

# CONSULTING IN STATISTICS

## 統計諮詢

指導老師: 王秀瑛 教授

醫療糾紛研究小組: 陳民捷 侯昌儒 高子瑗 溫俞婷 李家琦

### ● 資料背景

#### 全台醫療糾紛愈形嚴重 「醫師在法院上班，律師到醫院上班」

全台醫療糾紛案件，已到十分嚴重地步。衛生署醫事審議委員會醫事鑑定小組統計，國內醫療糾紛鑑定件數，從 1995 年的 197 件，增至 2010 年的 485 件，增加 1.46 倍，甚至 2009 年更高達 546 件。換言之，全台每天至少一名醫師被告。

#### 外科最常被告，其次內科以及婦產科

研究指出，從 2000 年元旦至 2008 年 6 月底，全台共有 312 位醫師被告刑事案件，80 名被判有罪，定罪率 25.6%，其中外科比率 26%最高、內科 21%其次、婦產科 16%第三

#### 狹義的醫療糾紛-----醫療事故

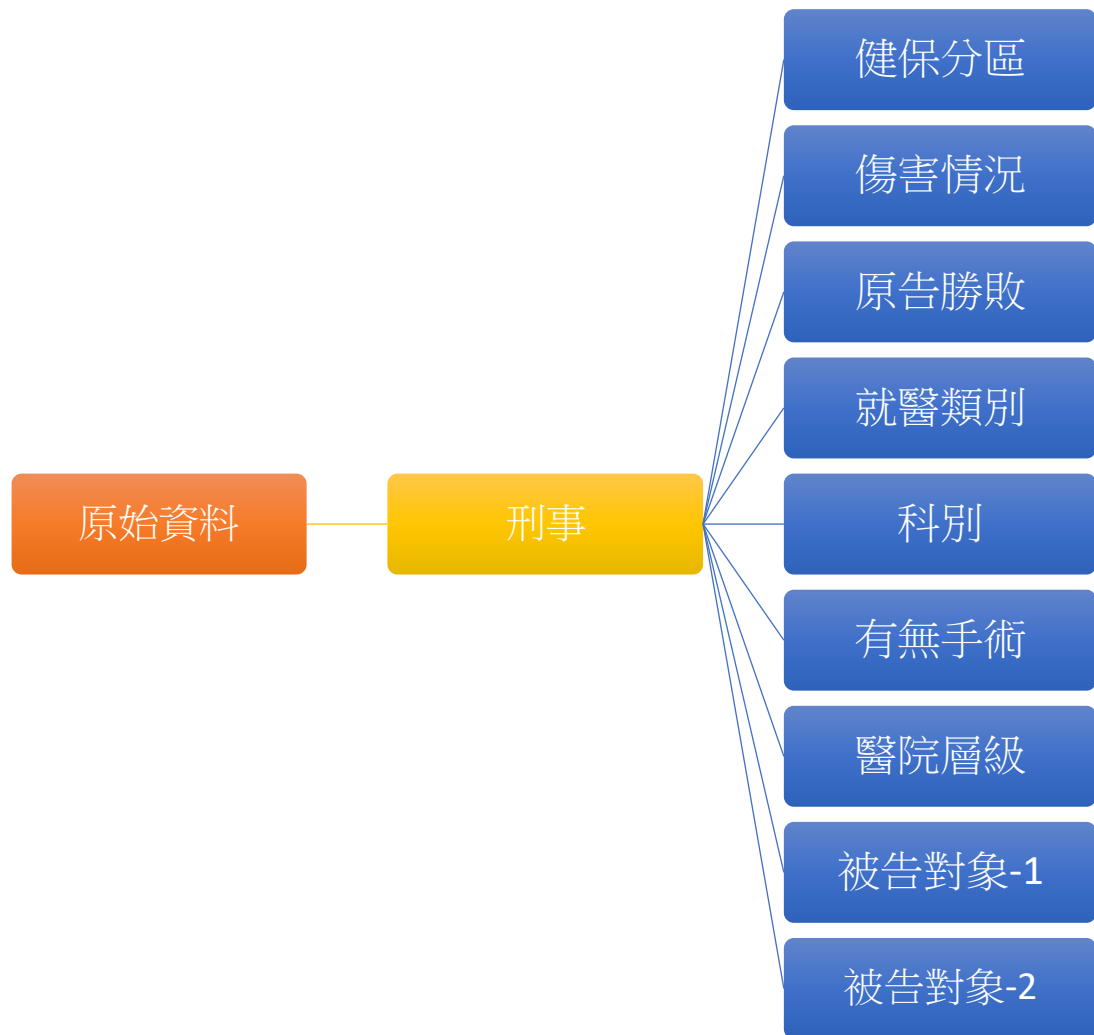
1. 醫療過失
2. 醫療不幸
3. 醫療意外

#### 台灣醫療糾紛傾向告刑事附帶民事

1. 刑法
  - 業務過失致死（公訴）
  - 業務過失重傷害（告訴乃論）
  - 業務過失輕傷害（告訴乃論）
2. 民法
  - 行為人責任
  - 醫療機構責任



