

## § 流病學原理資料分析課堂作業 3

流行病學資料分析 –  
SAS  
Practice 3  
學號：b07401048  
系級：醫學五

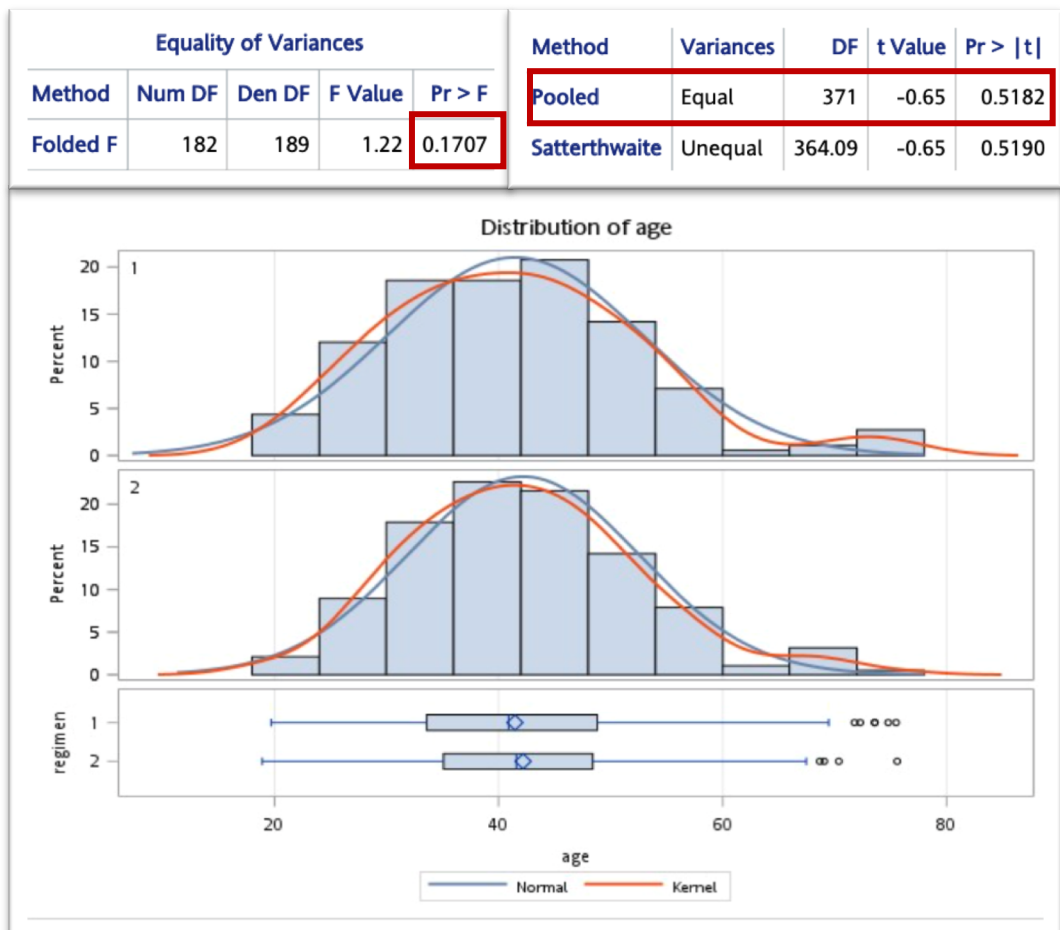
### 一. 請比較被隨機分派到 INH 與 RMP 的兩組 LTBI 病人？

#### (一) 年齡

##### 1. Results：

- (1) 由於年齡為等距變項，故以兩獨立樣本 t 檢定判斷 INH 與 RMP 兩組年齡是否有顯著差異。
- (2) 先以 folded F 判斷兩者之變異數是否相等（如下左圖）。虛無假說為兩者變異數沒有顯著差異，對立假說為兩者變異數有顯著差異。檢定結果之 p-value 為 0.1707，無法拒絕虛無假說，兩組變異數沒有顯著差異。
- (3) 之後以假定兩組變異數相等的 pooled two-sampled t-test 檢定兩組年齡平均數是否有顯著差異（如下右圖）。虛無假說為兩者沒有年齡平均數沒有顯著差異，對立假說為兩者年齡平均數有顯著差異。檢定結果之 p-value 為 0.5182，無法拒絕虛無假說，兩組年齡平均數沒有顯著差異。
- (4) 年齡為影響肝毒性風險的重要因素，在隨機分派的試驗中，是必須需確保實驗組與對照組的年齡相等，以免影響結果判讀。INH 與 RMP 兩組的年齡平均數沒有顯著差異，符合隨機分派的標準。

