

PHÂN TÍCH HỒI QUI LOGISTIC BẰNG PHẦN MỀM MEDCALC

Lê Đình Vãn

Bộ môn Giải Phẫu, Trường Đại học Y Dược Huế

Abstract

LOGISTIC REGRESSION ANALYSIS BY MEDCALC SOFTWARE

Regression analysis is a statistical method used to describe the relationship between two variables and to predict one variable from another. The Logistic regression is a tool used to find the best-fitting model for a relationship between one or more independent variables with a dependent variable that is dichotomous, or binary (death / life; disease / no disease; yes / no). Logistic regression generates the coefficients of a formula to predict a logit transformation of the probability of presence of the characteristic of interest:

Logit (p) = $\text{Ln}(\frac{p}{1-p}) = y$ Where $y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_i X_i$ ($X_0, \dots, X_i$: independent variable; $\beta_0, \beta_1, \beta_2, \dots, \beta_i$: Regression coefficients; $e^{\beta_1}, e^{\beta_2}, \dots, e^{\beta_i}$ are the odds ratios (O.R.) for these independent variables $X_1, X_2, \dots, X_i$ (X_i). Run madcalc software, and practice by click these button follow these command: Statistics... Regression... Logistic regression. The dialog display and after entering data in dialog and click OK button, the results are displayed.

1. ĐẠI CƯƠNG

Phân tích hồi quy có thể hiểu một cách nôm na là dự đoán một đại lượng chưa biết (Y) trên cơ sở các đại lượng đã biết (X), như khi biết chiều cao của cơ thể có thể dự đoán được trọng lượng cơ thể, phương trình dự đoán mối liên hệ trên gọi là phương trình hồi quy ("*Hồi quy (regression): là một từ được một nhà khoa học Anh, Francis Galton (1822–1911) sử dụng đầu tiên khi nghiên cứu mối liên hệ chiều cao của con cái và cha mẹ, nếu cha mẹ cao, theo quan niệm chung thì con cái sẽ cao lên, và sau nhiều thế hệ chiều cao sẽ đến vô cực! Galton tìm ra quy luật ngược lại, con sẽ không cao mà có khuynh hướng thấp lại để tiến về giá trị trung bình của nòi giống (hồi quy); khái niệm hồi quy được Udny Yule và Karl Pearson nghiên cứu và sử dụng trong thống kê.*")

Dạng đơn giản nhất là hồi quy tuyến tính cơ bản là:

$$y = a x + b$$

Trong đó y là biến phụ thuộc và x là biến độc lập, b là hằng số ví dụ các nhà nhân trắc khi nghiên cứu mối liên hệ giữa chiều cao và cân nặng đã tìm được phương trình hồi quy sau:

$$\text{Cân nặng (kg)} = 0.7 \times \text{chiều cao (cm)} + 22.5$$

Phương trình trên chỉ áp dụng khi y là đại lượng liên tục (biến định lượng liên tục).

Nhưng trong y học ngoài biến số định lượng liên tục còn các loại biến số khác, trong đó có biến định tính nhị phân: chết/sống; mắc bệnh/không mắc bệnh; có/không...

Khi tìm hiểu mối liên hệ giữa biến phụ thuộc là biến nhị phân với các biến độc lập khác thì không thể sử dụng hồi quy tuyến tính trên, mà phải sử dụng phân tích khác là phân tích hồi qui logistic (logistic regression analysis).

Hồi qui logistic sử dụng phép biến đổi logarit (logit transformation) của xác suất xuất hiện một hiện tượng (p)(xác suất bị bệnh...) với:

$$\text{logit (p)} = Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_i X_i$$

Cơ sở lý thuyết

Gọi p : Xác suất bị mắc bệnh mạch vành

Như vậy: $1-p$: Xác suất không bị bệnh là

Như vậy tỷ suất chênh (xác suất bị bệnh chia cho xác suất không bị bệnh) là:

$$odds = \frac{p}{1-p} \quad (1)$$

Tính logarit tự nhiên của odds ta có

$$\text{logit}(p) = \ln(odds) = \ln\left(\frac{p}{1-p}\right) = y \quad (2)$$

Trong đó: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_i X_i$

Đó là dạng của **phương trình hồi quy logistic**

Với: $X_0, \dots, X_i$ là các biến số độc lập (các yếu tố: tuổi, hút thuốc, giới...)

