

以下資料皆為模擬生成資料，供分析使用，不一定具有臨床意義。

1. lung_cancer_study

此資料檔運用於第一章及第十二章。資料是由美國 4 個研究肺癌的醫學中心整合出來的樣本資料中分析得來的。變數說明如下：

SITE：4 個研究中心

AGE：年齡(歲)

GENDER：性別(Female；Male)

gender：性別(0, Female；1, Male)

CHEMO：化療與否(0, No；1, Yes)

RT：放射治療(Yes；No；NA:未知)

rt：放射治療(0, 沒有；1, 有；NA, 未知)

Vital.Status：研究結束時存活狀態(Alive；Dead)

vital_status：研究結束時存活狀態(0, 存活；1, 死亡)

FIRST_PROGRESSION_OR_RELAPSE：術後第一次復發與否(Yes；No)

MONTHS_TO_FIRST_PROGRESSION：手術後到第一次復發時間(單位:月)

SMOKING：抽菸狀態(Never smoked, Smoked in the past, Currently smoking)

N_STAGE：癌症 N 分期(N0, 第 0 期；N1, 第一期；N1 or N2, 第一或二期)

T_STAGE：癌症 T 分期(T1, 第一期；T2 or T3, 第二或三期；T2, T3 or T4, 第二、三或四期)

SURVIVAL_MONTHS：存活時間(單位:月)

2. PPHN

在第四章及第六章中，有使用到此資料。此為新生兒持續性肺動脈高壓（persistent pulmonary hypertension of newborn, PPHN）之相關研究。此疾病多發生於足月兒或過期產兒，出現臨床上低血氧等之症狀，發生原因可能和子宮或生產時之因素有關。由於持續性肺動脈高壓之死亡率約 1/3，若能找出和死亡相關之危險因子，即可預防死亡之發生。在新生兒加護病房，新生兒持續性肺動脈高壓則納入欲收集之樣本，共有 131 名新生兒納入。變數說明如下：

性別：0,女生；1,男生

胎便吸入：出生時嬰兒是否吸入胎便(0,無；1,有)

早期破水：出生時母親羊水是否有早期破水情形(0,無；1,有)

剖婦產：出生時母親是否為剖婦產(0,否；1,是)

週數：母親之懷孕週數(單位：週)

體重：嬰兒之出生體重(單位：公克)

1 分鐘 apgar 分數：嬰兒之 1 分鐘 apgar 分數（Apgar 分數是針對出生新生兒健康狀況快速評核方法，分數愈高代表狀況愈佳，0-10 分。）

5 分鐘 apgar 分數：嬰兒之 5 分鐘 apgar 分數

PH：嬰兒之酸鹼值

AaDO₂：嬰兒之肺泡氣-動脈血氧分壓差，是判斷肺換氣功能正常與否的一個依據，是反映預后的一項重要指標（單位：kPa）

死亡：嬰兒在觀察過程中死亡與否(0,否；1,是)

3. survival

第五章中使用，此資料是關於補充抗壞血酸的癌症治療的研究，資料出自於1978年發表在 Proceedings of the National Academy of Science USA, 75, 4538-4542 論文，研究目的在重新評估該療法是否延長人類癌症的存活時間。樣本為患有胃、支氣管、結腸、卵巢及乳房的晚期癌症患者採用抗壞血酸，該研究的目的是想了解是否感染癌症的器官不同會影響其存活時間，有興趣的 endpoint 為存活天數，共有 64 名癌症患者。變數說明如下：

Survival：存活時間，單位：天

Organ：癌症之種類，Stomach(胃癌)、Bronchus(支氣管癌)、Colon(結腸癌)、Ovary(卵巢)、Breast(乳癌)

4. CVD_All

此資料在第七章及第八章使用。為基隆地區社區疾病篩檢研究資料，以了解可能影響心血管疾病各個變項之間的關係。變數說明如下：

心血管疾病：是否有心血管疾病(0,否；1,是)

年齡

性別(0,女生；1,男生)

追蹤時間(單位：年)

腰圍(單位：公分)

收縮壓(單位：mmHg)

舒張壓(單位：mmHg)

空腹血糖(單位：mg/dl)

高密度脂蛋白(單位：mg/dl)

三酸甘油酯(單位：mg/dl)

檳榔：是否有嚼檳榔習慣(0,否；1,是)

飲酒：是否有飲酒習慣(0,否；1,是)

家族病史：有無家族病史(0,無；1,有)

抽煙：是否有使用習慣(0,否；1,是)

抽菸量：每日抽菸量(0,每日菸量 0；1,每日菸量 1 包；2,每日菸量 2 包；3,每日菸量超過 2 包)

5. Danmark lung cancer

在第九章中，以一筆有關丹麥於 1968-1971 針對四個城市所作關於罹患肺癌的

資料為例，研究目的想知道是否不同的年齡層及地區會影響肺癌的發生率。研究中調查四個城市在六個年齡層的新發肺癌案例，其中包含每個城市中各年齡層的居民人數，通常我們稱此為各分群中涉險(risk exposure)人數(或人-年)，分群整理後共 24 筆。變數為：城市別、年齡層、居民數、案例數。

6. CVD_sampling5000

此資料為第十章之分析資料，為 CVD_All 資料中，隨機抽樣 5000 人之結果。