- 資料內容分析



- 議題探討

1. 何種**科別**最容易產生**醫療糾紛**
2. 何種情況下會影響**勝訴率** (ex: 某科別死亡率較高，導致勝訴率較高？)



- 統計分析方法
  - 相關性(卡方檢定)
  - Logistics regression
  - Hierarchical linear model

### Logistics regression

此方法為**羅吉斯迴歸** (Logistic regression ) 或是分類評定模型(Logit model )是離散選擇法模型之一，屬於多重變數分析範疇，是社會學、生物統計學、臨床、數量心理學、計量經濟學、市場營銷等統計實證分析的常用方法。

通常羅吉斯迴歸所探討結果的依變數是離散型，特別是其分類只有二類(例如“是與否”、“同意與不同意”、“成功與失敗”)或少數幾類時。

令 $p$ 表示某種事件成功的機率，它受因素 $x$ 的影響，即 $p$ 與 $x$ 之關係如下：

$$P = \frac{e^{f(x)}}{1 + e^{f(x)}}$$

$$\ln \frac{p}{1-p} = f(x) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k$$

### Hierarchical Linear Model (HLM)

此方法為**階層線性模式** (hierarchical linear modeling, 簡稱為 HLM) 或是多層次分析 (multi-level analysis) 是近一、二十年來開始流行的統計方法。

這種統計方法多半使用於教育研究，但後來也漸漸普及到社會學研究與其它領域。

#### ➤ Example

學生的學習成就可能由城市,老師和學生成績三個變量所影響。

#### ✧ 回歸模型

$$\text{學習成就} = \beta_0 + \beta_1 \text{城市} + \beta_2 \text{老師} + \beta_3 \text{成績}$$

但是，事實上城市有可能影響老師素質，而老師可能也會影響其學生成績的優劣。



✧ 階層線性模式

$$level\ 1: 學習成就 = \beta_0 + \beta_1 成績$$

$$level\ 2: \beta_0 = \gamma_{00} + \gamma_{01}W; \beta_1 = \gamma_{10} + \gamma_{11}W$$

這樣也就是學生成績影響了學習成就，但是在不同的 $W$ (老師、城市)有不同的 $\beta_0$ (截距)、 $\beta_1$ (斜率)

➤ 本次專案資料是否合適？

我們認為此方法適合這次專案，因為不同的地區，有不同的醫療資源，不同的醫療資源將影響醫療糾紛種類，

如果可以分層去看每個地區的醫療糾紛，觀察科別和其他變數影響醫療糾紛，會是一個較好的分析。

● 資料轉換(代碼&摘要統計)

健保分局

代碼	分局名稱	轄區範圍						案件總個數
1	台北分局	台北市	台北縣	基隆市	宜蘭縣	金門縣	連江縣	164
2	北區分局	桃園縣	新竹市	新竹縣	苗栗縣			57
3	中區分局	台中市	台中縣	彰化縣	南投縣			104
4	南區分局	雲林縣	嘉義市	嘉義縣	台南市	台南縣		33
5	東區分局	花蓮縣	台東縣					14
6	高屏分局	高雄市	高雄縣	屏東縣	澎湖縣			83

傷害情況

代碼	情況	案件總個數
1	非死亡	199
2	死亡	256



### 原告勝敗

代碼	結果	案件總個數
0	敗訴	246
1	勝訴	121
2	自訴不受理	24
3	公訴不受理	64

### 就醫類別

代碼	類別	案件總個數
1	門診	126
2	急診	44
3	住院	285

### 科別

代碼	科別	案件總個數
1	內科, 中醫科, 兒科	128
2	外科	104
3	急診科, 醫美科	73
4	婦產科	74
5	骨科, 牙科, 其他	76

### 有無手術

代碼	變項	案件總個數
0	無	186
1	有	204
2	侵入性檢查或治療	36
3	生產	29



### 醫院層級

代碼	層級	案件總個數
1	診所	116
2	地區醫院	74
3	區域醫院	139
4	醫學中心	126

### 被告對象-1

代碼	變項	案件總個數
0	醫事人員	44
1	內科	91
2	中醫科	9
3	外科	102
4	婦產科	60
5	兒科	13
6	急診科	55
7	骨科	27
8	牙科	18
9	醫美科	13
10	其他科	23

