* 流行病學原理：資料分析practice 5

流行病學資料分析 – SAS

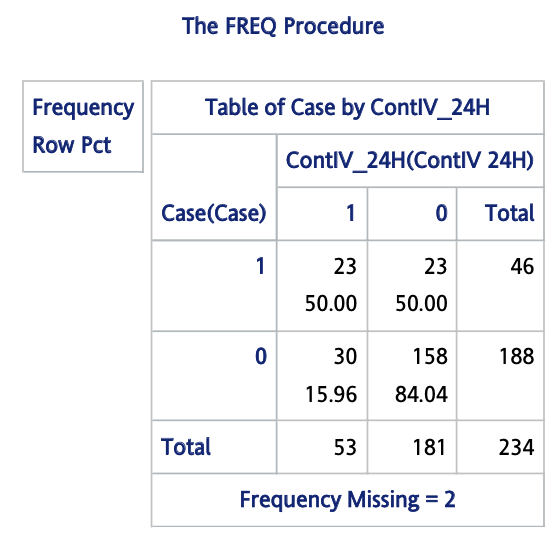
Practice 6

學號：b07401048

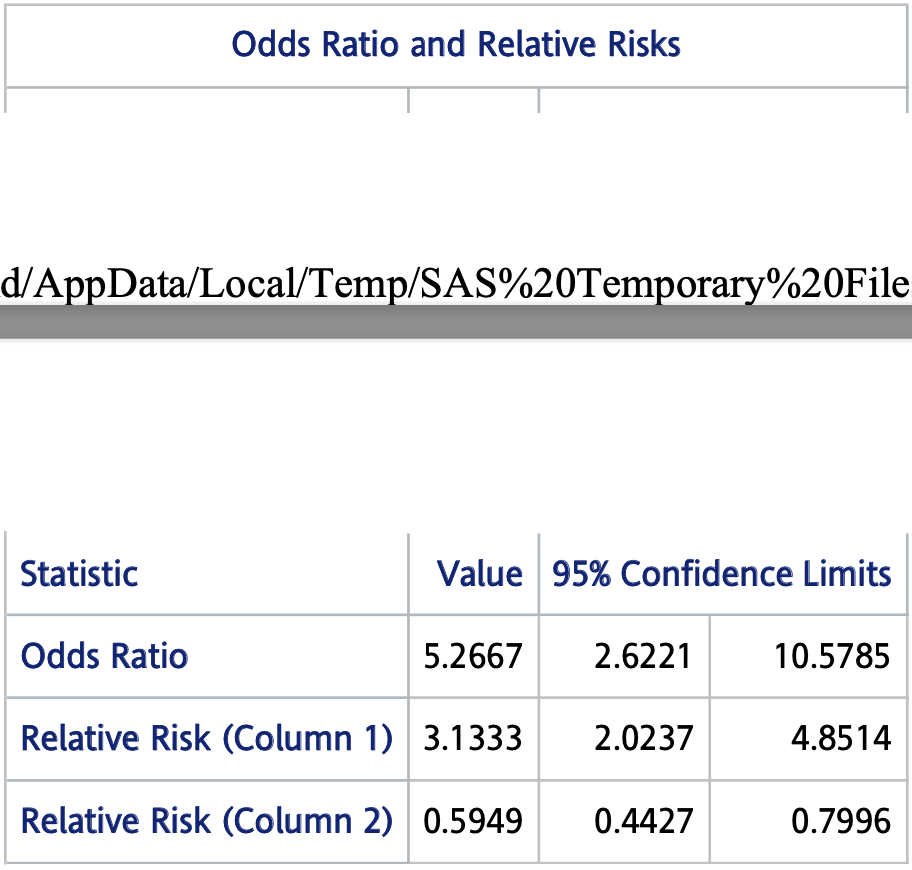
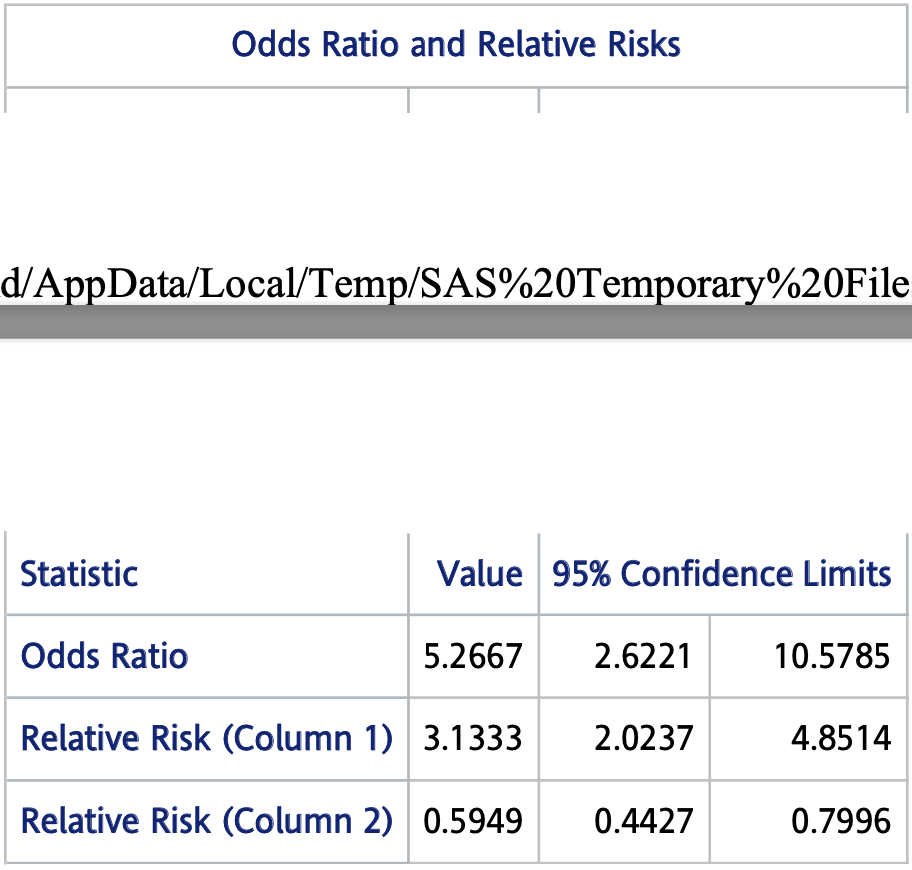
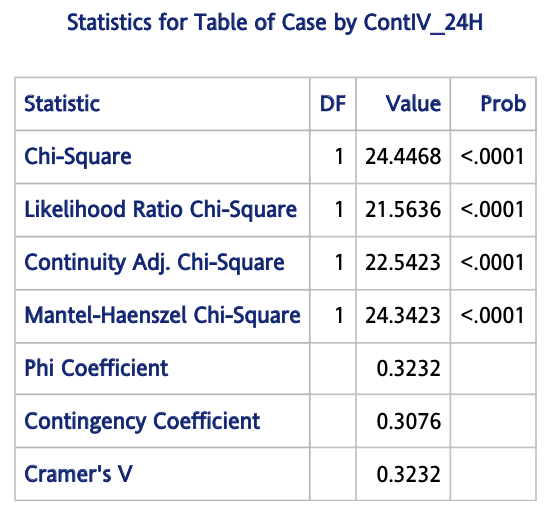
系級：醫學五

姓名：賴柏瑞

* 1. Odds ratio :
     1. 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：



* + 1. 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：



* + - 1. 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。
      2. 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 / nopercent nocol OR chisq;