$B_0, B_1, \beta_2, \dots, \beta_i$ là hệ số hồi quy

Từ (1) và (2) ta có:

$$odds = \frac{p}{1-p} = e^{\beta_0} \times e^{\beta_1 x_1} \times \dots \times e^{\beta_i x_i}$$

Như vậy ta thấy nếu ví dụ X_1 (tuổi) tăng thêm 1 đơn vị thì: odds tăng lên e^{β_1} lần.

Các đại lượng $e^{\beta_1}, e^{\beta_2}, \dots, e^{\beta_i}$ là tỷ suất chênh (odds) của các biến số $X_1, X_2, \dots, X_i$.

(odds là gì: odds là một từ tiếng Anh xuất phát từ các trường đua ngựa, trong đó odds được sử dụng như là tỷ lệ chấp bao nhiêu lần giữa các người cá độ trong đua ngựa ví dụ khi nói “ngựa đỏ (xích thố của Quan Vân Trường) chấp ngựa trắng (bạch mã của Kim Dung) 9 ăn 1 có nghĩa là odds = 9”).

2. PHÂN TÍCH HỒI QUY LOGISTIC BẰNG PHẦN MỀM MEDCALC

Có nhiều phần mềm có thể dùng để phân tích hồi quy, ở đây tôi trình bày những kinh nghiệm bản thân cách sử dụng phần mềm Medcalc, là phần mềm tốt cho thống kê y học.

Ví dụ: Một công trình nghiên cứu về mối liên quan giữa bệnh mạch vành với tuổi và tình trạng hút thuốc trên 76 người, có tập hợp dữ liệu như bảng 1

Các bước tiến hành

STT	Tình trạng bệnh (có =1)	Tuổi	Hút thuốc (có =1)	STT	Tình trạng bệnh (có =1)	Tuổi	Hút thuốc (có =1)
1	0	28	1		1	24	0
2	0	35	0		1	37	1
3	0	32	0		1	30	1
4	1	41	1		1	39	1
5	0	33	1		0	22	0
6	0	21	0		0	44	1
7	0	39	0		1	45	1
8	1	39	1		1	40	0
9	0	38	1		0	39	0

10	1	38	0	1	45	1
11	0	23	0	1	32	1
12	1	38	0	1	36	0
13	0	28	0	0	40	0
14	0	29	0	0	37	0
15	0	27	0	0	31	0
16	0	37	1	0	39	0
17	1	44	1	0	33	1
18	0	32	0	1	27	1
19	0	27	1	1	28	0
20	1	43	1	0	27	0
21	0	22	0	0	30	0
22	1	45	1	0	37	0
23	1	35	1	1	27	0
24	0	28	0	0	21	0
25	0	26	1	0	26	1
26	0	25	0	0	34	1
27	0	35	0	0	32	0
28	0	38	1	0	32	1
29	1	31	1	1	41	1
30	0	29	0	0	29	1
31	0	37	0	0	31	0
32	0	40	0	0	25	0
33	1	30	0	1	45	1
34	0	32	0	1	25	0
35	0	28	1	1	38	1
36	1	45	1	1	38	1
37	0	38	0	0	26	0
38	0	22	0	1	37	0

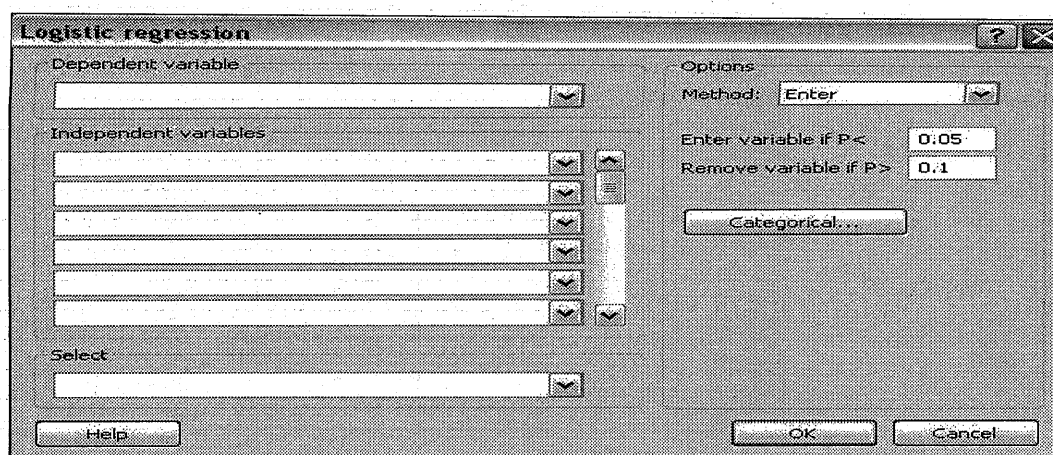
Bước 1. Nhập số liệu. Ta có bảng số liệu như hình 1