### 被告對象-2

代碼	對象	案件總個數
0	醫事人員	44
1	同科別	411



- 各科別的被告比例

科別	總計	被告比例
內科, 中醫科, 兒科	128	28%
外科	104	23%
骨科, 牙科, 其他	76	17%
婦產科	74	16%
急診科, 醫美科	73	16%
總計	455	

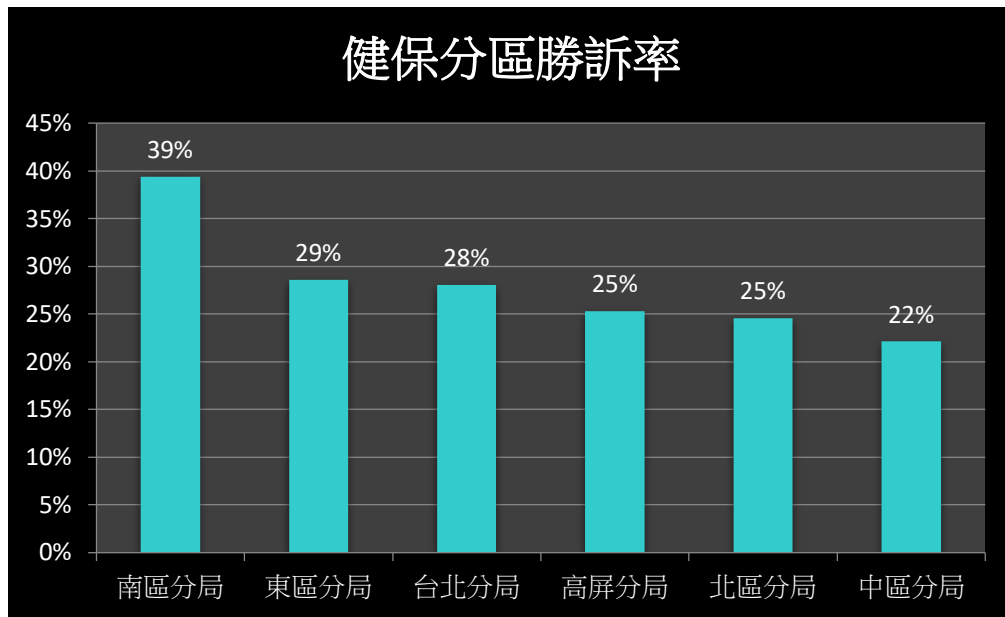
這裡可以看出被告是外科和內科為醫療糾紛事件的最大宗，可是相對於母體來說我們並不知道每個科別的醫療總數，很難直接斷言外科和內科是最容易被告的科別，例如，全國如果有 1000 名婦產科醫生，被告了 74 件，而有 2000 名外科醫生被告了 104 件，此時相對來說婦產科是相對容易被告的，總而言之，我們認為科別間是會影響被告機率，但我們並無法說明那些科別是相對容易被告的。

- 敘述統計(原告勝訴率)

#### 健保分局

健保分局	勝訴	非勝訴	總計	勝訴率
南區分局	13	20	33	39%
東區分局	4	10	14	29%
台北分局	46	118	164	28%
高屏分局	21	62	83	25%
北區分局	14	43	57	25%
中區分局	23	81	104	22%
總計	121	334	455	27%

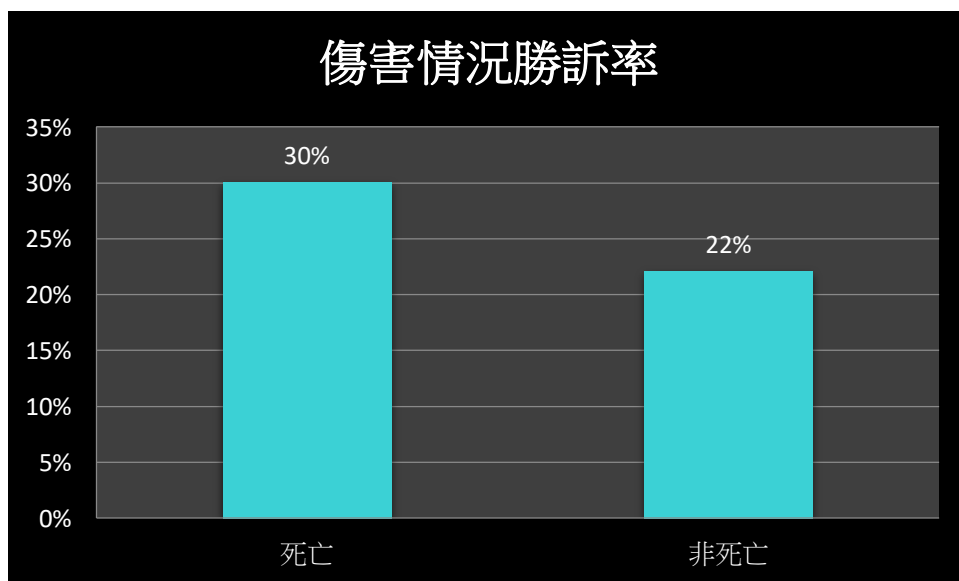




從這張圖中可以發現南區分局是相對勝率上比其他分局來的高，也就是說發生在南區醫療分局中的案子對於原告來說告贏的機率比較大，但是原因不明，可能是南區的醫療體系有關，但也非常可能是跟訴訟體系有關。

