$

## Chương 7. Hồi quy trong sinh thái QXTV

- Do đó, ước lượng tỷ lệ sai khác về bệnh rỗng ruột ở loài sao đen mọc ở hai vùng sinh thái khác nhau là

$$RR_i = e^{b^{\wedge}} = \exp(1,06251) = 2,89.$$

## Chương 7. Hồi quy trong sinh thái QXTV

- Khoảng tin cậy 95% của  $e^{b^{\wedge}}$ :

$$\begin{aligned} e^{b^{\wedge}} &= \exp[b^{\wedge} \pm 1,96 * Se(s^{b^{\wedge}})] \\ &= \exp(1,06251 \pm 1,96 * 0,157) \\ &= (2,13; 3,94). \end{aligned}$$

## Chương 7. Hồi quy trong sinh thái QXTV

- Như vậy, sử dụng mô hình 4 cũng dẫn đến câu trả lời tương tự như mô hình 1, nhưng độ tin cậy cao hơn (bởi vì khoảng tin cậy của mô hình 4 hẹp hơn).

## Chương 7. Hồi quy trong sinh thái QXTV

- So với mô hình 1, mô hình 4 có những ưu điểm gì?
- ✓ Đơn giản hơn, vì chứa ít tham số hơn.
- ✓ Có hiệu lực cao hơn, bởi vì mô hình 4 chỉ mô tả ảnh hưởng tuyến tính của tuổi.

## Chương 7. Hồi quy trong sinh thái QXTV

- Vì thế, mô hình 4 được chọn để mô tả tỷ lệ bệnh rỗng ruột ở loài sao đen mọc ở hai địa phương khác nhau.

Hết chương 7