Hình 1

Bước 2. Nhấp chuột vào “statistics” ta sẽ có hộp thoại theo hình 2

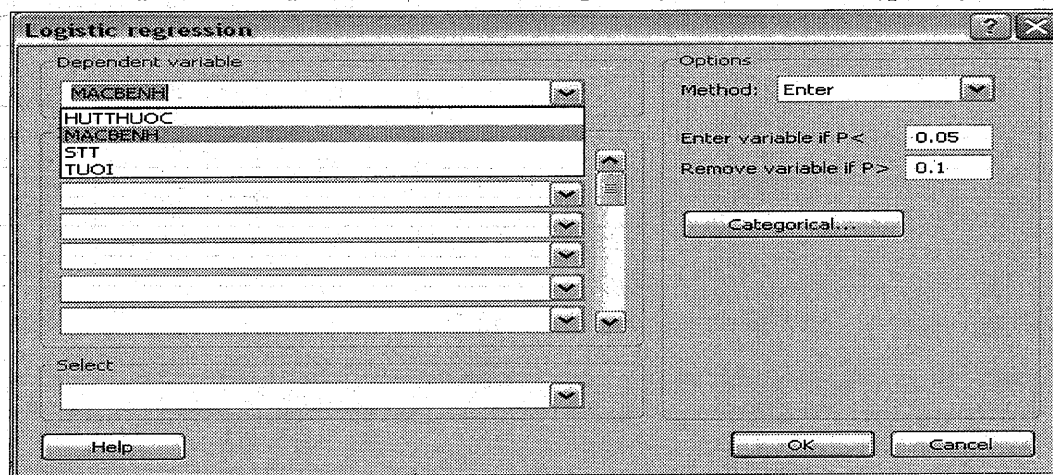
Hình 2

Bước 3. Nhấp chuột vào “Logistic regression” ta sẽ có hộp thoại theo hình 3



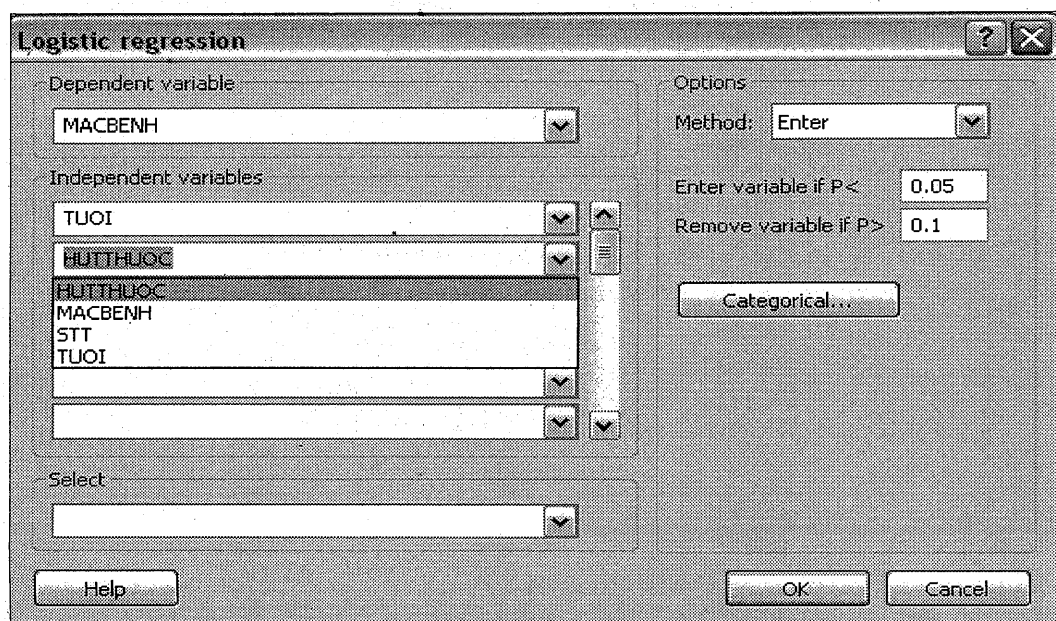
Hình 3

Bước 4. Ở hộp thoại “Dependent varianle” nhấp chuột vào  sẽ có hộp thoại như hình 4



