

調查方法與資料處理期末報告書

台灣地區不孕症患者壓力
與憂鬱傾向研究

統計碩士學位學程
R08H41001 廖容嫻

目錄

| | | |
|----------|---|----------|
| 1 | 緒論 | 1 |
| 1.1 | 研究問題與目的 | 1 |
| 1.2 | 研究架構 | 1 |
| 2 | 參考文獻 | 3 |
| 3 | 研究方法 | 4 |
| 3.1 | 資料來源 | 4 |
| 3.2 | 使用統計模型 | 4 |
| 3.2.1 | 多元羅吉斯迴歸模型 (Multinomial Logistic Regression) | 4 |
| 3.2.2 | 多元線性迴歸模型 (Multinomial Linear Regression) | 5 |
| 4 | 研究變數介紹與處理 | 6 |
| 4.1 | 研究變數 | 6 |
| 4.1.1 | 反應變數 | 6 |
| 4.1.2 | 解釋變數 | 7 |
| 4.2 | 探索性資料分析 | 7 |
| 4.3 | 遺失值差補 | 8 |
| 5 | 分析與研究 | 9 |
| 5.1 | 建立多元羅吉斯迴歸模型 | 9 |

| | | |
|----------|---------------------------|-----------|
| 5.2 | 建立多元線性迴歸模型 | 13 |
| 5.3 | 解釋自覺健康狀況的係數異常現象 | 15 |
| 6 | 結論 | 18 |

圖目錄

| | | |
|--------|---------------------------------|----|
| 圖 1.1 | 研究架構 | 2 |
| 圖 4.1 | 焦慮與憂鬱程度分類方式 | 6 |
| 圖 5.1 | 自覺身體狀況為解釋變數的簡單線性迴歸模型 | 15 |
| 圖 5.2 | 自覺身體狀況為解釋變數的簡單線性迴歸模型-刪去離群 值後 | 16 |
| 圖 6.1 | 焦慮與憂鬱-類別型 | 19 |
| 圖 6.2 | 焦慮與憂鬱-連續型 | 19 |
| 圖 6.3 | 年齡 | 20 |
| 圖 6.4 | 教育程度 | 20 |
| 圖 6.5 | 工作類別 | 20 |
| 圖 6.6 | 平均家庭月收入 | 21 |
| 圖 6.7 | 是否有小孩 | 21 |
| 圖 6.8 | 生育觀念 | 21 |
| 圖 6.9 | 自覺健康狀況 | 22 |
| 圖 6.10 | 近一年身體狀況 | 22 |
| 圖 6.11 | 近一年心理狀況 | 22 |
| 圖 6.12 | 是否有不良習慣 | 23 |
| 圖 6.13 | 是否有運動習慣 | 23 |

| | | |
|--------|-----------------------|----|
| 圖 6.14 | 目前受孕技術 | 23 |
| 圖 6.15 | 不孕年數 | 24 |
| 圖 6.16 | 接受人工生殖治療總次數 | 24 |
| 圖 6.17 | 接受試管嬰兒總次數 | 24 |

表目錄

| | | |
|-------|---|----|
| 表 4.1 | 本文使用的解釋變數 | 7 |
| 表 4.2 | 有遺失值的解釋變數列表 | 8 |
| 表 5.1 | 多元羅吉斯迴歸模型-放入全部的解釋變數（類別 0 對類別 1） | 10 |
| 表 5.2 | 多元羅吉斯迴歸模型-放入全部的解釋變數（類別 0 對類別 2） | 11 |
| 表 5.3 | 多元羅吉斯迴歸模型-放入重要解釋變數（類別 0 對類別 1） | 12 |
| 表 5.4 | 多元羅吉斯迴歸模型-放入重要解釋變數（類別 0 對類別 2） | 12 |
| 表 5.5 | 多元線性迴歸模型-放入全部的解釋變數 | 13 |
| 表 5.6 | 多元線性迴歸模型-使用項後消去法後得到的模型 | 14 |
| 表 5.7 | 自覺身體狀況為解釋變數的簡單線性迴歸模型 | 16 |
| 表 5.8 | 自覺身體狀況為解釋變數的簡單線性迴歸模型-刪去離群值後 | 17 |

第 1 章

緒論

1.1 研究問題與目的

本文的研究問題為，希望透過此份研究，可以發現影響因不孕問題前往不孕門診的女性得到焦慮或憂鬱症狀的原因。而研究目的則是希望透過了解其可能罹患焦慮或憂鬱症狀的因素，及早給予關懷與提供醫療資源，降低得到相關疾病的可能性。

