

GIỚI THIỆU MỘT SỐ PHƯƠNG PHÁP THỐNG KÊ NÂNG CAO



*PGS.TS. Hoàng Văn Minh
Hà nội- tháng 12 năm 2012*

Nội dung:

1. Khái niệm hồi quy
2. Hồi quy tuyến tính
3. Hồi quy logistic
4. Phân tích số liệu lặp lại
5. Phân tích sống còn

HỒI QUY

- ▶ Hồi qui là mô hình toán học thể hiện sự biến đổi của một biến số (biến phụ thuộc) theo một hay nhiều biến khác (biến độc lập = biến giải thích)
- ▶ Mục tiêu là ước lượng (hay dự đoán) giá trị của biến phụ thuộc trên cơ sở các giá trị đã biết của biến độc lập

HỒI QUY TUYẾN TÍNH

$$Y = a + bx_1 + cx_2 + dx_3 \dots$$

- ▶ Biến phụ thuộc là biến định lượng, phân bố chuẩn
- ▶ Biến độc lập (giải thích) có thể là định tính hoặc định lượng
- ▶ Các biến định tính cần chuyển sang các biến giả (có giá trị 1, 0)

Thực hành

use huyetap.dta, clear

xi: regress hatdtb i.gioitinh i.nhomtuoi

HỒI QUY LOGISTIC

$$\text{odds} = \frac{r}{1-r} = 0 \rightarrow +\infty$$

$$\text{logit} = \log(\text{odds}) = \ln\left(\frac{r}{1-r}\right) = -\infty \rightarrow +\infty$$

$$\log \text{it} = \ln\left(\frac{r}{1-r}\right) = y = b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n$$

HỒI QUY LOGISTIC

$$\log it = \ln\left(\frac{r}{1-r}\right) = y = b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n$$

- ▶ Biến phụ thuộc là biến nhị phân (1, 0)
- ▶ Biến độc lập có thể là định tính hoặc định lượng, biến định tính

HỒI QUY LOGISTIC

$$\log it = \ln\left(\frac{r}{1-r}\right) = y = b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n$$

$$odds = \exp(b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n) = e^{b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n} = e^{b_0} e^{b_1x_1} e^{b_2x_2} \dots e^{b_nx_n}$$

$$\frac{odds}{odds + 1} = \frac{1}{1 + odds^{-1}} = \frac{1}{1 + e^{-(b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n)}}$$

Thực hành

xi: logit benhcotim hutthuoc i.gioitinh i.nhomtuoi

xi: logistic benhcotim hutthuoc i.gioitinh i.nhomtuoi

PHÂN TÍCH SỐ LIỆU NHẮC LẠI

- ▶ Một biến số được đo nhiều lần
- ▶ So sánh các nhóm khác nhau

Group	Subject	Time-point			
		1	2	...	T
1	1	<input type="checkbox"/>	<input type="checkbox"/>	...	<input type="checkbox"/>
	⋮	⋮	⋮	⋮	⋮
	n_1	<input type="checkbox"/>	<input type="checkbox"/>	...	<input type="checkbox"/>
2	1	<input type="checkbox"/>	<input type="checkbox"/>	...	<input type="checkbox"/>
	⋮	⋮	⋮	⋮	⋮
	n_2	<input type="checkbox"/>	<input type="checkbox"/>	...	<input type="checkbox"/>

Thực hành

```
anova heartrate time drug
```

```
anova heartrate time id, repeated (time)
```

```
xi:regress heartrate i.drug i.time
```

```
xtset id
```

```
xi: xtgee heartrate i.drug i.time, family(normal) link(identity)
```

PHÂN TÍCH SỐNG CÒN

- t_i : thời gian bệnh nhân i^{th} tử vong
- $m[t]$: số bệnh nhân có $t_i > t$
- $d[t]$: số bệnh nhân có $t_i \leq t$
- Hàm sống còn (survival function)
 $S[t] = \Pr[t_i > t]$: xác suất sống đến ít nhất thời điểm t
- Hàm tử vong tích lũy (cum. mortality function)
 $D[t] = \Pr[t_i \leq t]$: xác suất tử vong trước thời điểm t

Nghiên cứu sống sót sau can thiệp

	id	survtime	died	treatment
	10	6	song	dieu tri
	11	9	song	dieu tri
	12	10	song	dieu tri
	13	11	song	dieu tri
	14	17	song	dieu tri
	15	19	song	dieu tri
	16	20	song	dieu tri
	17	25	song	dieu tri
	18	32	song	dieu tri
	19	32	song	dieu tri
	20	34	song	dieu tri
	21	35	song	dieu tri
	22	1	tu vong	chung
	23	1	tu vong	chung
	24	2	tu vong	chung
	25	2	tu vong	chung
	26	3	tu vong	chung
	27	4	tu vong	chung
	28	4	tu vong	chung
	29	5	tu vong	chung
	30	5	tu vong	chung
	31	8	tu vong	chung

Đường cong sống còn Kaplan-Meier

Hàm sống còn Kaplan-Meier

$$\hat{S}(t) = \hat{p}_1 \times \hat{p}_2 \times \dots \times \hat{p}_k = \prod_{k: t_k < t} \hat{p}_k = \prod_k \frac{n_k - d_k}{n_k}$$

Kiểm định logrank

Phân tích Logrank sẽ kiểm định cho biết xác suất sống còn của 2 nhóm có khác nhau có ý nghĩa thống kê hay không.

Phương trình hồi quy Cox

$$\frac{h(t)}{h_0(t)} = \exp(x'b)$$

$$\frac{h(t)}{h_0(t)} = \exp(X' B) = e^{b_1x_1 + b_2x_2 + \dots + b_px_p}$$

$$\begin{aligned} \text{hazard ratio}(t, x_1, x_0) &= \frac{h(t, x_1, \beta)}{h(t, x_0, \beta)} \\ &= e^{\beta(x_1 - x_0)} \end{aligned}$$

Thực hành

```
stset survtime, failure( died)
      stsum
      stsum, by( treatment)
      sts list
sts list, by( treatment) compare
      sts graph, yline(0.5)
      sts graph, by(treatment)
sts graph, gwood by(treatment)
      sts graph, by(treatment)na
      sts test treatment
      stcox treatment
```