##### 2. Figures：



### 3. Code :

```
/* homework 3 */
/* Q1 */
/* Q1-1 : AGE <- t-test */
title Q1_1;

proc ttest data = dataset1;
  class regimen;
  var age;
run;

/* extra: check if age is in normal distribution */
proc univariate data = dataset1 normal plot;
  var age;
run; /* actually not in norm, but we have big data (300 > 30), can
ignore
```

## (二)HBV

### 1. Result :

- (1) HBV 帶原（有無）為類別變項，故以卡方檢定判斷 RMP 與 INH 兩組之有無 HBV 帶原是否有顯著差異
- (2) 由左圖之列聯表可見 INH 組（regimen 1）HBV 帶原的比例為 13.11%，RMP 組（regimen 2）HBV 帶原的比例為 14.74%。
- (3) 以卡方檢定判斷兩組 HBV 帶原的比例是否有顯著差異，虛無假說為 HBV 帶原與分組為獨立事件，對立假說為 HBV 帶原與分組有相關性。檢定結果之 p-value 為 0.6512，無法拒絕虛無假說，表示 HBV 帶原與分組為獨立事件，兩組 HBV 帶原的比例沒有顯著差異。
- (4) HBV 帶原會導致慢性肝炎，與藥物肝毒性的感受性有重大關係，在隨機分派的試驗中需確保兩組的 HBV 帶原率相等，以免影響結果判讀。**INH 與 RMP 兩組的 HBV 帶原率沒有顯著差異**，符合隨機分派的標準。

### 2. Figures :

The FREQ Procedure				
Frequency Percent Row Pct Col Pct	Table of regimen by hbv			
	regimen(regimen)	hbv(hbv)		Total
		0	1	
1	159	24	183	
	42.63	6.43	49.06	
	86.89	13.11		
	49.53	46.15		
2	162	28	190	
	43.43	7.51	50.94	
	85.26	14.74		
	50.47	53.85		
Total	321	52	373	
	86.06	13.94	100.00	

Statistics for Table of regimen by hbv			
Statistic	DF	Value	Prob
Chi-Square	1	0.2044	0.6512
Likelihood Ratio Chi-Square	1	0.2047	0.6510
Continuity Adj. Chi-Square	1	0.0916	0.7622
Mantel-Haenszel Chi-Square	1	0.2039	0.6516
Phi Coefficient		0.0234	
Contingency Coefficient		0.0234	
Cramer's V		0.0234	

### 3. Code :

```
/* Q1-2 : HBV <- chi-square */  
title Q1_2;  
proc freq data = dataset1;  
    tables regimen * hbv / chisq;  
run;
```

### (三)HCV：

#### 1. Result：

- (1) HCV 帶原（有無）為類別變項，故以卡方檢定判斷 RMP 與 INH 兩組之有無 HCV 帶原率是否有顯著差異
- (2) 由左圖之列聯表可見 INH 組（regimen 1）HCV 帶原的比例為 20.77%，RMP 組（regimen 2）HCV 帶原的比例為 22.11%。
- (3) 以卡方檢定判斷兩組 HCV 帶原的比例是否有顯著差異，虛無假說為 HCV 帶原與分組為獨立事件，對立假說為 HCV 帶原與分組有相關性。檢定結果之 p-value 為 0.7526，無法拒絕虛無假說，表示 HCV 帶原與分組為獨立事件，兩組 HBV 帶原的比例沒有顯著差異。
- (4) HCV 會導致慢性肝炎，與藥物肝毒性的感受性有重大關係，在隨機分派的試驗中需確保兩組的 HCV 帶原率相等，以免影響結果判讀。INH 與 RMP 兩組的 HCV 帶原率沒有顯著差異，符合隨機分派的標準。

#### 2. Figures：

The FREQ procedure				
Frequency Percent Row Pct Col Pct	Table of regimen by hcv			
	regimen(regimen)	hcv(hcv)		
		0	1	Total
1	145	38	183	
	38.87	10.19	49.06	
	79.23	20.77		
	49.49	47.50		
2	148	42	190	
	39.68	11.26	50.94	
	77.89	22.11		
	50.51	52.50		
Total	293	80	373	
	78.55	21.45	100.00	