1.2 研究架構

圖 1.1 為本篇研究的架構。首先會對變數進行處理，之後會針對同樣的議題，使用兩種不同的模型進行分析，並比較不同模型下所做出來的結果有何不同。

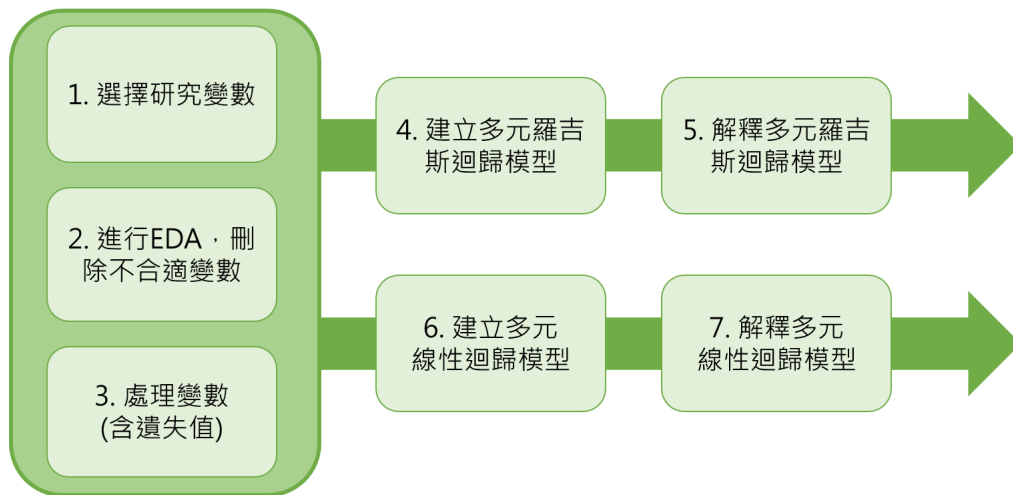


圖 1.1: 研究架構

第 2 章

參考文獻

由於本篇的反應變數為焦慮與憂鬱的程度，因此參考了設計「醫院焦慮與憂鬱量表」的 Zigmond & Snaith（1983）所做的建議，將得到的焦慮與憂鬱總分轉換成以三個類別的方式呈現。而本文也沿用了這樣的作法，並且將兩個面向的變數整合進行分析，詳細的做法請參考本文 4.1.1。

第 3 章

研究方法

3.1 資料來源

本文使用了學術調查研究資料庫（Survey Research Data Archive, SRDA）中的計畫所使用的資料，計畫名稱為：「台灣地區不孕症患者壓力、憂鬱傾向與精神疾病病發之結構方程模式、存活分析及資料探勘」，由中國醫藥大學醫務管理學系的王中儀教授於 2001-2013 年執行。而本研究所使用的資料是該計畫中的一份問卷的調查結果，此問卷為北中南三區共三家醫學中心或區域醫院不孕症門診患者（人工與非人工生殖）400 位，以結構式問卷進行橫斷調查，其中假設每位填寫問卷的受試者為隨機抽取。

3.2 使用統計模型

3.2.1 多元羅吉斯迴歸模型（Multinomial Logistic Regression）

多元羅吉斯迴歸模型適用於反應變數擁有 3 個以上的類別時，符合本文所設定的反應變數 ad（類別 0、類別 1 及類別 2）。其模型設定如下：

$$\text{logit}(Y) = \log\left(\frac{P(Y=1)}{P(Y=0)}\right) = \beta_0 + \beta_1 x_1 i + \beta_2 x_2 i + \dots + \beta_k x_k i + \varepsilon_i, i = 1, 2, \dots, k$$

$$\text{logit}(Y) = \log\left(\frac{P(Y=2)}{P(Y=0)}\right) = \beta_0 + \beta_1 x_1 i + \beta_2 x_2 i + \dots + \beta_k x_k i + \varepsilon_i, i = 1, 2, \dots, k$$

上式代表將類別 0 設為 reference group，分別與類別 1 和類別 2 進行比較。

3.2.2 多元線性迴歸模型 (Multinomial Linear Regression)

而多元線性迴歸模型則較常見，其反應變數為連續型態，且解釋變數均以一次項的方式呈現，不會有高次項出現。其模型設定如下：

$$Y = \beta_0 + \beta_1 x_1 i + \beta_2 x_2 i + \dots + \beta_k x_k i + \varepsilon_i, i = 1, 2, \dots, k$$

第 4 章

研究變數介紹與處理

4.1 研究變數

4.1.1 反應變數

本研究的反應變數來自 Zigmond & Snaith (1983) 的「醫院焦慮與憂鬱量表 (hospital anxiety and depression scale)」得分狀況，其分為憂鬱與焦慮兩個面向。題目總共 14 題，每題有 4 個選項代表 0-3 分，其中 7 題測量是否有焦慮症狀，另外 7 題則是測量是否有憂鬱症狀。因此每位受試者均能得到 0-21 分的焦慮分數與 0-21 分的憂鬱分數作為焦慮與憂鬱程度的測量。

