

CVD_All 資料中包含 66,489 人的臨床檢驗數據、家族、行為數據、及是否罹患CVD 的數據。

1. 請將” 壓差=收縮壓減舒張壓” 當成應變數，年齡當成自變數，做線性迴歸分析，並檢定年齡是否為壓差的風險因子?

答：

利用資料處理中的新變數建立增加變數”壓差”

R-web 分析步驟：

資料處理→新變數建立→輸入公式(壓差 = 收縮壓 - 舒張壓)並儲存於原檔後

分析結果

- 資料檔名稱：CVD_All
- 變數轉換公式：壓差 = 收縮壓 - 舒張壓
- 儲存位置：使用者個人資料檔 - CVD_All
- 處理時間：2.403 秒

分析年齡(自變數)對於壓差(依變數)的影響，可使用 R-web 中的迴歸模

式R-web 分析步驟：

分析方法→迴歸模式→迴歸分析

→步驟一：資料匯入(選取個人資料檔 CVD_ALL)

→步驟二：參數設定(依變數：壓差、自變數：年齡)

→進階選項(不須設定)

→開始分析

- 分析方法：迴歸分析
- 資料名稱：CVD_All
- 依變數名稱：壓差
- 自變數名稱：年齡
- 顯著水準：0.05
- 計算時間：28.186 秒
- 迴歸模式的變異數分析：

虛無假設：迴歸模式不顯著						
來源 source	平方和 sum of squares	自由 度 d.f.	均方和 mean square	F 檢定統 計量 F-statistic	臨界值 F(d.f.1,d.f.2,1- α)	p-值 ¹ p-value
迴歸 regression	2043759.61	1	2043759.61	11825.6764	3.8416	< 1e-04 ***
誤差 error	10922125.3402	63198	172.8239			
總和 total	12965884.9502	63199				
判定係數(R-square)：15.76 %						
調整判定係數(adjusted R-square)：15.76 %						

- 1：顯著性代碼： '***' : < 0.001, '**' : < 0.01, '*' : < 0.05, '#' : < 0.1

- 迴歸係數估計^I：

係數 coefficient	估計值 estimation	標準差 std. err.	t 檢定統計量 t-statistic	p 值 ^{II} p-value	參數的 95% 信賴區間 95% C.I. for estimations	
					下界 lower	上界 upper
(截距項)	26.0056	0.1837	141.5427	< 2.22e-16 ***	25.6454	26.3657
年齡	0.4093	0.0038	108.7459	< 2.22e-16 ***	0.4019	0.4167

- I：依變數為壓差，模式包含常數項

II：顯著性代碼： '***' : < 0.001, '**' : < 0.01, '*' : < 0.05, '#' : < 0.1

結論：

模式的變異數分析 P-值< 1e-04，表示此迴歸模式顯著，模式的參數估計在年齡變數的 P-值為< 2.22e-16，表示參數為顯著。故由以上結果可得知年齡為壓差的風險因子

2. 接續第 1 題，將性別當成做線性迴歸分析的干擾因子，分析年齡是否為風險因子？性別是否真的是干擾因子？針對壓差這個應變數，年齡和性別是否存有交互作用？

答：

在 1 的模式中加入性別，仍使用迴歸分析

R-web 分析步驟：

分析方法→迴歸模式→迴歸分析

- 步驟一：資料匯入(選取個人資料檔 CVD_ALL)
- 步驟二：參數設定(依變數：壓差、自變數：年齡、性別)
- 進階選項(不須設定)
- 開始分析

分析結果

- 迴歸係數估計^I：

係數 coefficient	估計值 estimation	標準差 std. err.	t 檢定統計量 t-statistic	p 值 ^{II} p-value	參數的 95% 信賴區間 95% C.I. for estimations	
					下界 lower	上界 upper
(截距項)	25.1539	0.1838	136.8223	< 2.22e-16 ***	24.7936	25.5142
年齡	0.3986	0.0037	106.4673	< 2.22e-16 ***	0.3913	0.4059
性別(1)	3.6272	0.1076	33.7186	< 2.22e-16 ***	3.4163	3.838

I：依變數為壓差，模式包含常數項

II：顯著性代碼： ‘***’ :< 0.001, ‘**’ :< 0.01, ‘*’ :< 0.05, ‘#’ :< 0.1

在 1 的模式中加入性別及年齡*性別交互作用項，仍使用迴歸分析

R-web 分析步驟：

分析方法→迴歸模式→迴歸分析

→步驟一：資料匯入(選取個人資料檔 CVD_ALL)