Statistics for Table of regimen by hcv			
Statistic	DF	Value	Prob
Chi-Square	1	0.0994	0.7526
Likelihood Ratio Chi-Square	1	0.0994	0.7525
Continuity Adj. Chi-Square	1	0.0358	0.8500
Mantel-Haenszel Chi-Square	1	0.0991	0.7529
Phi Coefficient		0.0163	
Contingency Coefficient		0.0163	
Cramer's V		0.0163	

#### 3. Code：

```
/* Q1-3 : */  
title Q1_3;  
proc freq data = dataset1;  
    tables regimen * hcv / chisq;  
run;
```

#### (四)收案時的肝指數是否有異常

##### 1. Result :

- (1) 收案時肝指數（GPT）是否有異常（有無）為類別變項，故以卡方檢定判斷 RMP 與 INH 兩組病人在收案時肝指數有異常的比例是否有顯著差異。
- (2) 由左圖之列聯表可見 INH 組（regimen 1）收案時肝指數有異常的比例為 21.31%，RMP 組（regimen 2）收案時肝指數有異常的比例為 20.53%。
- (3) 以卡方檢定判斷兩組收案時肝指數有異常的比例是否有顯著差異，虛無假說為收案時肝指數有異常與分組為獨立事件，對立假說為收案時肝指數有異常與分組有相關性。檢定結果之 p-value 為 0.8521，無法拒絕虛無假說，表示收案時的肝指數異常與分組為獨立事件，兩組收案時肝指數異常的比例沒有顯著差異。
- (4) 肝指數可代表 baseline 的肝功能，與藥物肝毒性的感受性有重大關係，在隨機分派的試驗中需確保兩組 baseline 肝功能相等，以免影響結果判讀。**INH 與 RMP 兩組收案時肝指數異常率沒有顯著差異**，符合隨機分派的標準。

##### 2. Figures :

The FREQ Procedure				
Frequency Percent Row Pct Col Pct	Table of regimen by GPT_UNL			
	regimen(regimen)	GPT_UNL(GPT_UNL)		
		0	1	Total
	1	144	39	183
		38.61	10.46	49.06
		78.69	21.31	
		48.81	50.00	
	2	151	39	190
		40.48	10.46	50.94
		79.47	20.53	
		51.19	50.00	
	Total	295	78	373
		79.09	20.91	100.00