而本研究所使用的是分別將焦慮與憂鬱兩個面向再做分類，將單一面向分為三個級別：0-7 分為無焦慮或憂鬱、8-11 分為有焦慮或憂鬱的傾向以及 12-21 分為有焦慮或憂鬱症狀。再依據兩個面向不同程度的組合，分成 0、1 與 2 三個類別，將此變數定義為 ad，其分類方式請參考下圖 4.1。

| 焦慮 \ 憂鬱 | 0-7 | 8-11 | 12-21 |
|---------|-----|------|-------|
| 0-7 | 0 | 1 | 1 |
| 8-11 | 1 | 1 | 2 |
| 12-21 | 1 | 2 | 2 |

圖 4.1: 焦慮與憂鬱程度分類方式

而本文還會使用到兩個面向的分數加總作為另一個模型的解釋變數，將此變數定義為 **ad-sum**。

4.1.2 解釋變數

本研究的解釋變數分為以下兩個部分，為「基本資料」與「不孕症疾病狀況」的相關變數，如下表 4.1 所示。

表 4.1: 本文使用的解釋變數

| 基本資料 | | 不孕症疾病狀況 |
|---------|---------|-------------|
| 年齡 | 生育觀念 | 目前受孕技術 |
| 教育程度 | 自覺健康狀況 | 不孕年數 |
| 職業類別 | 近一年身體狀況 | 接受人工生殖治療總月數 |
| 平均家庭月收入 | 近一年心理狀況 | 接受人工生殖治療總次數 |
| 婚姻年數 | 是否有不良習慣 | 接受試管嬰兒總次數 |
| 是否有小孩 | 是否有運動習慣 | |

4.2 探索性資料分析

探索性資料分析的目的在於，可以事前檢查是否有不合適的變數不應被放入模型中，或是變數過度集中在某一類別，造成分析上可能會產生錯誤，經由探索性資料分析皆能預先得知，並且針對該變數進行適當處理，以減少建立模型的時間成本並增加模型可信度。

經由探索性資料分析，本文對原始資料進行的更動為：不使用性別變數，因為在 400 筆樣本中，男性僅有 24 個樣本，因此將男性的樣本刪除。

而其餘的解釋變數的探索性資料分析結果將會放置於附錄。

4.3 遺失值差補

有遺失值的變數如下列表：

表 4.2: 有遺失值的解釋變數列表

| 變數名稱 | 遺失值個數 | 處理方式 |
|-------------|-------|--------------|
| 平均家庭月收入 | 3 | 依照相似樣本補差值 |
| 近一年身體狀況 | 2 | 依據他題回答推測此題答案 |
| 近一年心理狀況 | 1 | 依據他題回答推測此題答案 |
| 是否有運動習慣 | 1 | 依據他題回答推測此題答案 |
| 不孕年數 | 1 | 刪除有缺漏值的樣本 |
| 接受人工生殖治療總月數 | 143 | 因缺漏過多，刪除該變數 |
| 接受人工生殖治療總次數 | 6 | 刪除有缺漏值的樣本 |
| 接受試管嬰兒總次數 | 3 | 刪除有缺漏值的樣本 |

關於上表 4.2 中的「依據他題回答推測此題答案」，是指因受試者已勾選以上皆非，因此該向缺漏可以直接經由邏輯判斷被推測出來。而在不孕年數、接受人工生殖治療總次數與接受試管嬰兒總次數中會直接「刪除該缺漏值的樣本」的原因是因為這幾向缺漏的樣本均重疊，因此刪除不會對估計有太大的影響，故刪之。

處理完遺失值後，最終會剩下 370 個樣本進行分析。

第 5 章

分析與研究

5.1 建立多元羅吉斯迴歸模型

將保留下來的解釋變數放入多元羅吉斯迴歸模型中，得到的結果如表 5.1 與表 5.2。

表 5.1 呈現的是類別 0 對類別 1 的比較，即相對於未有焦慮或憂鬱症狀者，有焦慮或憂鬱傾向者的特徵為何；同樣的，表 5.2 呈現的是類別 0 對類別 2 的比較，即相對於未有焦慮或憂鬱症狀者，有焦慮或憂鬱症者的特徵為何。而從表 5.1 與表 5.2 中可以找到一些較為重要的變數，將其挑選出來並另外建立多元羅吉斯迴歸模型，結果如 5.3 與表 5.4。