Hình 4

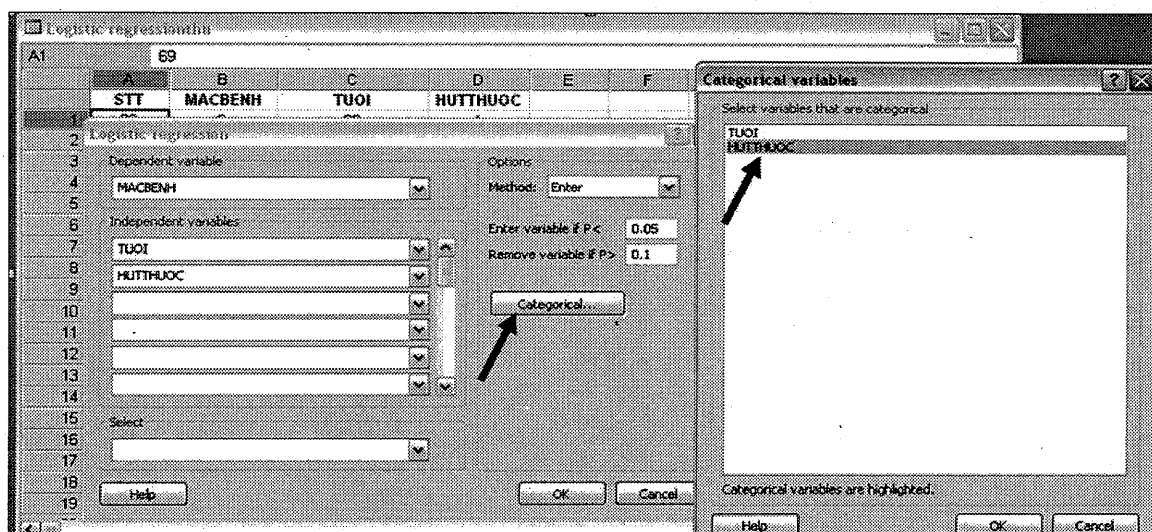
Bước 5. Ở hộp thoại “Independent varianle” thực hiện tương tự để có được hình 5



Hình 5

Nếu biến độc lập có biến định tính (hút thuốc ở ví dụ trên) thì phải thực hiện lệnh sau

Nhấp vào ô **Categorical...** ta sẽ có hộp thoại hình 6:



Hình 6

Nhấp chuột vào mục HUTTHUOC để xác định đó là biến số định tính, nhấp vào ô “OK” sẽ có kết quả như hình 7

Logistic regression

Dependent Y	MACBENH
Method	Enter
Sample size	76
Cases with Y=0	47 (61.84%)
Cases with Y=1	29 (38.16%)

Overall Model Fit

Null model - 2 Log Likelihood	101.054
Full model - 2 Log Likelihood	85.421
Chi-square	15.634
DF	2
Significance level	P = 0.0004

Coefficients and Standard Errors

Variable	Coefficient	Std. Error	P
TUOI	0.1078	0.04541	0.0176
HUTTHUOC=1	1.1398	0.5345	0.0330
Constant	-4.6493		

Odds Ratios and 95% Confidence Intervals

Variable	Odds Ratio	95% CI
TUOI	1.1138	1.0190 to 1.2175
HUTTHUOC=1	3.1262	1.0967 to 8.9115

Classification table (cut-off value p=0.5)

Actual group	Predicted group		Percent correct
	0	1	
Y = 0	40	7	85.11 %
Y = 1	15	14	48.28 %
Percent of cases correctly classified			71.05 %

ROC curve analysis

Area under the ROC curve (AUC)	0.754
Standard Error	0.0598
95% Confidence Interval	0.642 to 0.846

Hình 7

Kết quả ở bảng trên có nhiều phần, các bảng cần xét là:

- Overall model fit

+ Chú ý hàng significance level cho biết giá trị của p; nếu $p < 0.05$, thì trong các biến số độc lập có ít nhất một biến ảnh hưởng đến biến độc lập.