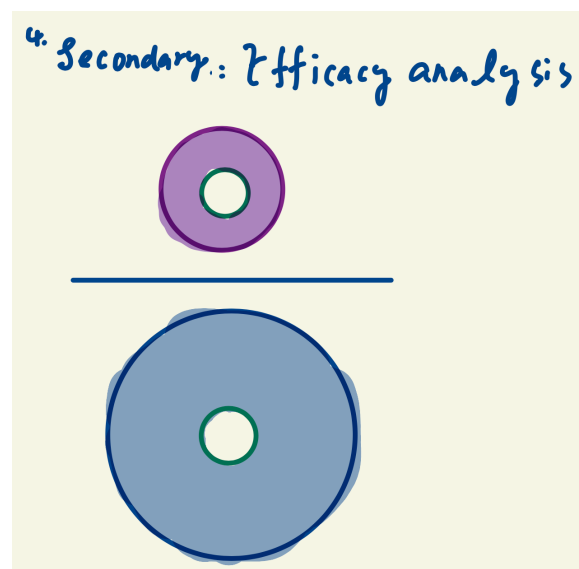
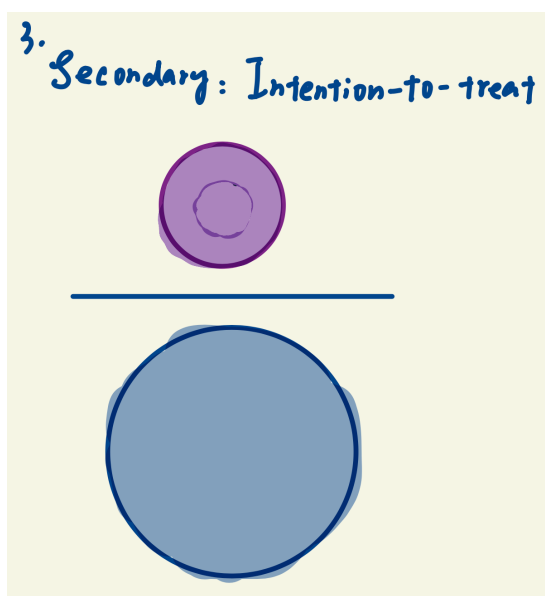
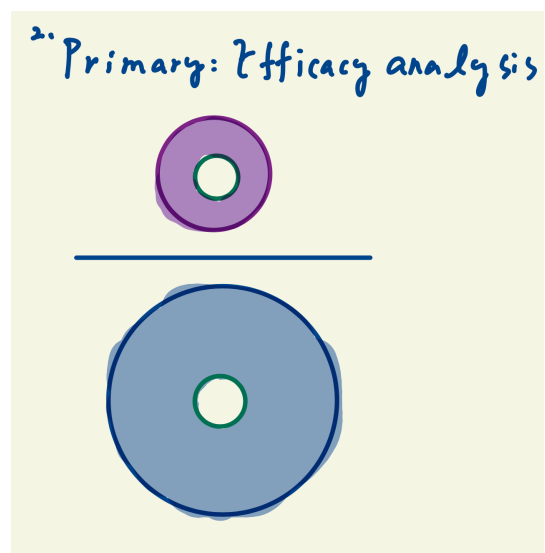
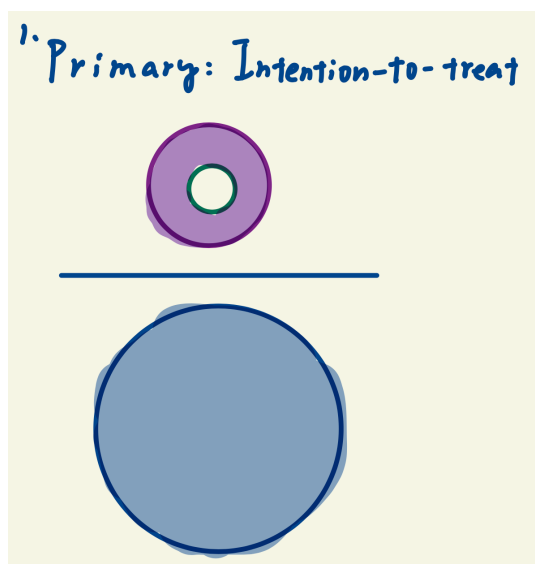
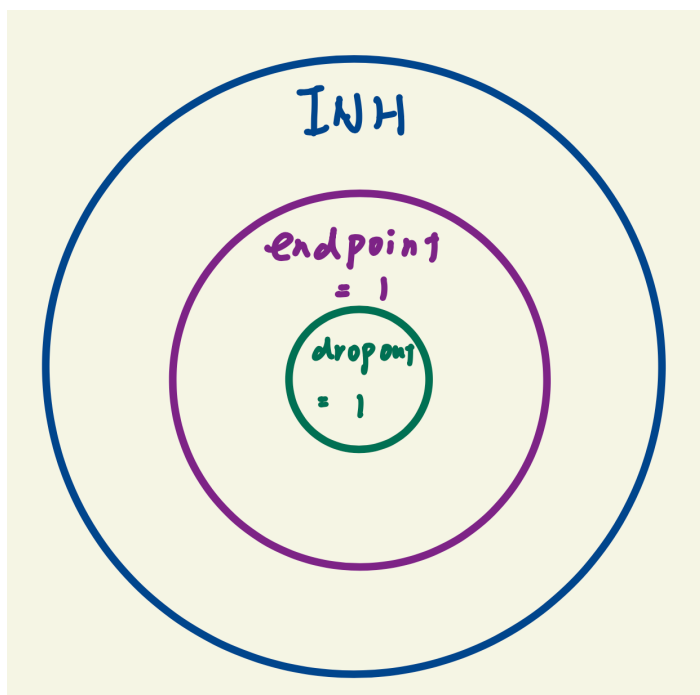
Statistics for Table of regimen by GPT_UNL			
Statistic	DF	Value	Prob
Chi-Square	1	0.0347	0.8521
Likelihood Ratio Chi-Square	1	0.0347	0.8521
Continuity Adj. Chi-Square	1	0.0035	0.9529
Mantel-Haenszel Chi-Square	1	0.0347	0.8523
Phi Coefficient		-0.0097	
Contingency Coefficient		0.0097	
Cramer's V		-0.0097	