**run**;

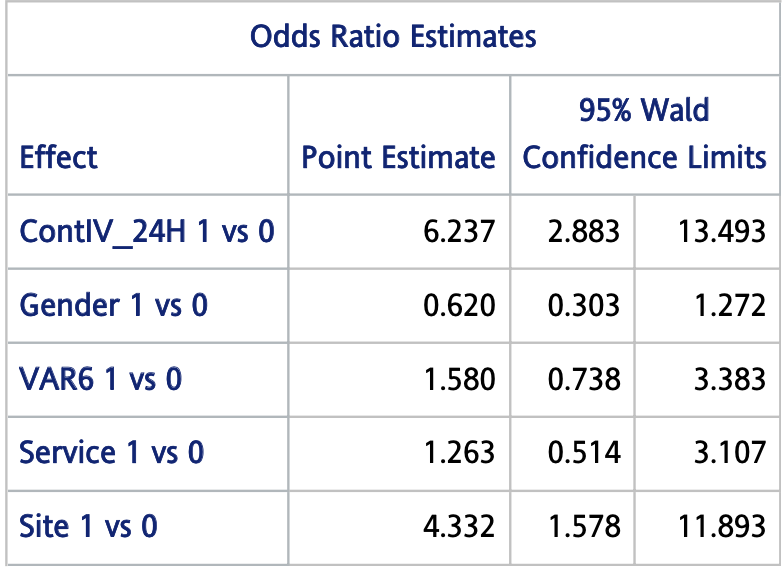
**proc** **logistic** data = dataset4;

class contiv\_24h(ref = "0")/ param = ref;

model case(event = "1") = contiv\_24h;

**run**;

* 1. Logistic regression：
     1. 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：



* + - 1. 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")

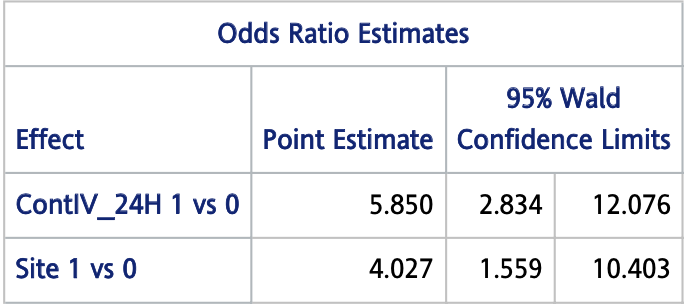
service(ref = "0")

site(ref = "0") / param = ref;

model case = contiv\_24h gender VAR6 service site;

**run**;

* + 1. 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：



* + - 1. 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

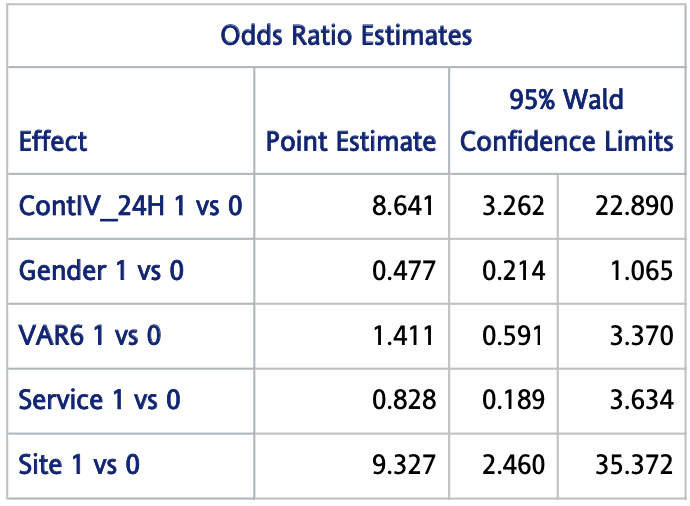
slentry = 0.15

slstay = 0.15

details;

**run**;

* 1. Conditional logistic regression ：
     1. 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：



* + - 1. Descriptions
      2. 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")