#### 傷害情況

傷害情況	非勝訴	勝訴	總計	勝訴率
死亡	179	77	256	30%
非死亡	155	44	199	22%
總計	334	121	455	27%



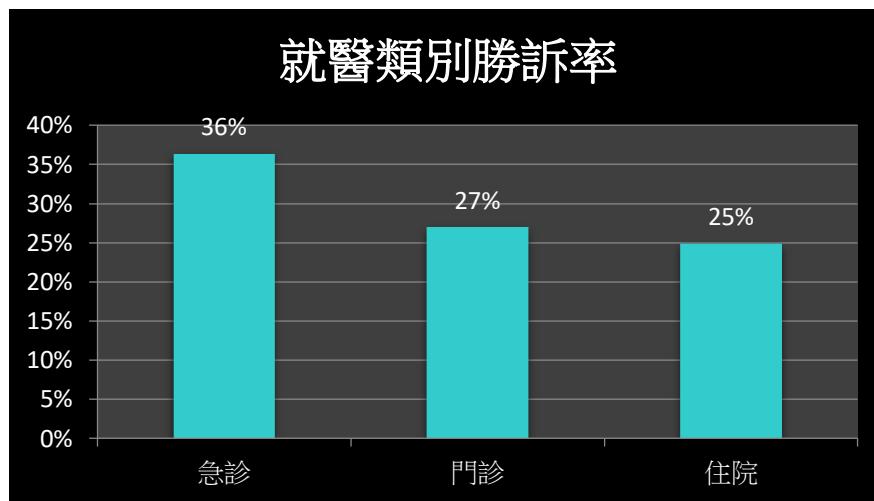
在傷害情況中，可以發現**死亡**的勝訴率最高





## 就醫類別

就醫類別	勝訴	非勝訴	總計	勝訴率
急診	16	28	44	36%
門診	34	92	126	27%
住院	71	214	285	25%
總計	121	334	455	73%

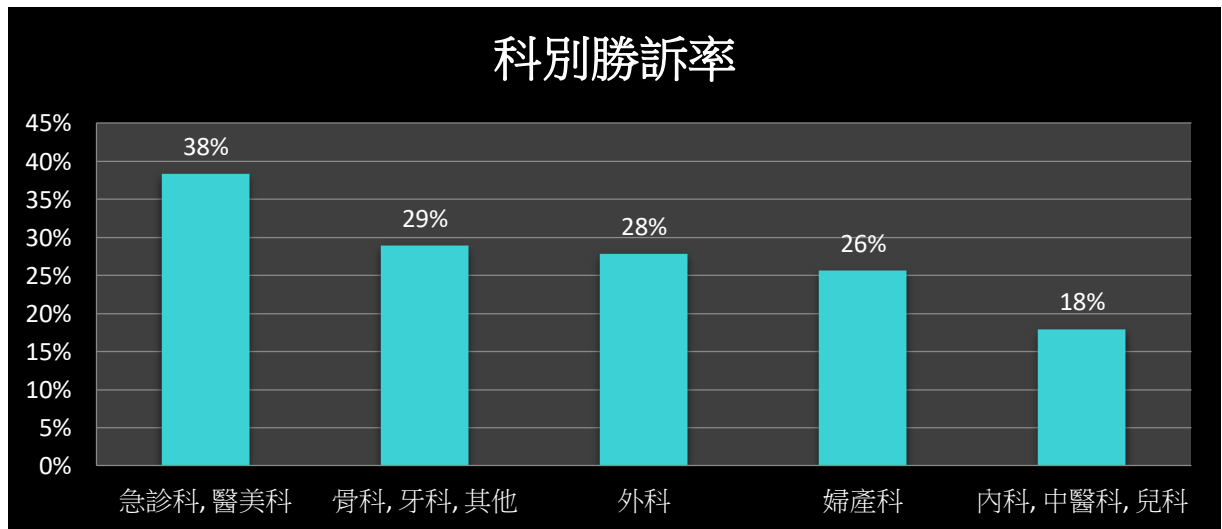


就醫類別的勝訴率急診最高，但相差比例不算大，在統計中就醫類別與勝訴率無關

## 科別

科別	非勝訴	勝訴	總計	勝訴率
急診科, 醫美科	45	28	73	38%
骨科, 牙科, 其他	54	22	76	29%
外科	75	29	104	28%
婦產科	55	19	74	26%
內科, 中醫科, 兒科	105	23	128	18%
總計	334	121	455	27%