表 5.1: 多元羅吉斯迴歸模型-放入全部的解釋變數（類別 0 對類別 1）

| 變數名稱 | 係數估計值 | 顯著水準 |
|--|------------|------|
| 截距項 | -0.5977400 | |
| 年齡 | 0.2579660 | |
| 教育程度 | 0.0633335 | |
| 職業類別 | -0.0039753 | |
| 平均家庭月收入 | -0.0902069 | |
| 是否有小孩 | 0.5949013 | |
| 生育觀念 | 0.1649432 | |
| 自覺健康狀況 | -0.5450134 | * |
| 婚姻年數 | 0.0178154 | |
| 近一年身體狀況 | 0.1711550 | |
| 近一年心理狀況 | 0.2186345 | ** |
| 是否有不良習慣 | -0.9486644 | |
| 是否有運動習慣 | -0.0974749 | |
| 目前受孕技術 | 0.0386467 | |
| 不孕年數 | -0.0590230 | |
| 接受人工生殖治療總次數 | 0.0254717 | |
| 接受試管嬰兒總次數 | -0.0572066 | |
| McFadden R-squared: 0.0936 | | |
| Likelihood ratio test : chisq = 59.963 (p.value = 0.0019664) | | |

表 5.2: 多元羅吉斯迴歸模型-放入全部的解釋變數（類別 0 對類別 2）

| 變數名稱 | 係數估計值 | 顯著水準 |
|--|------------|------|
| 截距項 | 0.6572147 | |
| 年齡 | 0.3182480 | |
| 教育程度 | -0.3426564 | |
| 職業類別 | -0.3037196 | |
| 平均家庭月收入 | 0.2807946 | * |
| 是否有小孩 | 0.4657360 | |
| 生育觀念 | 0.0774809 | |
| 自覺健康狀況 | -0.6453463 | * |
| 婚姻年數 | -0.0280376 | |
| 近一年身體狀況 | 0.4508694 | |
| 近一年心理狀況 | 0.2417390 | * |
| 是否有不良習慣 | 0.3060692 | |
| 是否有運動習慣 | -0.3966730 | |
| 目前受孕技術 | -0.1243989 | |
| 不孕年數 | 0.0598678 | |
| 接受人工生殖治療總次數 | 0.1624855 | |
| 接受試管嬰兒總次數 | -0.3382146 | * |
| McFadden R-squared: 0.0936 | | |
| Likelihood ratio test : chisq = 59.963 (p.value = 0.0019664) | | |

以下是由經過篩選的變數所建立的多元羅吉斯迴歸模型，其解釋如下：

1. 在 0 對 1 與 0 對 2 兩個模型中，自覺健康狀況的係數均顯著為負，代表自覺健康狀況越好，這些人落入有焦慮或憂鬱傾向與有焦慮或憂鬱症的機率較無焦慮與憂鬱者高。這似乎不符合常理，因此在後面的小節會另外再進行討論。

2. 在 0 對 1 與 0 對 2 兩個模型中，近一年心理狀況的係數均顯著為正，代表認為心理狀況不好者，落入有焦慮或憂鬱傾向與有焦慮或憂鬱症的機率較無焦慮與憂鬱者高，此結果符合我們的期望。
3. 在 0 對 2 的模型中，接受人工生殖治療總次數的係數為正，代表治療總次數越多，落入焦慮或憂鬱症的機率較無焦慮與憂鬱者高，此結果符合我們的期望。

表 5.3: 多元羅吉斯迴歸模型-放入重要解釋變數（類別 0 對類別 1）

| 變數名稱 | 係數估計值 | 顯著水準 |
|---|-----------|------|
| 截距項 | 0.754462 | |
| 自覺健康狀況 | -0.622372 | ** |
| 近一年心理狀況 | 0.193825 | * |
| 接受人工生殖治療總次數 | 0.010809 | |
| McFadden R-squared: 0.058054 | | |
| Likelihood ratio test : chisq = 37.191 (p.value = 1.0613e-05) | | |

表 5.4: 多元羅吉斯迴歸模型-放入重要解釋變數（類別 0 對類別 2）

| 變數名稱 | 係數估計值 | 顯著水準 |
|---|-----------|------|
| 截距項 | 0.689416 | |
| 自覺健康狀況 | -0.847985 | ** |
| 近一年心理狀況 | 0.216462 | * |
| 接受人工生殖治療總次數 | 0.198918 | . |
| McFadden R-squared: 0.058054 | | |
| Likelihood ratio test : chisq = 37.191 (p.value = 1.0613e-05) | | |

5.2 建立多元線性迴歸模型

由於使用 Zigmond & Snaith (1983) 的建議將受試者的精神疾病狀況做分組，可能會消除組內的差距。因此本文決定將反應變數改為原始總得分 ad-sum，觀察解釋變數的解釋能力是否提升。而放入所有解釋變數的多元線性迴歸模型如下表 5.5。

