

## § 流行病學原理：資料分析 practice 5

流行病學資料分析 –  
SAS  
Practice 6  
學號：b07401048  
系級：醫學五

### 一. Odds ratio :

(一) Case 與 control 組中各有多少病人的接受連續點滴注射超過 24 小時？

#### 1. Result :

- (1) Case 組共有 23 人的靜脈留置導管接受連續點滴注射超過 24 小時，佔 case 組的 50%。
- (2) Control 組共 30 人的靜脈留置導管接受連續點滴注射超過 24 小時，佔 control 組的 15.96%。

#### 2. Figure :

The FREQ Procedure

Frequency Row Pct	Table of Case by ContIV_24H			
	Case(Case)	ContIV_24H(ContIV 24H)		Total
		1	0	
1		23 50.00	23 50.00	46
0		30 15.96	158 84.04	188
Total		53	181	234
Frequency Missing = 2				

(二) contIV\_24H 這個暴露變項的 odds ratio 為何？是否達統計上的顯著？

1. Result :

(1) Odds ratio 為 2.6221

(2) 達統計上顯著

2. Figure :

Odds Ratio and Relative Risks			
Statistic	Value	95% Confidence Limits	
Odds Ratio	5.2667	2.6221	10.5785
Relative Risk (Column 1)	3.1333	2.0237	4.8514
Relative Risk (Column 2)	0.5949	0.4427	0.7996

Statistics for Table of Case by ContIV\_24H

Statistic	DF	Value	Prob
Chi-Square	1	24.4468	<.0001
Likelihood Ratio Chi-Square	1	21.5636	<.0001
Continuity Adj. Chi-Square	1	22.5423	<.0001
Mantel-Haenszel Chi-Square	1	24.3423	<.0001
Phi Coefficient		0.3232	
Contingency Coefficient		0.3076	
Cramer's V		0.3232	

3. Description :

(1) 兩組在 contIV\_24H 這個變項的 odds ratio 為 5.2667。95% CI：1.6221-10.5785，顯示達統計上顯著。

(2) 以 chi-square 檢定「是 case 還是 control 組」與「是否暴露於 contIV\_24H」兩個事件是否為獨立事件。結果顯示檢定值為 24.4468，p-value 小於 0.001，故可拒絕虛無假說，「是 case 還是 control 組」與「是否暴露於 contIV\_24H」並非獨立事件。

4. Code for Q1 :

```
/* hw5 cohort study */
dm "odsresult" clear;
dm "log" clear;
proc import datafile = "C:\Users\Raymond\Desktop\sas data\Dataset 4.xlsx"
  dbms = xlsx
  out = work.dataset4
  replace;
run;

/* q1 : odds ratio */
title "q1";

proc sort data = dataset4 out = dataset4_sorted;
  by descending case descending contIV_24H;
```

```
run;

proc freq data = dataset4_sorted order = data;
  tables case*contIV_24h / nopercnt nocol OR chisq;
run;

proc logistic data = dataset4;
  class contiv_24h(ref = "0") / param = ref;
  model case(event = "1") = contiv_24h;
run;
```

## 二. Logistic regression :

(一) No selection logistic regression 分析調整 gender, age over 65, service, site 的干擾後，contIV\_24H 的 adjusted odds ratio 為何？

### 1. Result :

(1) 兩組的 adjusted odds ratio 為 6.237，95%CI：2.883-13.493，達統計上顯著。

### 2. Figures :

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
ContIV_24H 1 vs 0	6.237	2.883	13.493
Gender 1 vs 0	0.620	0.303	1.272
VAR6 1 vs 0	1.580	0.738	3.383
Service 1 vs 0	1.263	0.514	3.107
Site 1 vs 0	4.332	1.578	11.893

### 3. Code :

```
/* q2 : logistic regression */
title "q2-1";
/* no selection */
proc logistic data = dataset4;
  class case(ref = "0")
    contiv_24h(ref = "0")
    gender(ref = "0")
    VAR6(ref = "0")
    service(ref = "0")
    site(ref = "0") / param = ref;
  model case = contiv_24h gender VAR6 service site;
run;
```

(二) Stepwise selection logistic regression 分析調整 gender, age over 65, service, site 的干擾後，contIV\_24H 的 adjusted odds ratio 為何？

1. Result :

(1) 兩組的 adjusted odds ratio 為 5.850，95%CI：2.834 - 12.076，達統計上顯著。

2. Figures :

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
ContIV_24H 1 vs 0	5.850	2.834	12.076
Site 1 vs 0	4.027	1.559	10.403

3. Code :

```
title "q2-2";
/* stepwise selection */
proc logistic data = dataset4;
  class case(ref = "0")
    contiv_24h(ref = "0")
    gender(ref = "0")
    VAR6(ref = "0")
    service(ref = "0")
    site(ref = "0") / param = ref;
  model case = contiv_24h gender VAR6 service site /
  selection = stepwise
  slentry = 0.15
  slstay = 0.15
  details;
run;
```

### 三. Conditional logistic regression :

(一) No selection conditional logistic regression 分析調整 gender, age over 65, service, site 的干擾後, contIV\_24H 的 adjusted odds ratio 為何?