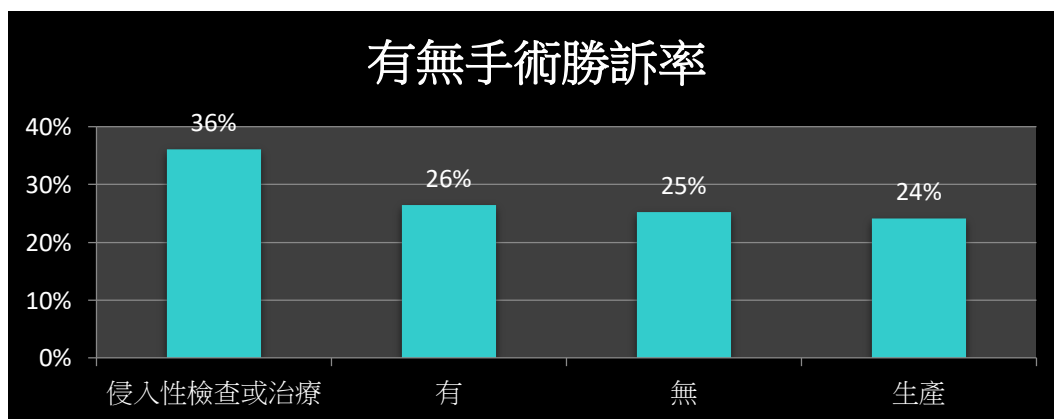




在科別中，我們發現不同科別有不同的勝率，也發現科別與勝訴率有關，依照統計的觀點，各個科別的樣本個數太少，如果提高樣本個數我想會有相當不同的結果

### 有無手術

有無手術	勝訴	非勝訴	總計	勝訴率
侵入性檢查或治療	13	23	36	36%
有	54	150	204	26%
無	47	139	186	25%
生產	7	22	29	24%
總計	121	334	455	27%

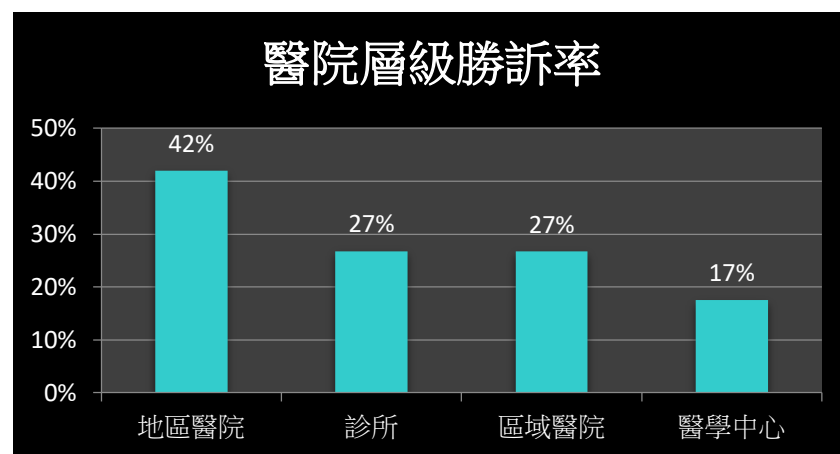


有無手術的勝訴率侵入性檢查治療最高，但相差比例不算大，在統計中**有無手術與勝訴率無關**



## 醫院層級

醫院層級	勝訴	非勝訴	總計	勝訴率
地區醫院	31	43	74	42%
診所	31	85	116	27%
區域醫院	37	102	139	27%
醫學中心	22	104	126	17%
總計	121	334	455	27%