表 5.5: 多元線性迴歸模型-放入全部的解釋變數

| 變數名稱 | 係數估計值 | 顯著水準 |
|---|-----------|------|
| 截距項 | 18.590847 | *** |
| 年齡 | 0.556272 | |
| 教育程度 | -0.796840 | |
| 職業類別 | 0.027380 | |
| 平均家庭月收入 | -0.020347 | |
| 是否有小孩 | 0.960838 | |
| 生育觀念 | 0.235578 | |
| 自覺健康狀況 | -2.162483 | *** |
| 婚姻年數 | -0.006869 | |
| 近一年身體狀況 | 0.603323 | |
| 近一年心理狀況 | 0.539695 | ** |
| 是否有不良習慣 | -0.433010 | |
| 是否有運動習慣 | -0.025879 | |
| 目前受孕技術 | -0.105457 | |
| 不孕年數 | 0.136331 | |
| 接受人工生殖治療總次數 | 0.470932 | * |
| 接受試管嬰兒總次數 | -0.308627 | |
| Adjusted R-squared: 0.1225 | | |
| F-statistic: 3.575 on 20 and 349 DF, p-value: 6.268e-07 | | |

而由於變數太多，因此使用項後消去法後得到以下精簡版的多元線性迴歸

模型，如表 5.6所示。對於多元線性迴歸模型的解釋為：

1. 年齡的係數顯著為正，代表隨著年齡越高越有可能進入為焦慮或憂鬱的狀態。此結果很合理，因為隨著年紀的增長，懷孕的難度會增加，心理上承受的壓力可能更大。
2. 教育程度的係數為負，代表隨著教育程度越低越有可能進入為焦慮或憂鬱的狀態。有可能是因為教育程度較低者，更需要仰賴丈夫及其家庭，若無法生育可能會造成社會上的批評。
3. 自覺健康狀況的係數十分顯著為負，代表隨著自覺健康狀況越好越有可能進入為焦慮或憂鬱的狀態。這點一樣在後面的小節為大家說明。
4. 近一年心理狀況顯著為正，代表近一年心理狀況越差，越有可能進入為焦慮或憂鬱的狀態。此與預想相同。
5. 接受人工生殖治療總次數顯著為正，代表接受人工生殖治療總次數越多，越有可能進入為焦慮或憂鬱的狀態。此也與預想相同。

表 5.6: 多元線性迴歸模型-使用項後消去法後得到的模型

| 變數名稱 | 係數估計值 | 顯著水準 |
|--|---------|------|
| 截距項 | 20.2332 | *** |
| 年齡 | 0.7434 | * |
| 教育程度 | -0.7910 | |
| 自覺健康狀況 | -2.4070 | *** |
| 近一年心理狀況 | 0.5282 | ** |
| 接受人工生殖治療總次數 | 0.4575 | * |
| Adjusted R-squared: 0.1413 | | |
| F-statistic: 11.12 on 6 and 363 DF, p-value: 2.156e-11 | | |

經由多元線性迴歸模型可以發現到，原本在多元羅吉斯迴歸模型中不顯著的變數「年齡」與「教育程度」，都因此變得顯著了（教育程度的 p 值為

0.1191，也在顯著邊緣)。因此發現使用連續型的 Y，在這樣的情況下的確能讓解釋變數解釋更多，顯著的變數與顯著的程度都明顯上升。

5.3 解釋自覺健康狀況的係數異常現象

由於在前面兩個不同的模型中發現，「自覺健康狀況」此一係數的表現十分異常，因此特地將之提出並討論。而本文有以下兩種推測：

第一，下圖 5.1 為一簡單線性迴歸模型，可以發現在自覺健康程度選擇 2（即選擇「健康」）的最上方有兩個點重合為一點，她們的焦慮家憂鬱分數均為 29 分，是此次試驗中的最高分。經由合理推斷，在其他人都均勻選擇每一類別時，因為她們都選了「2（健康）」，加上迴歸模型是十分容易受到離群值影響的，因此造成結果的偏差。

第二，有可能是受試者在填寫問卷時，會下意識地認為詢問健康狀況是指「身體層面的健康狀況」，而非心理層面的。因此會與我們的想樣有落差。為了驗證是否為第一種可能性，故將這兩筆離群的樣本刪除，並且再建立一次模型，察看結果是否會有所不同。