→步驟二：參數設定(依變數：壓差、自變數：年齡、性別)

→進階選項(點選年齡、性別進入交互作用項)

→開始分析

分析結果

迴歸係數估計 I：

係數 coefficient	估計值 estimation	標準差 std. err.	t 檢定統計量 t-statistic	p 值 ^{II} p-value	參數的 95% 信賴區間 95% C.I. for estimations	
					下界 lower	上界 upper
(截距項)	22.3875	0.2316	96.6665	< 2.22e-16 ***	21.9336	22.8414
年齡	0.4589	0.0048	94.7622	< 2.22e-16 ***	0.4494	0.4684
性別(1)	10.6583	0.3754	28.388	< 2.22e-16 ***	9.9224	11.3942
年齡 * 性別(1)	-0.1485	0.0076	-19.5414	< 2.22e-16 ***	-0.1634	-0.1336

I：依變數為壓差，模式包含常數項

II：顯著性代碼： ‘***’ :<0.001, ‘**’ :<0.01, ‘*’ :<0.05, ‘#’ :<0.1

結論：

1. 控制性別後，年齡係數估計值為 0.40 ($p < 2.22 \times 10^{-16}$)，在顯著水準 0.05 下，年齡是為顯著影響壓差的因子
2. 未調整性別的年齡係數估計值為 0.41 ($p < 2.22 \times 10^{-16}$)，與調整性別後的估計值相差不大，性別並沒有對年齡與壓差間的關係造成明顯的干擾
3. 檢定年齡和性別之係數估計值是否為 0 的 $p < 2.22 \times 10^{-16}$ ，在顯著水準 0.05 下，年齡和性別對壓差的影響有統計上顯著的交互作用

3. CVD_All 資料中有追蹤時間的紀錄，請使用這個資料及卜瓦松模型估計 CVD 的每年發生率為何？

答：

計算 CVD 每年的發生率，因每筆資料中有該個體患病與否及其追蹤時間，為了讓測量發生次數的時間長度一致，利用 R-web 中的新變數建立產生新變數年發生次數，再以摘要統計計算年發生率

R-web 分析步驟：

分析方法→新變數建立

→步驟一：資料匯入(選取個人資料檔 CVD_ALL)

→步驟二：參數設定(年發生次數=心血管疾病/追蹤時間)

→開始分析

R-web 分析步驟： 分析方法→摘要統計

→步驟一：資料匯入(選取個人資料檔 CVD_ALL)

→步驟二：參數設定(變數：年發生次數)

→進階選項(不須設定)

→開始分析

分析結果

- 資料名稱：CVD_All
- 變數名稱：年發生次數
- 計算時間：0.278 秒
- 摘要統計表¹：

變數名稱 Variable	年發生次數
平均數 Mean	0.0231

- 1：摘要統計皆不包含遺失值

4. 請檢定壓差 (壓差=收縮壓減舒張壓) 是否為 CVD 發生的風險因子? 年齡 (或性別) 是否為干擾因子? 年齡 (或性別) 和壓差是否存在交互作用?

答 :

CVD 發生與否為一類別變數, 故檢定壓差為 CVD 的風險因子可使用獨立兩樣本檢定或是邏輯斯迴歸, 因為後續加入年齡及交互作用項, 故此處選擇邏輯斯迴歸。

此題分為五個部份處理

- (1) 檢定壓差是否為 CVD 發生的風險因子, 將壓差放入模式中分析
- (2) 判斷年齡是否為干擾因子
- (3) 判斷性別是否為干擾因子
- (4) 檢定年齡與壓差的交互作用, 將壓差及年齡變數放入模式中, 並設定交互作用項
- (5) 檢定性別與壓差的交互作用, 將壓差及性別變數放入模式中, 並設定交互作用項

(1) 檢定壓差是否為 CVD 發生的風險因子，將壓差放入模式中分析 R-web 分析步驟：

分析方法→迴歸模式→邏輯斯迴歸分析

→步驟一：資料匯入(選取個人資料檔 CVD_ALL)

→步驟二：參數設定(依變數：心血管疾病，自變數：壓差)

→進階選項(不須設定)

→開始分析

分析結果

- 分析方法：邏輯斯迴歸分析
- 資料名稱：CVD_All
- 依變數名稱：心血管疾病
- 自變數名稱：壓差
- 計算時間：6.892 秒
- 模式係數估計^I：