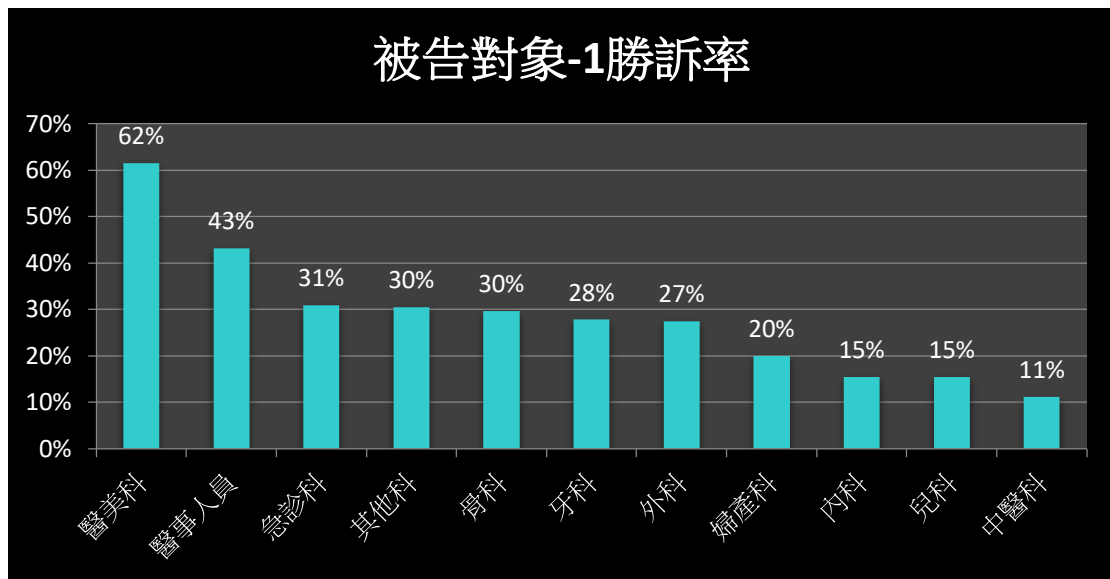


醫院層級的勝訴率看起來有相當大的不同，在統計的檢定中，也發現醫院層級與勝訴率有關，也就是說從表格中可以看出，地區醫院的勝率是最高的，最低是醫中心，其原因值得深入討論。

## 被告對象-1

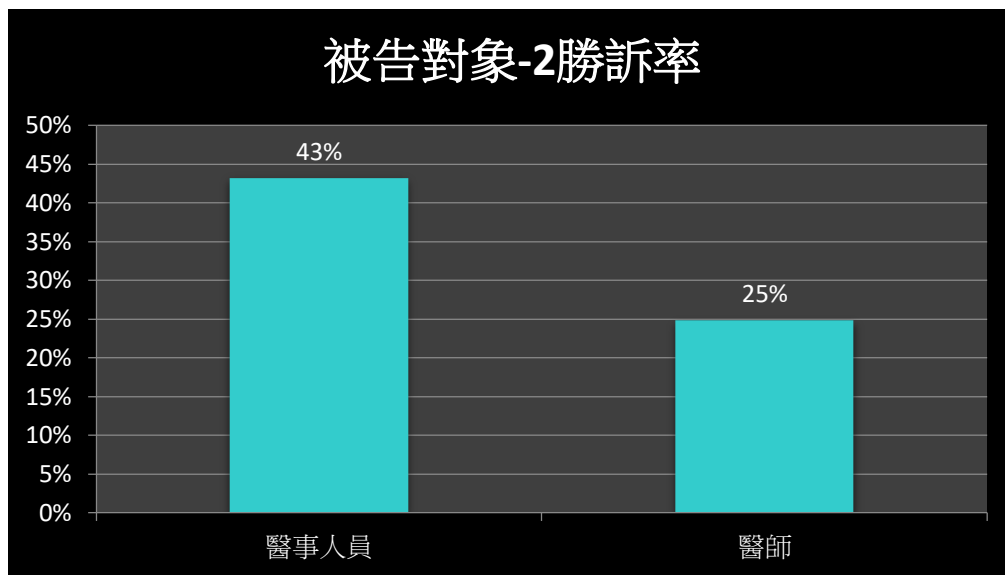
被告對象-1	勝訴	非勝訴	總計	勝訴率
醫美科	8	5	13	62%
醫事人員	19	25	44	43%
急診科	17	38	55	31%
其他科	7	16	23	30%
骨科	8	19	27	30%
牙科	5	13	18	28%
外科	28	74	102	27%
婦產科	12	48	60	20%
內科	14	77	91	15%
兒科	2	11	13	15%
中醫科	1	8	9	11%
總計	121	334	455	27%





### 被告對象-2

被告對象-2	勝訴	非勝訴	總計	勝訴率
醫事人員	19	25	44	43%
醫師	102	309	411	25%
總計	121	334	455	27%



我們將被告對象分成了醫事人員與醫師，可以發現的兩者比例相當不同，在統計檢定中也驗證了這件事，被告對象與勝訴率是相當有關係的，也就是說醫事人員被告之後，相對起醫生來說來得容易被告成功，其背後原因值得有專業知識的人員去探討研究。



- 敘述統計(科別死亡比例 vs 勝訴率)

科別	非死亡	死亡	總計	死亡比例	勝訴率
急診科, 醫美科	43	85	128	66%	18%
骨科, 牙科, 其他	48	56	104	54%	28%
外科	24	49	73	67%	38%
婦產科	31	43	74	58%	26%
內科, 中醫科, 兒科	53	23	76	30%	29%
總計	199	256	455	56%	27%