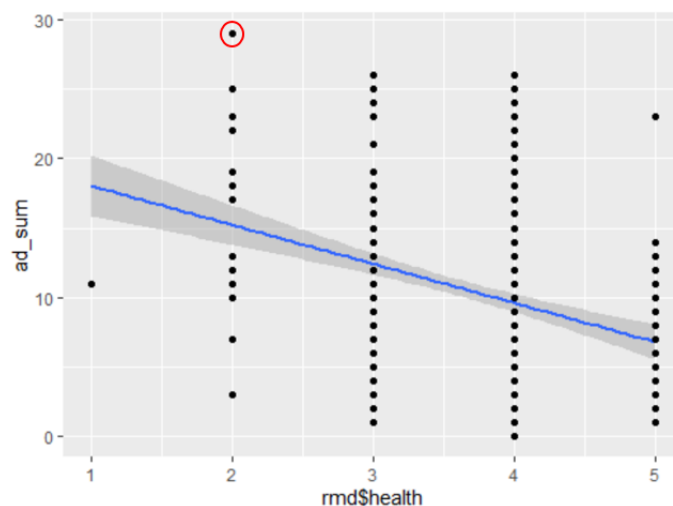


圖 5.1: 自覺身體狀況為解釋變數的簡單線性迴歸模型

表 5.7: 自覺身體狀況為解釋變數的簡單線性迴歸模型

| 變數名稱 | 係數估計值 | 顯著水準 |
|--|---------|------|
| 截距項 | 20.7290 | *** |
| 自覺健康狀況 | -2.7741 | *** |
| Multiple R-squared: 0.1081 | | |
| F-statistic: 44.61 on 1 and 368 DF, p-value: 8.892e-11 | | |

在表 5.8 中可發現，雖然刪去了兩位 ad-sum 為 29 的受試者，但係數依然十分顯著為負，只是係數從-2.72 增加了一點至-2.55。若此實驗可信，則經此可證明，確實自覺健康狀況越差的人，不一定焦慮或憂鬱程度就會很嚴重。或是如剛剛的第二個論點所述，因為我們的認知與受試者有所不同而導致的結果。

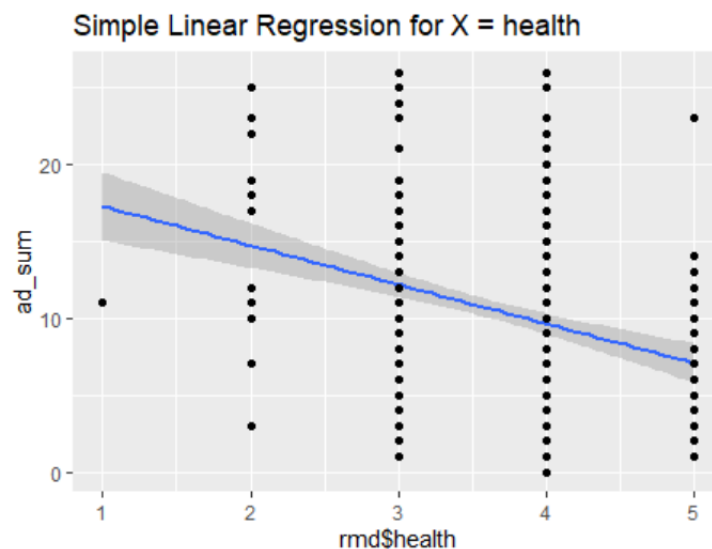


圖 5.2: 自覺身體狀況為解釋變數的簡單線性迴歸模型-刪去離群值後

表 5.8: 自覺身體狀況為解釋變數的簡單線性迴歸模型-刪去離群值後

| 變數名稱 | 係數估計值 | 顯著水準 |
|--|---------|------|
| 截距項 | 19.8243 | *** |
| 自覺健康狀況 | -2.5452 | *** |
| Adjusted R-squared: 0.09053 | | |
| F-statistic: 37.53 on 1 and 366 DF, p-value: 2.322e-09 | | |