- Coefficients and Standard Errors

Có hệ số hồi quy của tuổi hút thuốc:

$$+ \beta_{\text{tuoi}} = 0.1078 \text{ với } p = 0.0176 < 0.05$$

$$+ \beta_{\text{hutthuoc}} = 1.1398 \text{ với } p = 0.033 < 0.05$$

$$\text{Hằng số} = -4.6498$$

- Odds Ratios and 95% Confidence Intervals

Tỷ suất chênh của các biến số độc lập

$$+ \text{OR tuổi} = 1.1138 \text{ với KTC 95\% (1.0190 - 1.2175).}$$

$$+ \text{OR hút thuốc} = 3.1262 \text{ KTC 95\% (1.0967 - 8.9115)}$$

Có một số người sử dụng đường cong ROC để đánh giá mối liên quan, nên phần mềm medcalc có thêm phần kết quả của phân tích đường cong ROC.

Phương trình hồi quy logistic của bài tập này:

$$Y = -4.6498 + 0.1078 * \text{tuoi} + 1.1398 * \text{hutthuoc}$$

(Hút thuốc có hai giá trị là 1 và 0)

Nếu một người 38 tuổi có hút thuốc, ta có thể tính logit (p) như sau

$$Y = -4.6498 + 0.1078 * 38 + 1.1398 * 1 = 0.5864$$

Từ công thức (1) và (2) ta có thể tính p như sau:

$$p = \frac{1}{1 + e^{-Y}}$$

$$P = 0.64$$

Có thể tra cứu giá trị p bằng cách tra cứu ở bảng sau với logit (p) = 0.5864 thì p = 0.64

Có thể tải phần mềm ở địa chỉ <http://bmgiaphauhue.org/tailieu/MedCalc8.0.1.0/>

p	logit(p)	p	logit(p)	p	logit(p)	p	logit(p)
0.01	-4.5951	0.26	-1.0460	0.51	0.0400	0.76	1.1527
0.02	-3.8918	0.27	-0.9946	0.52	0.0800	0.77	1.2083
0.03	-3.4761	0.28	-0.9445	0.53	0.1201	0.78	1.2657
0.04	-3.1781	0.29	-0.8954	0.54	0.1603	0.79	1.3249
0.05	-2.9444	0.30	-0.8473	0.55	0.2007	0.80	1.3863
0.06	-2.7515	0.31	-0.8001	0.56	0.2412	0.81	1.4500
0.07	-2.5867	0.32	-0.7538	0.57	0.2819	0.82	1.5163

0.08	-2.4423	0.33	-0.7082	0.58	0.3228	0.83	1.5856
0.09	-2.3136	0.34	-0.6633	0.59	0.3640	0.84	1.6582
0.10	-2.1972	0.35	-0.6190	0.60	0.4055	0.85	1.7346
0.11	-2.0907	0.36	-0.5754	0.61	0.4473	0.86	1.8153
0.12	-1.9924	0.37	-0.5322	0.62	0.4895	0.87	1.9010
0.13	-1.9010	0.38	-0.4895	0.63	0.5322	0.88	2.9924
0.14	-1.8153	0.39	-0.4473	0.64	0.5754	0.89	2.0907
0.15	-1.7346	0.40	-0.4055	0.65	0.6190	0.90	2.1972
0.16	-1.6582	0.41	0.3640	0.66	0.6633	0.91	2.3136
0.17	-1.5856	0.42	-0.3228	0.67	0.7082	0.92	2.4423
0.18	-1.5163	0.43	-0.2819	0.68	0.7538	0.93	2.5867
0.19	-1.4500	0.44	-0.2412	0.69	0.8001	0.94	2.7515
0.20	-1.3863	0.45	-0.2007	0.70	0.8473	0.95	2.9444
0.21	-1.3249	0.46	-0.1603	0.71	0.8954	0.96	3.1781
0.22	-1.2657	0.47	-0.1201	0.72	0.9445	0.97	3.4761
0.23	-1.2083	0.48	-0.0800	0.73	0.9946	0.98	3.8918
0.24	-1.1527	0.49	-0.0400	0.74	1.0460	0.99	4.5951
0.25	-1.0986	0.50	0.0000	0.75	1.0986		

TÀI LIỆU THAM KHẢO

1. Alvan R. Feinstein (2002). Principles of Medical Statistics. Copyright 2002 by Chapman & Hall/CRC.
2. David F. Penson (2006). *Clinical Research Methods for Surgeons*. 2006 Humana Press Inc.
3. Dawson, Beth; Trapp, Robert G (2004). *Basic & Clinical Biostatistics*, 4th Edition Copyright 2004 McGraw-Hill.
4. MedCalc Software (2010). MedCalc for Windows. statistics for biomedical research, software manual.
5. Michel Huguier (2000). *Biostatistiques au quotidien*. Elsevier
6. R. B. D'Agostino (2004). *Tutorials in Biostatistics*, Volume 1: Statistical Methods in Clinical Studies. Edited by Boston University, USA. Copyright 2004 John Wiley & Sons Ltd.
7. Tom Fawcett (2005). *An introduction to ROC analysis*. Institute for the Study of Learning and Expertise, 2164 Staunton Court, Palo Alto, CA 94306, USA. Available online 19 December 2005.