##### 3. Code :

```
/* Q1-4 : */  
title Q1_4;  
proc freq data = dataset1;  
    tables regimen * GPT_UNL / chisq;  
run;
```

二. 請比較以下評估項目 RMP 與 INH 兩組是否有統計上的顯著差異？

(一)簡圖：



## (二) Primary endpoint, by intention-to-treat

### 1. Result :

- (1) 本試驗的 primary endpoint 定義為「因藥物肝毒性或其他嚴重副作用而被迫停藥」，intention-to-treat 則表示以收案時該組人數表示該組的總人數。
- (2) 是否達到 primary endpoint（有無）為類別變項，故以卡方檢定判斷 RMP 與 INH 兩組病人達到 primary endpoint 的比例是否有顯著差異。
- (3) 由左圖之列聯表可見 INH 組（regimen 1）達到 primary endpoint 的比例為 12.02%，RMP 組（regimen 2）達到 primary endpoint 的比例為 2.11%。
- (4) 同時，也可以 risk 敘述：INH 組病人在 INH 療程的時間內，發生 primary endpoint 的 risk 為 12.02%；RMP 組在 RMP 療程的時間內，發生 primary endpoint 的 risk 為 2.11%，兩組的 risk difference 為 9.92%。
- (5) 以卡方檢定判斷兩組達到 primary endpoint 的比例是否有顯著差異，虛無假說為達到 primary endpoint 與分組為獨立事件，對立假說達到 primary endpoint 與分組有相關性。檢定結果之 p-value 為 0.0002，拒絕虛無假說，表示兩組達到 primary endpoint 的比例有顯著差異。

### 2. Figures :

Column 2 Risk Estimates						
	Risk	ASE	95% Confidence Limits		Exact 95% Confidence Limits	
Row 1	0.1202	0.0240	0.0731	0.1673	0.0769	0.1763
Row 2	0.0211	0.0104	0.0006	0.0415	0.0058	0.0530
Total	0.0697	0.0132	0.0439	0.0955	0.0460	0.1005
Difference	0.0992	0.0262	0.0478	0.1505		
Difference is (Row 1 - Row 2)						

Frequency Percent Row Pct Col Pct				
Table of regimen by primary				
regimen(regimen)	primary			Total
	0	1		
1	161	22		183
	43.16	5.90		49.06
	87.98	12.02		
	46.40	84.62		
2	186	4		190
	49.87	1.07		50.94
	97.89	2.11		
	53.60	15.38		
Total	347	26		373
	93.03	6.97		100.00

Statistics for Table of regimen by primary			
Statistic	DF	Value	Prob
Chi-Square	1	14.1363	0.0002
Likelihood Ratio Chi-Square	1	15.3902	<.0001
Continuity Adj. Chi-Square	1	12.6484	0.0004
Mantel-Haenszel Chi-Square	1	14.0984	0.0002
Phi Coefficient		-0.1947	
Contingency Coefficient		0.1911	
Cramer's V		-0.1947	

### 3. Code :

```
/* Q2 */
/* Q2-1 primary endpoint by intention-to-treat*/
title Q2_1;
data primary;
    set dataset1;

    if endpoint - dropout = 1 then primary = 1;
    else primary = 0;
run;

proc print data = primary;
run;
/* chisquare + risk difference */
proc freq data = primary;
    tables regimen * primary / chisq riskdiff;
run;
```