1. Result :

(1) 兩組 adjusted odds ratio 為 8.641, 95% CI : 2.262 – 22.890, 達統計上顯著。

2. Figures :

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
ContIV_24H 1 vs 0	8.641	3.262	22.890
Gender 1 vs 0	0.477	0.214	1.065
VAR6 1 vs 0	1.411	0.591	3.370
Service 1 vs 0	0.828	0.189	3.634
Site 1 vs 0	9.327	2.460	35.372

3. Descriptions

4. Code :

```
/* q3 : conditional logistic regression */
title "q3-1";
/* no selection */
proc logistic data = dataset4;
  class case(ref = "0")
    contiv_24h(ref = "0")
    gender(ref = "0")
    VAR6(ref = "0")
    service(ref = "0")
    site(ref = "0") / param = ref;
  model case = contiv_24h gender VAR6 service site;
  strata strata;
run;
```

(二) Stepwise selection conditional selection logistic regression 分析調整 gender, age over 65, service, site 的干擾後，contIV\_24H 的 adjusted odds ratio 為何？

1. Result :

(1) 兩組 adjusted odds ratio 為 9.496，95% CI：3.700 – 24.375，達統計上顯著。

2. Figures :

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
ContIV_24H 1 vs 0	9.496	3.700	24.375
Gender 1 vs 0	0.475	0.213	1.057
Site 1 vs 0	9.612	2.585	35.743

3.

(三) Code :

```
title "q3-2";
/* stepwise selection */
proc logistic data = dataset4;
  class case(ref = "0")
    contiv_24h(ref = "0")
    gender(ref = "0")
    VAR6(ref = "0")
    service(ref = "0")
    site(ref = "0") / param = ref;
  model case = contiv_24h gender VAR6 service site/
  selection = stepwise
  slentry = 0.15
  slstay = 0.15
  details;
  strata strata;
run;
```

#### 四. 對本研究而言，適用 **Logistic regression** 還是 **conditional logistic regression** ?

- (一)線性迴歸的其中一個基本假設是「每一筆資料需彼此獨立」，才可以線性迴歸模式來分析及估計。
- (二)本筆資料中，case 與 control 並非完全獨立。根據 dataset4 的背景說明：對照的選擇則從該醫學中心的住院病人(來源族群 source population)依病房及日期 1:4 個別配對(case: control = 1:4, matched individually by ward and date)以 risk set sampling 方式取樣。
- (三)因此，對本研究而言，並不符合 Logistic regression 的基本假設。需在指令中加入 strata 來配對，以 conditional logistic regression 來估計。

#### 五. 第三題算出的 **adjusted odds ratio** 的意義是什麼？

- (一)首先，第三題有 no selection 與 stepwise selection 兩個小題。首先解釋 selection 的意義：
  1. 線性迴歸的基本假設，除了上題提及的「每一筆資料需彼此獨立外」，尚有「每一個自變項須完全獨立」。然而，流行病學的資料中常常有自變項不完全獨立情形，比如 dataset4 中，有「age」以及「age > 65」兩個變項，顯然兩者並非獨立的變項。
  2. 因此，在流行病學的研究中，若要以線性迴歸進行分析，需要進行 selection 的步驟，以挑選最有代表性的變項，並刪除不需要的或者重複的變項，以此估計出正確的統計結果。
  3. 此外，在迴歸模式中，若可減少重疊或不必要變項，可提昇統計模型的檢力。
  4. 也就是說，本題統計出的 adjusted odds ratio，應以第二小題經過 stepwise selection 的結果為主。
- (二)Stepwise selection：
  1. 設定 entry 及 remove 的顯著水準（本題都設定為 0.15）後，以 entry 及 remove 兩個步驟篩選具有顯著差異的變項。
  2. 本題以上述顯著水準篩選的變項有 contIV\_24H, site 及 gender。

Summary of Stepwise Selection								
Step	Effect		DF	Number In	Score Chi-Square	Wald Chi-Square	Pr > ChiSq	Variable Label
	Entered	Removed						
1	ContIV_24H		1	1	26.6367		<.0001	ContIV 24H
2	Site		1	2	13.8154		0.0002	Site
3	Gender		1	3	3.4128		0.0647	Gender

#### (三)流行病學意義：

1. 在經過上述 stepwise selection 後，挑選出 3 個具顯著差異的變項。
2. 以 conditional Logistic regression 調整上述變項的干擾作用，並計算 contIV\_24H 的 adjusted odds ration。



3. 結果為 9.496，95% CI：3.700 – 24.375，達統計上顯著。表示 case 組暴露於「靜脈留置導管接受連續點滴注射超過 24 小時」顯著高於 control 組。