第 6 章

結論

經由以上眾多的模型，我們可以發現，在知道不孕後需要注意的女性族群為：

1. 年齡較高
2. 教育程度較低
3. 自覺近一年心理狀況較差者
4. 接受多次人工生殖治療

並且，自覺身體狀況較好者，並不一定就不會有焦慮或憂鬱的症狀，不能因此而鬆懈。

希望經由這份研究，能夠讓醫療機構多注意與觀察不孕婦女的心理狀況，及早的關心與透過醫療資源的支持，相信一定能帶給這樣的族群更多正向的力量。

附錄

附錄為所有反應變數與解釋變數的探索性資料分析。

反應變數

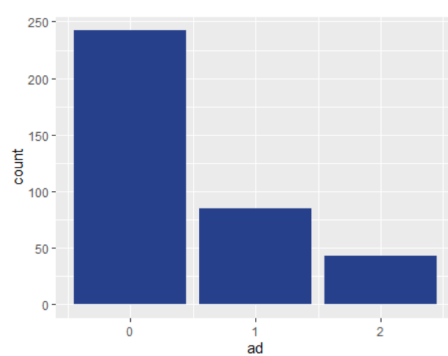


圖 6.1: 焦慮與憂鬱-類別型

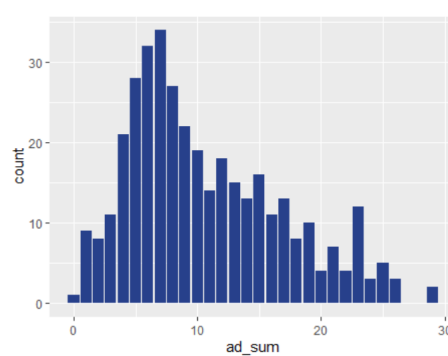


圖 6.2: 焦慮與憂鬱-連續型

解釋變數

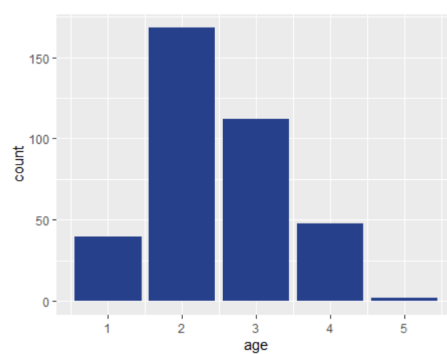


圖 6.3: 年齡

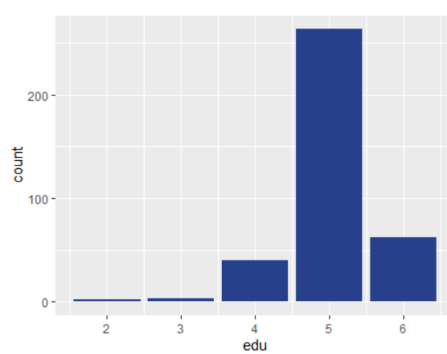


圖 6.4: 教育程度

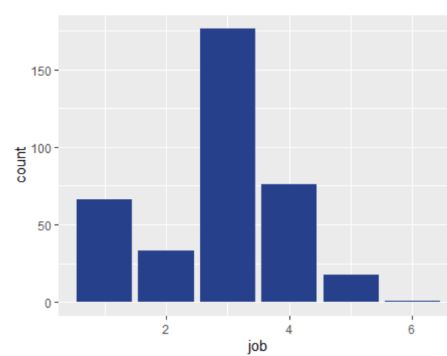


圖 6.5: 工作類別

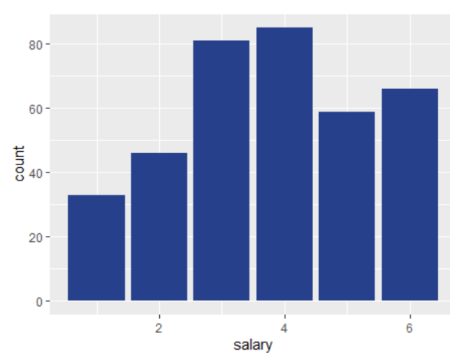


圖 6.6: 平均家庭月收入

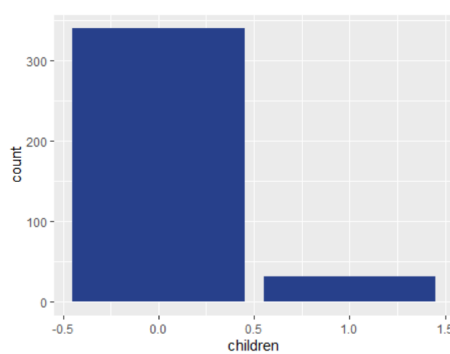