這是統計出不同的科別會被提告相害程度不同，我們發現以急診醫美科，如果醫死人相對來說被提告的機率就會大得多。

- 統計方法

### 相關性

使用卡方檢定計算各變數對勝敗訴是否有顯著影響

	健保分局	傷害情況	就醫類別	科別	有無手術	醫院層級	被告對象
X-squared	4.2358	3.2443	2.574,	10.3879	1.9289,	14.2569,	5.9578,
df	5	1	2	4	3	3	1
p-value	0.516	0.07167	0.2761	0.03438	0.5873	0.002576	0.01465

結論:

發現科別跟醫院層級跟被告對象有顯著的影響

(傷害情況 p-value 感覺也滿小,用 Logistics regression 下去檢驗)

### Logistics regression

經由卡方檢定所篩選出的顯著影響勝敗訴之變數(醫院層級跟被告對象)，本組進而使用 logistics regression 找出最適合的模型

- 傷害情況

代碼	情況	虛擬變數
1	非死亡	傷害情況 3=0
2	死亡	傷害情況 3=1



➤ 科別

代碼	科別	虛擬變數
1	內科, 中醫科, 兒科	科別 2=0, 科別 3=0, 科別 4=0, 科別 5=0
2	外科	科別 2=1, 科別 3=0, 科別 4=0, 科別 5=0
3	急診科, 醫美科	科別 2=0, 科別 3=1, 科別 4=0, 科別 5=0
4	婦產科	科別 2=0, 科別 3=0, 科別 4=1, 科別 5=0
5	骨科, 牙科, 其他	科別 2=0, 科別 3=0, 科別 4=0, 科別 5=1

➤ 醫院層級

代碼	層級	虛擬變數
1	診所	醫院層級 2=0, 醫院層級 3=0, 醫院層級 4=0
2	地區醫院	醫院層級 2=1, 醫院層級 3=0, 醫院層級 4=0
3	區域醫院	醫院層級 2=0, 醫院層級 3=1, 醫院層級 4=0
4	醫學中心	醫院層級 2=0, 醫院層級 3=0, 醫院層級 4=1

➤ 被告對象-2

代碼	對象	虛擬變數
0	醫事人員	被告對象 1=0
1	同科別	被告對象 1=1

去尋找最好模型，做了下面幾個迴歸式子。

model1=glm(WL~ 醫院層級+被告對象+科別 F+傷害 F, family=binomial) AIC: 513.68
model2=glm(WL~ 醫院層級+被告對象+科別 F, family=binomial) AIC: 513.95
model2=glm(WL~ 醫院層級+被告對象+科別 F, family=binomial) AIC: 513.95
model3=glm(WL~ 醫院層級+被告對象+傷害 F, family=binomial) AIC: 517.51
model4=glm(WL~ 被告對象+科別 F+傷害 F, family=binomial) AIC: 519.09
model5=glm(WL~ 醫院層級+被告對象, family=binomial) AIC: 518.01



```
model6=glm(WL~ 醫院層級+被告對象+as.factor(科別傷害)
```

(注:科別傷害是有科別與其交互作用)

AIC: 541.87

然後比較 AIC 之後發現，model1 的表現比較好，他的係數模型如下。

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-0.9748	0.4337	-2.247	0.02462 *
醫院層級 2	0.5327	0.3342	1.594	0.11103
醫院層級 3	-0.1229	0.3096	-0.397	0.69139
醫院層級 4	-0.6069	0.3363	-1.804	0.07115 .
被告對象 1	-0.8395	0.3542	-2.370	0.01777 *
as.factor(科別)2	0.7874	0.3329	2.365	0.01803 *
as.factor(科別)3	1.0724	0.3467	3.093	0.00198 **
as.factor(科別)4	0.3362	0.3650	0.921	0.35690
as.factor(科別)5	0.7125	0.3610	1.974	0.04841 *
as.factor(傷害)3	0.4065	0.2378	1.709	0.08740 .

從這個模型可以發現，**傷害情況**其實加入迴歸模型中比較好的。

代碼	情況	案件總個數
1	非死亡	199
2	死亡	256

代碼	科別	案件總個數
1	內科, 中醫科, 兒科	128
2	外科	104
3	急診科, 醫美科	73
4	婦產科	74
5	骨科, 牙科, 其他	76

代碼	層級	案件總個數
1	診所	116
2	地區醫院	74
3	區域醫院	139
4	醫學中心	126



代碼	對象	案件總個數
0	醫事人員	44
1	醫生	411

### 使用時機:

1. 應變數為類別型變數。
2. 主要針對二項式(binomial)的應變數亦可推廣至應變數為多項式時使用。
3. 迴歸模式中包含一個應變數及多個自變數。

### 使用方法:

根據 model1

CASE1

(診所,醫師,外科,死亡) 代碼為(2,1,2,2)