PHẪU THUẬT NỘI SOI UNG THƯ ĐẠI TRÀNG TRÁI TẠI BỆNH VIỆN TRUNG ƯƠNG HUẾ

Lê Mạnh Hà, Phạm Trung Vỹ, Phạm Minh Đức
Bộ môn Ngoại, Trường Đại học Y Dược Huế

Tóm tắt

Mục đích: Đánh giá đặc điểm lâm sàng, cận lâm sàng và kết quả điều trị bằng phẫu thuật nội soi ung thư đại tràng (ĐT) trái tại Bệnh viện Trung ương Huế. **Đối tượng và phương pháp:** Gồm 45 bệnh nhân (BN) ung thư đại tràng trái được phẫu thuật cắt đại tràng trái, đại tràng sigma và nạo vét hạch nội soi tại Bệnh viện Trung ương Huế từ tháng 9/2007 đến tháng 9/2010. **Kết quả:** Tuổi trung bình $57,2 \pm 16,8$, tuổi từ 40 - 60 chiếm 51,1%. Nam giới chiếm 62,2%, nữ giới 37,8%, tỷ lệ nam/nữ 1,65/1. Đau bụng là triệu chứng thường gặp nhất 93,3%, rối loạn tiêu hóa 31,1%, phân có máu 53,3%, sờ thấy u ở bụng trái 13,3%. Nội soi cho hình ảnh sùi loét chiếm tỷ lệ cao nhất 42,2% trong đó khối u ở đại tràng sigma 60,9%. Cắt đoạn đại tràng sigma 62,2%, cắt nửa đại tràng trái 17,8%, kích thước khối u trên 5cm là 60,8%. Giai đoạn (GD) ung thư: GD III 64,4% trong đó ung thư biểu mô tuyến 97,8%. Tăng CEA trước mổ là 57,8%, sau mổ 3 tháng là 25,8% và sau mổ 6 tháng là 33,3%. Nhiễm trùng vết mổ 4,4%, rò rỉ miệng nối 2,2%, không có trường hợp nào phải mổ lại và không có tử vong sau mổ. **Kết luận:** Phẫu thuật nội soi ung thư đại tràng trái cho thấy tính an toàn, dễ thực hiện, ít biến chứng và khẳng định được tính ưu việt của nó.

Abstract

LAPAROSCOPIC LEFT HEMICOLECTOMY FOR LEFT COLON CANCER AT HUE CENTRAL HOSPITAL

Le Manh Ha, Pham Trung Vy, Pham Minh Duc

Purpose: Evaluation of clinic, subclinic characteristics and results from laparoscopic left hemicolectomy for left colon cancer at Hue Central Hospital. **Methods:** Consist of 45 patients with left colon cancer who were operated to laparoscopic left hemicolectomy, sigmoidectomy and lymphadectomy at Hue Centre Hospital from september 2007 to september 2010. **Results:** Age: average $57,2 \pm 16,8$, the most age range from 40 - 60, to take 51,1%. Sex: male 62,2%, female 37,8%, rate male/female 1,65/1. The most symptom was abdominal pain 93,3%, alimentary disorder 31,1%, blood stool 53,3%, touch up the tumors 13,3%. Colonoscopy: malignant ulcer tumors were highest rate 42,2%, tumor positions at sigmoid colon 60,9%. Laparoscopic sigmoidectomy 62,2%, left hemicolectomy 17,8%, size of the tumor upper 5 cm was 60,8%. Third stage of TNM 64,4% and adenocarcinoma was 97,8%. Elevation of CEA serum level upper 5ng/ml: Preoperative was 57,8%, postoperative of three months was 25,8%, postoperative of six months was 33,3%. Wound infection was 4,4%, anastomotic leakage 2,2%, and without reoperative case and without surgical mortality. **Conclusions:** Laparoscopic left hemicolectomy for left colon cancer is safe, feasible, few complications and its advantages.