係數 coefficient	估計值 estimation	標準差 std. err.	華德檢定 統計量 Wald- statistic	p 值 ^{II} p- value	參數的 95% 信賴區間 95% C.I. for estimations		估計值的指數(勝 算比) Exp(coef.)(odds Ratio) (僅針對類別變 數)	Exp(coef.) 的 95% 信賴 區間	
					下界 lower	上界 upper		下界 lower	上界 upper
(截距項)	-3.6208	0.0446	6585.1168	< 1e- 04 ***	- 3.7084	- 3.5335	---	---	---
壓差	0.0281	8e-04	1105.487	< 1e- 04 ***	0.0265	0.0298	---	---	---

I：依變數為心血管疾病

II：顯著性代碼： '***' :< 0.001, '**' :< 0.01, '*' :< 0.05, '#' :< 0.1

(2) 判斷年齡是否為干擾因子，將壓差及年齡變數放入模式中分析 R-web
分析步驟：

分析方法→迴歸模式→邏輯斯迴歸分析

→步驟一：資料匯入(選取個人資料檔 CVD_ALL)

→步驟二：參數設定(依變數：心血管疾病，自變數：壓差、年齡)

→進階選項(不須設定)

→開始分析

分析結果

- 分析方法：邏輯斯迴歸分析
- 資料名稱：CVD_All
- 依變數名稱：心血管疾病
- 自變數名稱：壓差, 年齡
- 計算時間：9.337 秒
- 模式係數估計¹：

係數 coefficient	估計值 estimation	標準差 std. err.	華德檢定 統計量 Wald- statistic	p 值 II p-value	參數的 95% 信賴區間 95% C.I. for estimations		估計值的指數(勝算 比) Exp(coef.)(odds Ratio) (僅針對類別變數)	Exp(coef.) 的 95% 信賴區間	
					下界 lower	上界 upper		下界 lower	上界 upper
(截距項)	-6.0556	0.0674	8072.0631	< 1e-04 ***	- 6.1883	-5.9241	---	---	---
壓差	0.0024	0.001	6.1678	0.013 *	5e-04	0.0043	---	---	---
年齡	0.07	0.0012	3373.618	< 1e-04 ***	0.0676	0.0724	---	---	---

未調整年齡之壓差 OR	$\exp(0.0281)=1.028 (p<10^{-4})$
調整年齡之壓差 OR	$\exp(0.0024)=1.002 (p=0.013)$

(3)判斷性別是否為干擾因子，將性別及年齡變數放入模式中分析 R-web
分析步驟：

分析方法→迴歸模式→邏輯斯迴歸分析

→步驟一：資料匯入(選取個人資料檔 CVD_ALL)

→步驟二：參數設定(依變數：心血管疾病，自變數：壓差、性別)

→進階選項(不須設定)

→開始分析

分析結果

- 分析方法：邏輯斯迴歸分析
- 資料名稱：CVD_All
- 依變數名稱：心血管疾病
- 自變數名稱：性別, 壓差
- 計算時間：9.325 秒
- 模式係數估計^I：

係數 coefficient	估計值 estimation	標準差 std. err.	華德檢定 統計量 Wald- statistic	p 值 ^{II} p-value	參數的 95% 信賴 區間 95% C.I. for estimations		估計值的指數(勝 算比) Exp(coef.)(odds Ratio) (僅針對類別變 數)	Exp(coef.) 的 95% 信賴區間	
					下界 lower	上界 upper		下界 lower	上界 upper
(截距項)	-3.6202	0.045	6465.9211	< 1e-04 ***	-3.7087	-3.5322	---	---	---
壓差	0.0281	9e-04	1089.4314	< 1e-04 ***	0.0265	0.0298	---	---	---
性別(1)	-0.0024	0.0285	0.0073	0.9319	-0.0585	0.0534	0.9976	0.9432	1.0549

I：依變數為心血管疾病

II：顯著性代碼： '***' :< 0.001, '**' :< 0.01, '*' :< 0.05, '#' :< 0.1

未調整性別之壓差 OR	$\exp(0.0281)=1.028 (p<10^{-4})$
調整性別之壓差 OR	$\exp(0.0281)=1.028 (p<10^{-4})$