$f(x) = -0.9748 + 0 + 0.7874 - 0.8395 + 0.4065 = -0.6204$

$P(\text{勝算}) = 0.3496905$

CASE2

(區域醫院,醫事人員,急診科,非死亡) 代碼為(3,0,3,1)

$f(x) = -0.9748 - 0.1229 + 0 + 1.0724 + 0 = -0.0253$

$p(\text{勝算}) = 0.4936753$

### HLM

在羅吉斯迴歸之後，不同的醫院層級有不同的羅吉斯迴歸

編號	健保分局	傷害情況	原告勝敗	就醫類別	科別	有無手術	醫院層級	被告對象
id	local	damage	WL	Medic_cat	division	operation	level	defendant

接著做階層線性模型，所以選出 damage， division， level 和 defendant

- Model 1：不同醫院層級有不同的被告對象的影響程度





```

$醫院層級
(Intercept) 被告對象1 科別F2 科別F3 科別F4 科別F5 傷害F3
1 -1.016426 -0.7672171 0.7903249 1.111828 0.3527162 0.7411896 0.4515277
2 -1.016426 -0.5769085 0.7903249 1.111828 0.3527162 0.7411896 0.4515277
3 -1.016426 -0.9952585 0.7903249 1.111828 0.3527162 0.7411896 0.4515277
4 -1.016426 -1.2652950 0.7903249 1.111828 0.3527162 0.7411896 0.4515277

```

➤ Model 2：不同醫院層級有不同的科別的影響程度

```

$醫院層級
(Intercept) 被告對象1 科別F2 科別F3 科別F4 科別F5 傷害F3
1 -0.9080602 -0.9023602 0.9067211 1.1929654 0.2264030 0.7688077 0.4149042
2 -0.3900066 -0.9023602 0.5586252 0.7781549 -0.4485313 0.9146086 0.4149042
3 -1.3402534 -0.9023602 1.1233274 1.4591626 0.8056314 0.6159401 0.4149042
4 -1.2898679 -0.9023602 0.5736313 0.8605747 0.8528543 0.4118342 0.4149042

```

➤ Model 3：不同醫院層級有不同的傷害的影響程度

```

$醫院層級
(Intercept) 被告對象1 科別F2 科別F3 科別F4 科別F5 傷害F3
1 -0.9990826 -0.8875755 0.7030649 0.9732551 0.3726742 0.7186848 0.2694611
2 -0.9991437 -0.8875755 0.7030649 0.9732551 0.3726742 0.7186848 0.9467645
3 -0.9991126 -0.8875755 0.7030649 0.9732551 0.3726742 0.7186848 0.6017003
4 -0.9990450 -0.8875755 0.7030649 0.9732551 0.3726742 0.7186848 -0.1475767

```

➤ Model 4 不同醫院層級只有不同的截距

```

$醫院層級
(Intercept) 被告對象1 科別F2 科別F3 科別F4 科別F5 傷害F3
1 -0.9902330 -0.8680984 0.7678362 1.086789 0.3622181 0.7442738 0.4170675
2 -0.6611916 -0.8680984 0.7678362 1.086789 0.3622181 0.7442738 0.4170675
3 -1.0667015 -0.8680984 0.7678362 1.086789 0.3622181 0.7442738 0.4170675
4 -1.3754598 -0.8680984 0.7678362 1.086789 0.3622181 0.7442738 0.4170675

```

比較這 4 種模型

➤ ANOVA



```

> anova(ghlm1,ghlm2,ghlm3,ghlm4)
Data:
Models:
ghlm4: WL ~ 被告對象 + 科別F + 傷害F + (1 | 醫院層級)
ghlm1: WL ~ 被告對象 + 科別F + 傷害F + (被告對象 | 醫院層級)
ghlm3: WL ~ 被告對象 + 科別F + 傷害F + (傷害F | 醫院層級)
ghlm2: WL ~ 被告對象 + 科別F + 傷害F + (科別F | 醫院層級)
      Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
ghlm4   8 518.32 551.28 -251.16   502.32
ghlm1  10 522.46 563.66 -251.23   502.46 0.0000      2    1.0000
ghlm3  10 519.96 561.16 -249.98   499.96 2.5007      0    <2e-16 ***
ghlm2  22 542.59 633.24 -249.29   498.59 1.3651     12    0.9999
---

```

最後從 AIC 與 BIC 的比較中發現模型 4 比較好，所以得到以下的結論。  
 不同的醫院層級是會影響原告在醫療糾紛中的訴訟勝率，但是不同的醫院層級必  
 不會改變，被告對象，科別或傷害程度對於原告對於訴訟勝率的影響程度，簡而  
 言之，這筆資料並不適合對於醫院層級來使用階層線性模型，原本的羅吉斯迴歸  
 已是一個相當好個迴歸模型。

## 使用方法