### (三)Primary endpoint, by efficacy analysis

#### 1. Result :

- (1) 本試驗的 primary endpoint 定義為「因藥物肝毒性或其他嚴重副作用而被迫停藥」，efficacy 則表示不納入中途 dropout 的人，以完成療程的病人為該組的人數。
- (2) 是否達到 primary endpoint（有無）為類別變項，故以卡方檢定判斷 RMP 與 INH 兩組病人達到 primary endpoint 的比例是否有顯著差異。
- (3) 由左圖之列聯表可見 INH 組完成療程的病人中（regimen 1）達到 primary endpoint 的比例為 13.41%，RMP 組完成療程的病人中（regimen 2）達到 primary endpoint 的比例為 2.40%。
- (4) 同時，也可以 risk 敘述：INH 組病人在完成 INH 療程的時間內，發生 primary endpoint 的 risk 為 13.41%；RMP 組在完成 RMP 療程的時間內，發生 primary endpoint 的 risk 為 2.40%，兩組的 risk difference 為 11.02%。
- (5) 以卡方檢定判斷兩組達到 primary endpoint 的比例是否有顯著差異，虛無假說為達到 primary endpoint 與分組為獨立事件，對立假說達到 primary endpoint 與分組有相關性。檢定結果之 p-value 為 0.0002，拒絕虛無假說，表示兩組達到 primary endpoint 的比例有顯著差異。

#### 2. Figures :

Column 2 Risk Estimates						
	Risk	ASE	95% Confidence Limits		Exact 95% Confidence Limits	
Row 1	0.1341	0.0266	0.0820	0.1863	0.0860	0.1960
Row 2	0.0240	0.0118	0.0008	0.0471	0.0066	0.0602
Total	0.0785	0.0148	0.0496	0.1075	0.0520	0.1130
Difference	0.1102	0.0291	0.0531	0.1673		
Difference is (Row 1 - Row 2)						

The FREQ Procedure				
Frequency Percent Row Pct Col Pct	Table of regimen by primary			
	regimen(regimen)	primary		
		0	1	Total
	1	142 42.90 86.59 46.56	22 6.65 <b>13.41</b> 84.62	164 49.55
	2	163 49.24 97.60 53.44	4 1.21 <b>2.40</b> 15.38	167 50.45
	Total	305 92.15	26 7.85	331 100.00

Statistics for Table of regimen by primary			
Statistic	DF	Value	Prob
Chi-Square	1	13.8814	0.0002
Likelihood Ratio Chi-Square	1	15.1387	<.0001
Continuity Adj. Chi-Square	1	12.4007	0.0004
Mantel-Haenszel Chi-Square	1	13.8395	0.0002
Phi Coefficient		-0.2048	
Contingency Coefficient		0.2006	
Cramer's V		-0.2048	

### 3. Code :

```
/* Q2-2 primary endpoint by efficacy analysis */  
title Q2_2;  
data primary_efficacy;  
    set dataset1;  
  
    if dropout = 1 then delete;  
  
    if endpoint = 1 then primary = 1;  
    else primary = 0;  
  
run;  
  
proc print data = primary_efficacy;  
run;  
  
proc freq data = primary_efficacy;  
    tables regimen * primary / chisq riskdiff;  
run;
```

#### (四)Secondary endpoint, by intention-to-treat

##### 1. Result :

- (1) 本試驗的 secondary endpoint 定義為「因任何原因而停藥（包含 drop out）」，intention-to-treat 則表示以收案時該組人數表示該組的總人數。
- (2) 是否達到 secondary endpoint（有無）為類別變項，故以卡方檢定判斷 RMP 與 INH 兩組病人達到 secondary endpoint 的比例是否有顯著差異。
- (3) 由左圖之列聯表可見 INH 組（regimen 1）達到 secondary endpoint 的比例為 22.40%，RMP 組（regimen 2）達到 secondary endpoint 的比例為 14.21%。
- (4) 同時，也可以 risk 敘述：INH 組病人在 INH 療程的時間內，發生 secondary endpoint 的 risk 為 22.40%；RMP 組在 RMP 療程的時間內，發生 primary endpoint 的 risk 為 14.21%，兩組的 risk difference 為 8.19%。
- (5) 以卡方檢定判斷兩組達到 secondary endpoint 的比例是否有顯著差異，虛無假說為達到 secondary endpoint 與分組為獨立事件，對立假說達到 secondary endpoint 與分組有相關性。檢定結果之 p-value 為 0.0405，拒絕虛無假說，表示兩組達到 secondary endpoint 的比例有顯著差異。

##### 2. Figures :

Column 2 Risk Estimates						
	Risk	ASE	95% Confidence Limits		Exact 95% Confidence Limits	
Row 1	0.2240	0.0308	0.1636	0.2845	0.1658	0.2914
Row 2	0.1421	0.0253	0.0925	0.1918	0.0958	0.2000
Total	0.1823	0.0200	0.1431	0.2215	0.1444	0.2253
Difference	0.0819	0.0399	0.0037	0.1601		
Difference is (Row 1 - Row 2)						