(4) 檢定年齡與壓差的交互作用，將壓差及年齡變數放入模式中，並設定交互作用項 R-web 分析步驟：

分析方法→迴歸模式→邏輯斯迴歸分析

- 步驟一：資料匯入(選取個人資料檔 CVD_ALL)
- 步驟二：參數設定(依變數：心血管疾病，自變數：壓差、年齡)
- 進階選項(選入壓差*年齡的交互作用)
- 開始分析

分析結果

- 分析方法：邏輯斯迴歸分析
- 資料名稱：CVD_All
- 依變數名稱：心血管疾病
- 自變數名稱：壓差, 年齡, 壓差 * 年齡
- 計算時間：12.132 秒
- 模式係數估計^I：

係數 coefficient	估計值 estimation	標準差 std. err.	華德檢定統 計量 Wald- statistic	p 值 ^{II} p-value	參數的 95% 信賴區 間 95% C.I. for estimations		估計值的指數(勝 算比) Exp(coef.)(odd s Ratio) (僅針對類別變 數)	Exp(coef.) 的 95% 信賴區間	
					下界 lower	上界 upper		下界 lower	上界 upper
(截距項)	-6.2946	0.2184	830.4348	< 1e-04 ***	-6.7241	-5.8679	---	---	---
壓差	0.0076	0.0046	2.7372	0.098 .	-0.0014	0.0167	---	---	---
年齡	0.074	0.0037	404.4388	< 1e-04 ***	0.0668	0.0812	---	---	---
壓差 * 年齡	-1e-04	1e-04	1.3289	0.249	-2e-04	1e-04	0.9999	0.9998	1.0001

I：依變數為心血管疾病

II：顯著性代碼： '***' : < 0.001, '**' : < 0.01, '*' : < 0.05, '# ' : < 0.1

(5) 檢定性別與壓差的交互作用，將壓差及性別變數放入模式中，並設定交互作用項

R-web 分析步驟：

分析方法→迴歸模式→邏輯斯迴歸分析

→步驟一：資料匯入(選取個人資料檔 CVD_ALL)

→步驟二：參數設定(依變數：心血管疾病，自變數：壓差、性別)

→進階選項(選入壓差*性別的交互作用)

→開始分析

分析結果

- 分析方法：邏輯斯迴歸分析
- 資料名稱：CVD_All
- 依變數名稱：心血管疾病
- 自變數名稱：性別, 壓差, 性別 * 壓差
- 計算時間：12.047 秒
- 模式係數估計¹：

係數 coefficient	估計值 estimation	標準差 std. err.	華德檢定統 計量 Wald- statistic	p 值 ^{II} p-value	參數的 95% 信賴 區間 95% C.I. for estimations		估計值的指數(勝 算比) Exp(coef.)(odds Ratio) (僅針對類別變 數)	Exp(coef.) 的 95% 信賴區間	
					下界 lower	上界 upper		下界 lower	上界 upper
(截距項)	-3.6513	0.0554	4350.0078	< 1e-04 ***	-3.7601	-3.543	---	---	---
壓差	0.0288	0.0011	712.3608	< 1e-04 ***	0.0267	0.0309	---	---	---
性別(1)	0.0844	0.0943	0.8025	0.3703	-0.1005	0.2691	1.0881	0.9044	1.3088
壓差 * 性別(1)	-0.0017	0.0018	0.934	0.3338	-0.0052	0.0017	0.9983	0.9948	1.0017

I：依變數為心血管疾病

II：顯著性代碼： '***' :< 0.001, '**' :< 0.01, '*' :< 0.05, '#' :< 0.1

小結：

(1)檢定壓差是否為 CVD 發生的風險因子，將壓差放入模式中分析；模式中依變數為心血管疾病，自變數為壓差，在邏輯斯迴歸中壓差的參數估計值為 0.0281，P 值 $<1e-4$ ，顯示壓差為顯著的風險因子。

(2)

未調整年齡之壓差 OR	$\exp(0.0281)=1.028$ ($p < 10^{-4}$)
調整年齡之壓差 OR	$\exp(0.0024)=1.002$ ($p=0.013$)

比較未調整年齡與調整年齡後之 OR 估計值(1.028;1.002),OR 值估計有些微差異，年齡有造成些微干擾

(3)

未調整性別之壓差 OR	$\exp(0.0281)=1.028$ ($p < 10^{-4}$)
調整性別之壓差 OR	$\exp(0.0281)=1.028$ ($p < 10^{-4}$)

