



兩樣本檢定vs資料分析

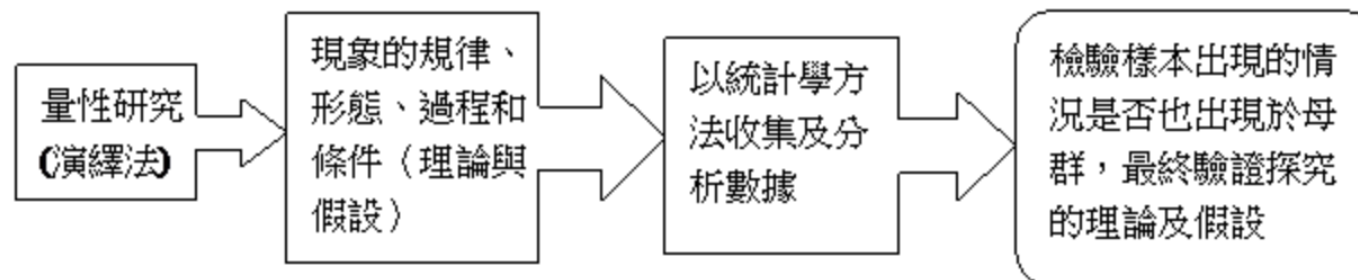
Google表單，結果分享

大綱

- ▶ 敘述性統計 (google表單分享)
- ▶ 分析的方法 (質性與量化分析)
- ▶ 資料處理與資料的型態
 - ▶ 資料來源、問卷題項
 - ▶ 抽樣方式、Data clean
 - ▶ 資料的型態與使用的檢定
 - ▶ 敘述性統計 (google表單分享)
 - ▶ 推論統計 (two population)
- ▶ 五種題型練習
 - ▶ 作業3提醒

分析的方法（質性與量化分析）

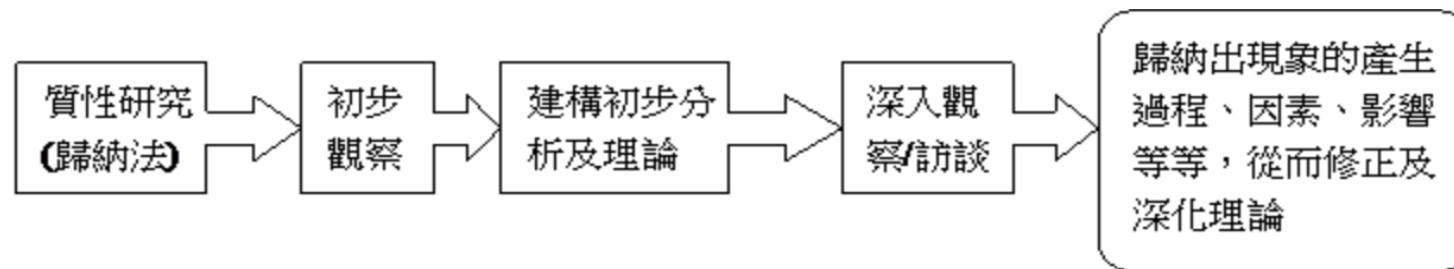
- ▶ 量性研究目的在於「描述」某現象所出現的規律、形態、過程和條件，並「檢驗」這些在樣本出現的情況是否也在探究對象的母群出現。
- ▶ 量性探究法的理念：將樣本的分析、探究結果，**推論**至母群。又稱「演繹法」（Deduction）
- ▶ 推論邏輯與一些統計學原理相當類近，故量化研究從抽樣大小、資料類型，分析方法皆以符合統計學要求為依據。



資料來源：http://www.cuhk.edu.hk/soc/isonline/ies/3/ies_l3a_samp.htm ,本研究整理

分析的方法（質性與量化分析）

- ▶ 質性研究的目的，並非「檢驗」和「推論」樣本的結果，推論回母體。而是「了解」相同探究議題，在不同個案所出現的經過，並且「比較」個案間異同之處，最後歸納整理出不同的因素。
- ▶ 質性探究所用的不是「演繹法」，而是「歸納法」（Induction）。
- ▶ 常使用個案訪談法，深入、仔細觀察受測者對談中的內容，並進行歸納。
 - ▶ EX:全球教養方式大不同（文化、家庭成員）
 - ▶ 成功者10大特質（守時、事前規劃、與上司溝通能力）



資料的來源

▶ 一手資料

▶ 問卷

▶ 訪談

▶ 實驗

▶ 二手資料

▶ 報刊與雜誌

▶ 網上文章

Google表單

- ▶ 封閉式題項 (Closed-ended Questions)：預設了有限答項的提問

性別 *

男

女

組別 *

科組

企組

其他...