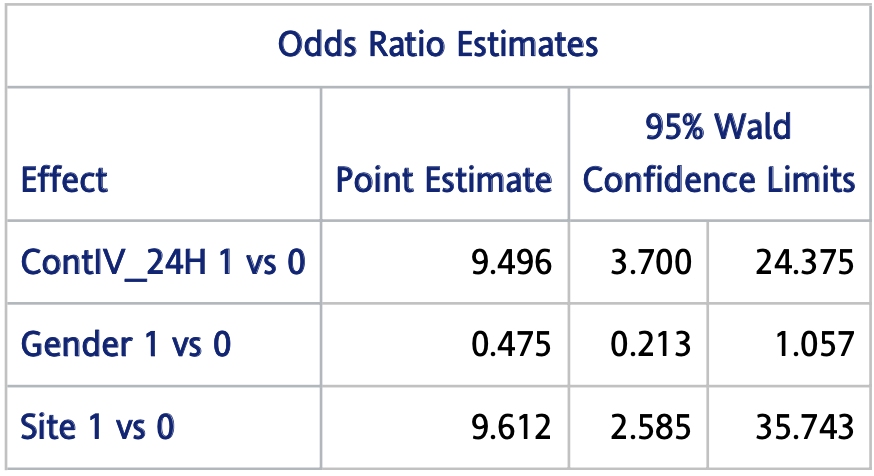
site(ref = "0") / param = ref;

model case = contiv\_24h gender VAR6 service site;

strata strata;

**run**;

* + 1. 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：



* + 1. 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

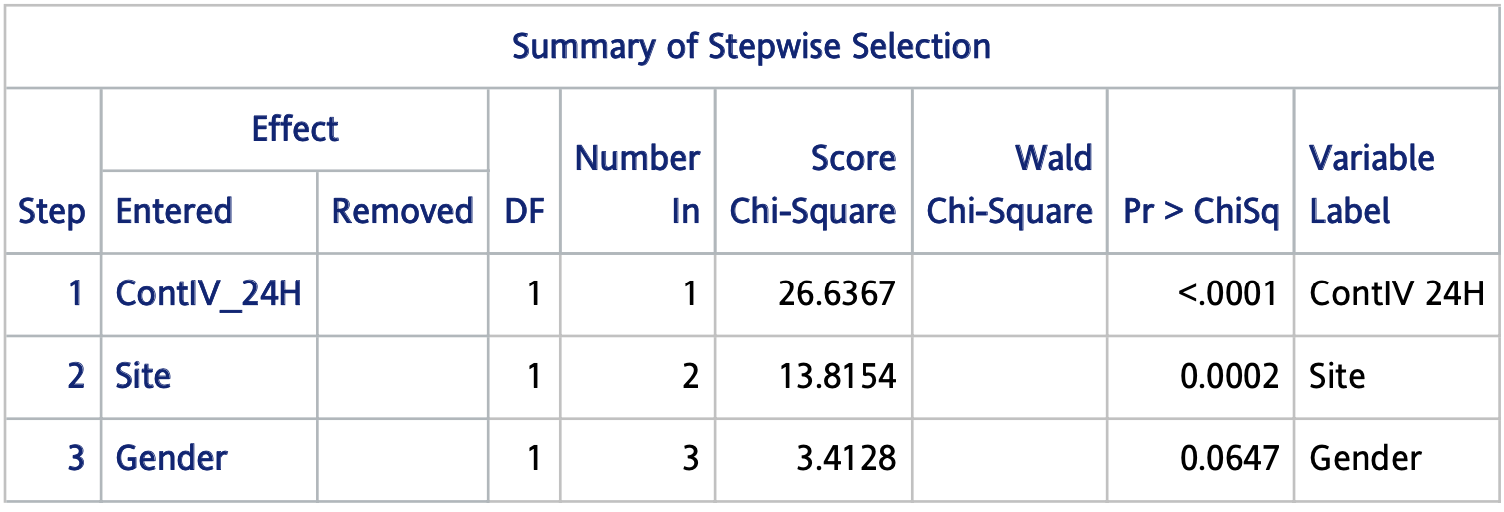
slentry = 0.15

slstay = 0.15

details;

strata strata;

**run**;

* 1. 對本研究而言，適用Logistic regression還是conditional logistic regression？
     1. 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。
     2. 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。
     3. 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。
  2. 第三題算出的adjusted odds ratio的意義是什麼？
     1. 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。
     2. 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。
     3. 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。