比較未調整性別與調整性別後之 OR 估計值(1.028;1.028),OR 值估計並無差異，性別沒有造成干擾。

(4)檢定年齡與壓差的交互作用，將壓差及年齡變數放入模式中，並設定壓差與年齡的交互作用項；模式中壓差與年齡交互作用項的參數估計值所對應的 P 值為 $0.249 > 0.05$ ，交互作用不顯著，顯示此模式中不需加入壓差與年齡的交互作用項；年齡和壓差對心血管疾病發生的影響不存在交互作用。

(5)檢定性別與壓差的交互作用，將壓差及性別變數放入模式中，並設定壓差與性別的交互作用項；模式中壓差與性別交互作用項的參數估計值所對應的 P 值為 $0.3338 > 0.05$ ，交互作用不顯著，顯示性別和壓差對心血管疾病發生的影響不存在交互作用。

5. 接續第 4 個問題，針對性別做分層分析，並比較 4 和 5 的結論？

答：

(1) 要將資料依性別做分層分析，此處必須先對資料進行處理，進行資料篩選，區分為男性與女性兩個資料檔

R-web 分析步驟：

資料處理→資料篩選→

→步驟一：資料匯入(選取個人資料檔 CVD_ALL)

→步驟二：參數設定(篩選條件：性別=1，另存新檔 CVD_ALL_M)

→開始處理

相同的方式處理性別為女性的資料，另存新檔 CVD_ALL_F

(2) 個別對資料進行邏輯斯迴歸分析，CVD 發生與否為依變數，壓差為自變數。

R-web 分析步驟：

分析方法→迴歸模式→邏輯斯迴歸分析

→步驟一：資料匯入(選取個人資料檔 CVD_ALL)

→步驟二：參數設定(依變數：心血管疾病，自變數：壓差)

→進階選項(不須設定)

→開始分析

再以相同的方式處理 CVD_ALL_F

分析結果

- 分析方法：邏輯斯迴歸分析
- 資料名稱：CVD_ALL_M
- 依變數名稱：心血管疾病
- 自變數名稱：壓差
- 計算時間：2.604 秒
- 模式係數估計 ^I：

係數 coefficient	估計值 estimation	標準差 std. err.	華德檢定統計量 Wald-statistic	p 值 ^{II} p-value	參數的 95% 信賴區間 95% C.I. for estimations		估計值的指數 (勝算比) Exp(coef.)(odds Ratio) (僅針對類別變數)	Exp(coef.) 的 95% 信賴區間	
					下界 lower	上界 upper		下界 lower	上界 upper
(截距項)	-3.5668	0.0763	2185.6755	< 1e-04 ***	-3.7167	-3.4177	---	---	---
壓差	0.0271	0.0014	378.0515	< 1e-04 ***	0.0243	0.0298	---	---	---

I：依變數為心血管疾病

II：顯著性代碼： '***' : < 0.001, '**' : < 0.01, '*' : < 0.05, '#' : < 0.1

- 分析方法：邏輯斯迴歸分析
- 資料名稱：CVD_ALL_F
- 依變數名稱：心血管疾病
- 自變數名稱：壓差
- 計算時間：4.183 秒
- 模式係數估計¹：

係數 coefficient	估計值 estimation	標準差 std. err.	華德檢定 統計量 Wald- statistic	p 值 ¹¹ p-value	參數的 95% 信賴區間 95% C.I. for estimations		估計值的指數(勝 算比) Exp(coef.)(odds Ratio) (僅針對類別變 數)	Exp(coef.) 的 95% 信賴區間	
					下界 lower	上界 upper		下界 lower	上界 upper
(截距項)	-3.6513	0.0554	4350.0078	< 1e-04 ***	-3.7601	-3.543	---	---	---
壓差	0.0288	0.0011	712.3608	< 1e-04 ***	0.0267	0.0309	---	---	---

結論:

	男性	女性	未調整性別
壓差之 OR 估計值	1.027 ($p < 10^{-4}$)	1.029 ($p < 10^{-4}$)	1.028 ($p < 10^{-4}$)

1. 男性中的分析，壓差之 OR 估計值為 1.027 ($p < 10^{-4}$)；女性中的分析，壓差之 OR 估計值為 1.029 ($p < 10^{-4}$)。男、女分層分析的結果相似，性別與壓差間對 CVD 發生的影響沒有交互作用
2. 未調整性別與性別分層分析之壓差 OR 估計值相差不大，性別不是干擾因子
3. 與第 4 題的分析結果與結論一致