THE FREQ PROCEDURE				
Frequency Percent Row Pct Col Pct	Table of regimen by endpoint			
	regimen(regimen)	endpoint(endpoint)		
		0	1	Total
	1	142	41	183
		38.07	10.99	49.06
		77.60	22.40	
		46.56	60.29	
	2	163	27	190
		43.70	7.24	50.94
		85.79	14.21	
		53.44	39.71	
	Total	305	68	373
		81.77	18.23	100.00

Statistic	DF	Value	Prob
Chi-Square	1	4.1984	0.0405
Likelihood Ratio Chi-Square	1	4.2187	0.0400
Continuity Adj. Chi-Square	1	3.6667	0.0555
Mantel-Haenszel Chi-Square	1	4.1871	0.0407
Phi Coefficient		-0.1061	
Contingency Coefficient		0.1055	
Cramer's V		-0.1061	

### 3. code

```
/* Q2-3 secondary endpoint by intention-to-treat */  
title Q2_3;  
proc freq data = dataset1;  
    tables regimen * endpoint / chisq riskdiff;  
run;
```

## (五) Secondary endpoint, by efficacy analysis

### 1. Result :

- (1) 本試驗的 secondary endpoint 定義為「因任何原因而停藥」，efficacy 則表示不納入中途 dropout 的人，以完成療程的病人為該組的人數。
- (2) 是否達到 secondary endpoint（有無）為類別變項，故以卡方檢定判斷 RMP 與 INH 兩組病人達到 secondary endpoint 的比例是否有顯著差異。
- (3) 由左圖之列聯表可見 INH 組完成療程的病人中（regimen 1）達到 secondary endpoint 的比例為 13.41%，RMP 組完成療程的病人中（regimen 2）達到 secondary endpoint 的比例為 2.40%。
- (4) 同時，也可以 risk 敘述：INH 組病人在完成 INH 療程的時間內，發生 secondary endpoint 的 risk 為 13.41%；RMP 組在完成 RMP 療程的時間內，發生 secondary endpoint 的 risk 為 2.40%，兩組的 risk difference 為 11.02%。
- (5) 以卡方檢定判斷兩組收案時肝指數有異常的比例是否有顯著差異，虛無假說為達到 secondary endpoint 與分組為獨立事件，對立假說達到 secondary endpoint 與分組有相關性。檢定結果之 p-value 為 0.0002，拒絕虛無假說，表示兩組達到 secondary endpoint 的比例有顯著差異。

### 2. Figures :

Column 2 Risk Estimates						
	Risk	ASE	95% Confidence Limits		Exact 95% Confidence Limits	
Row 1	0.1341	0.0266	0.0820	0.1863	0.0860	0.1960
Row 2	0.0240	0.0118	0.0008	0.0471	0.0066	0.0602
Total	0.0785	0.0148	0.0496	0.1075	0.0520	0.1130
Difference	0.1102	0.0291	0.0531	0.1673		
Difference is (Row 1 - Row 2)						

The FREQ Procedure				
Frequency Percent Row Pct Col Pct	Table of regimen by secondary			
	regimen(regimen)	secondary		
		0	1	Total
	1	142	22	164
		42.90	6.65	49.55
		86.59	13.41	
		46.56	84.62	
	2	163	4	167
		49.24	1.21	50.45
		97.60	2.40	
		53.44	15.38	
	Total	305	26	331
		92.15	7.85	100.00

Statistic	DF	Value	Prob
Chi-Square	1	13.8814	0.0002
Likelihood Ratio Chi-Square	1	15.1387	<.0001
Continuity Adj. Chi-Square	1	12.4007	0.0004
Mantel-Haenszel Chi-Square	1	13.8395	0.0002
Phi Coefficient		-0.2048	
Contingency Coefficient		0.2006	
Cramer's V		-0.2048	

### 3. Code :

```
/* Q2-4 secondary endpoint by efficacy analysis */  
title Q2_4;  
data secondary_efficacy;  
    set dataset1;  
  
    if dropout = 1 then delete;  
  
    if endpoint = 1 then secondary = 1;  
    else secondary = 0;  
run;  
  
proc print data = secondary_efficacy;  
run;  
  
proc freq data = secondary_efficacy;  
    tables regimen * secondary / chisq riskdiff;  
run;
```