戶籍地 *

北部

⋮ 中部

南部

東部

- ▶ 開放式題項 (Open-ended Questions)：沒有預設答項的提問

上學期最喜歡科目 *

簡答文字

.....

上學期最討厭科目 *

簡答文字

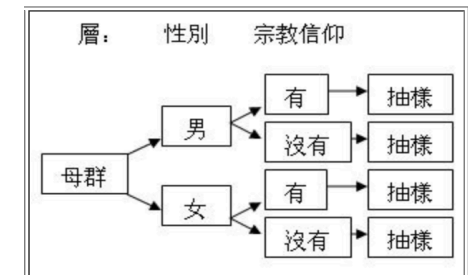
.....

抽樣方式

- ▶ 機率抽樣法 (probabilistic sampling methods)
 - ▶ 簡易隨機抽樣 (simple random sampling)
 - ▶ 系統抽樣 (systematic sampling)
 - ▶ 分層抽樣 (stratified sampling)
- ▶ 非機率抽樣法 (non-probabilistic sampling methods) :
 - ▶ 便利抽樣 (convenience sampling)
 - ▶ 配額抽樣 (quota sampling)
 - ▶ 滾雪球抽樣 (snowball sampling)
 - ▶ 立意抽樣/判斷抽樣 (purposeful sampling)

抽樣方式

- ▶ 機率抽樣法 (probabilistic sampling methods)：假設母體中，每個成員被抽選的機率是一樣的，並且可以透過抽樣來獲取相似的推論。最常見的包括簡易隨機抽樣、系統抽樣、分層抽樣。
 - ▶ 簡易隨機抽樣 (simple random sampling)
 - ▶ 隨機挑選路人填寫
 - ▶ 易受發卷者主觀偏好影響
 - ▶ 系統抽樣 (systematic sampling)
 - ▶ 每 x 個元素抽取樣本。例如：每五個走進7-11顧客為單位，則相關的樣本便是 1、6、11和 16
 - ▶ 分層抽樣 (stratified sampling)
 - ▶ 假設我們要解釋不同人對同性婚姻的的看法，需要抽樣檢驗。
 - ▶ 但是其可能會受性別與宗教影響



抽樣方式

- ▶ 非機率抽樣法 (non-probabilistic sampling methods)：在現實情況下中，機率抽樣並不一定可行。因此只能進行非機率抽樣。
 - ▶ 便利抽樣 (convenience sampling)
 - ▶ 考量時間與和成本下，抽選最容易接觸或聯絡的單位作樣本
 - ▶ 所得之樣本推論，不宜過份概括至更大的母群
 - ▶ 滾雪球抽樣 (snowball sampling)
 - ▶ 由 A 傳給 5 人, 這 5 人在各自傳給 5 個人.....
 - ▶ 立意抽樣 (purposeful sampling)
 - ▶ 根據研究者對母群的認識，去抽選他認為最具代表性的樣本。

Data clean

- ▶ 需先檢測是否有outliner, 亂填的問卷
- ▶ 檢測樣本是否符合iid假設 (後面章節會在教)

資料的型態與使用的檢定

$$Y = X_1 + X_2 + X_3 \dots$$

工作類型 = 性別

自變數(X) /依變數(Y)	類別	連續變數
類別	卡方檢定	多變量分析 (本課程不會學到)
連續變數	t 檢定(兩組) ANOVA(兩組以上)	相關分析 迴歸分析

成績 = 性別 (男生/女生) → t 檢定
成績 = 組別 (企/科/ 其他) → ANOVA

成績 = 性別 + 玩FB時間 + 微積分成績

資料的型態與使用的檢定

成績 = 性別 (男生/女生) → t 檢定

女生	男生
58	80
86	66
70	67
73	77
98	77
93	92
96	76
92	38
92	88
99	76
20	28
98	99
98	84
95	98
93	70
74	83
55	95
97	93
74	96
84	81
96	86
99	56
51	96
90	98
64	97
92	72