圖 6.7: 是否有小孩

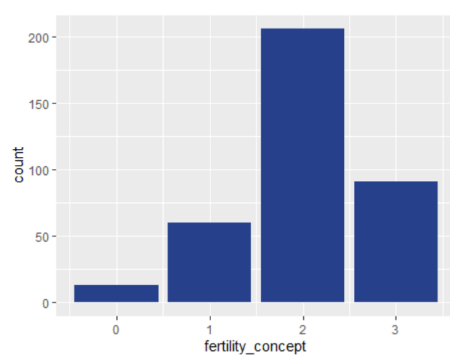


圖 6.8: 生育觀念

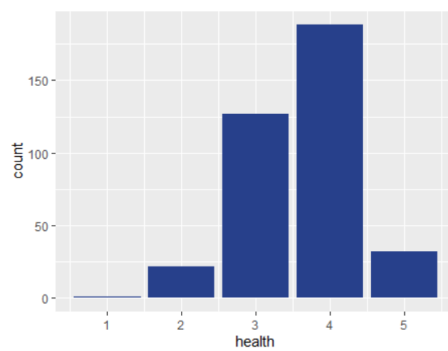


圖 6.9: 自覺健康狀況

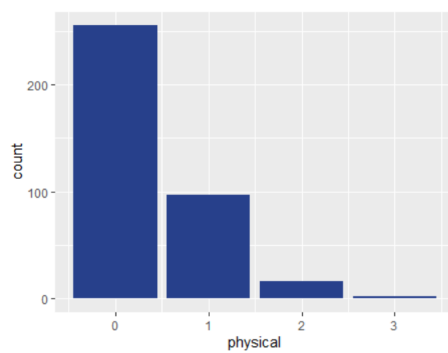


圖 6.10: 近一年身體狀況

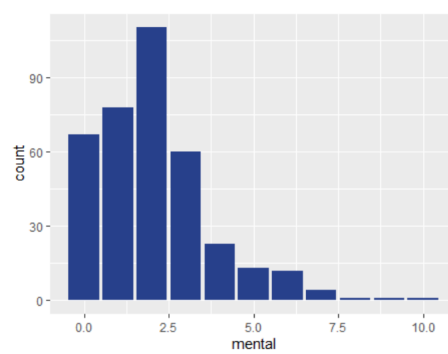


圖 6.11: 近一年心理狀況

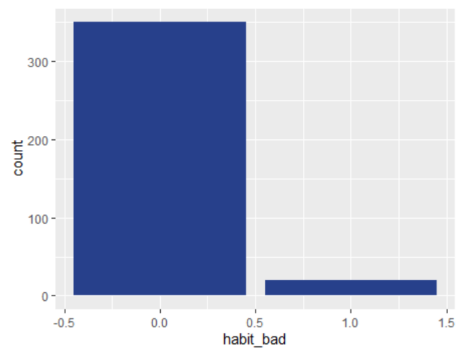


圖 6.12: 是否有不良習慣

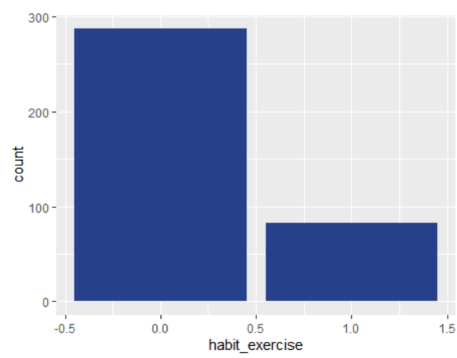


圖 6.13: 是否有運動習慣

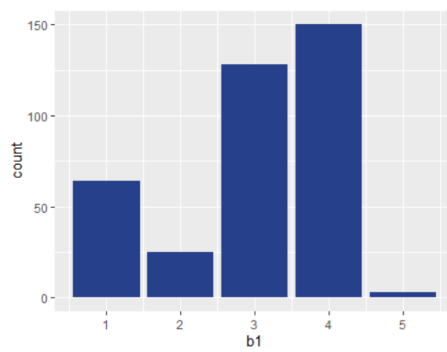


圖 6.14: 目前受孕技術

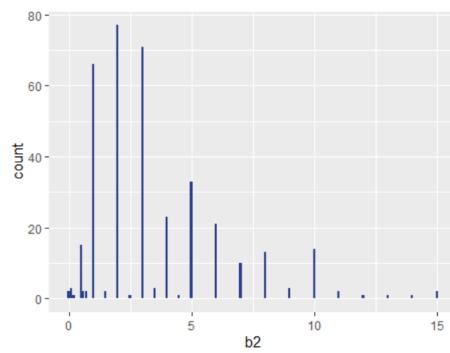


圖 6.15: 不孕年數

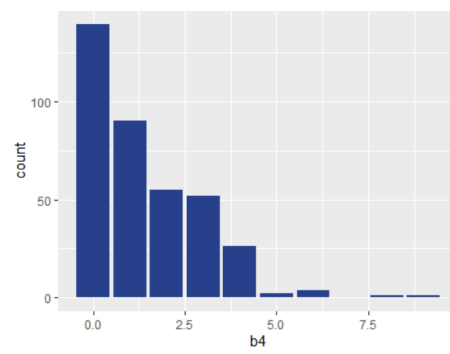


圖 6.16: 接受人工生殖治療總次數

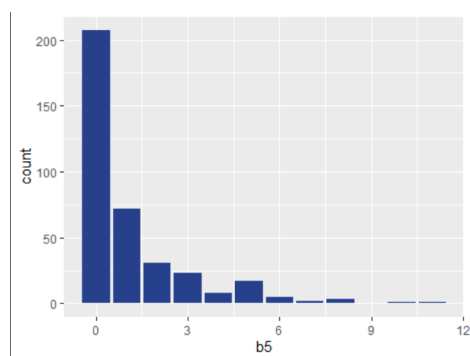


圖 6.17: 接受試管嬰兒總次數