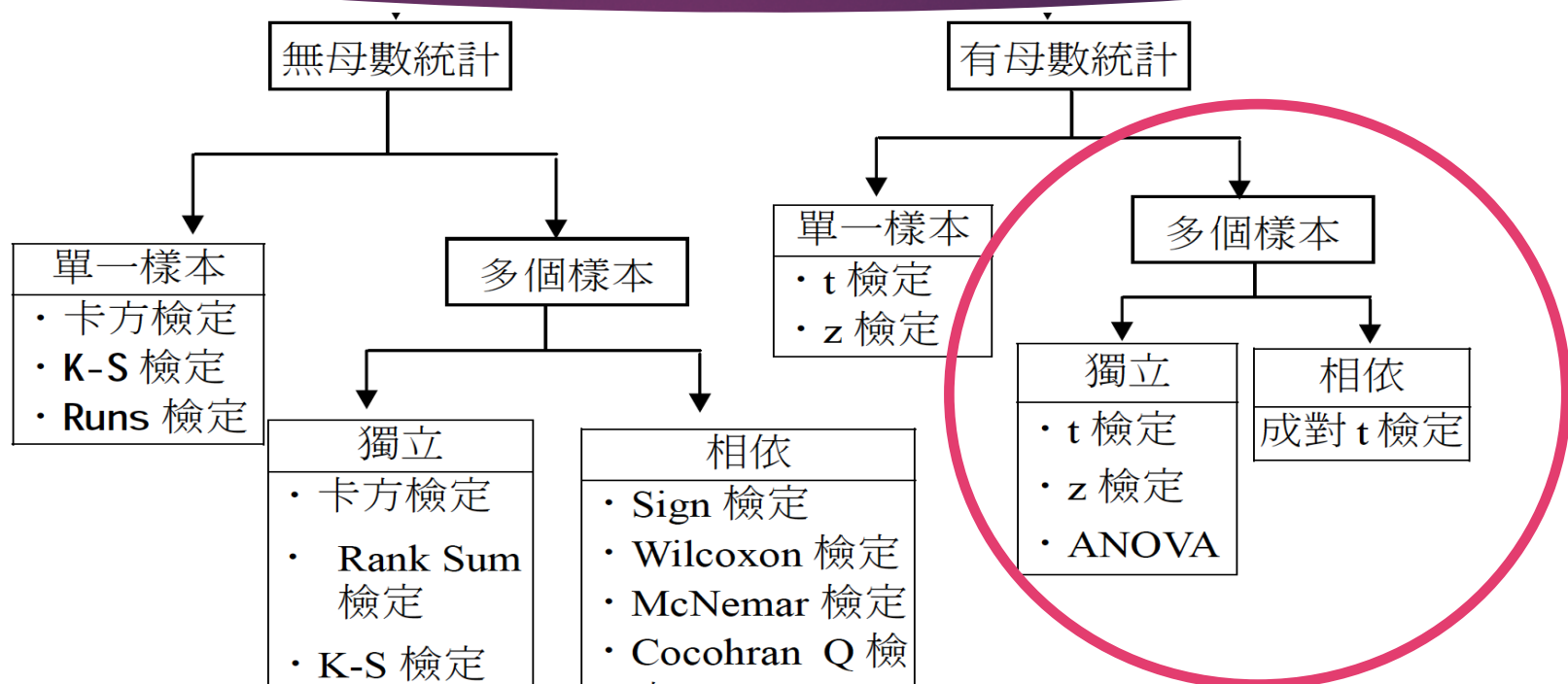
肝指數 = 組別 (drug1,2,3,4) → one factor ANOVA

成績 = 性別 + 教育程度 (國/高/ 大學) → two factors ANOVA

Group	Drug 1	Drug 2	Drug 3	Drug 4
1	6.6	12.6	2.7	8.7
2	7.1	3.5	2.4	9.3
3	7.5	4.4	6.5	10
4	9.9	7.5	16.2	12.6
5	13.8	6.4	8.3	10.6
6	13.9	13.5	5.4	15.4
7	15.9	16.9	15.4	16.3
8	14.3	11.4	17.1	18.9
9	16	16.9	7.7	13.7
10	16.3	14.8	16.1	19.4
11	14.6	18.6	9	18.5
12	18.7	21.2	24.3	21.1
13	17.3	10	9.3	19.3
14	19.6	17	19.2	21.9
15	20.7	21	18.7	22.1
16	18.4	27.2	18.9	19.4
17	21.5	26.8	7.9	25.4
18	20.4	28	23.8	26.5
19	21.9	31.7	8.8	22.2
20	22.5	11.9	26.7	23.5
21	21.5	28.7	25.2	19.6
22	25.2	29.5	27.3	30.1
23	23	22.2	17.6	26.6
24	23.7	19.5	25.6	24.5
25	28.4	31.2	26.1	27.4

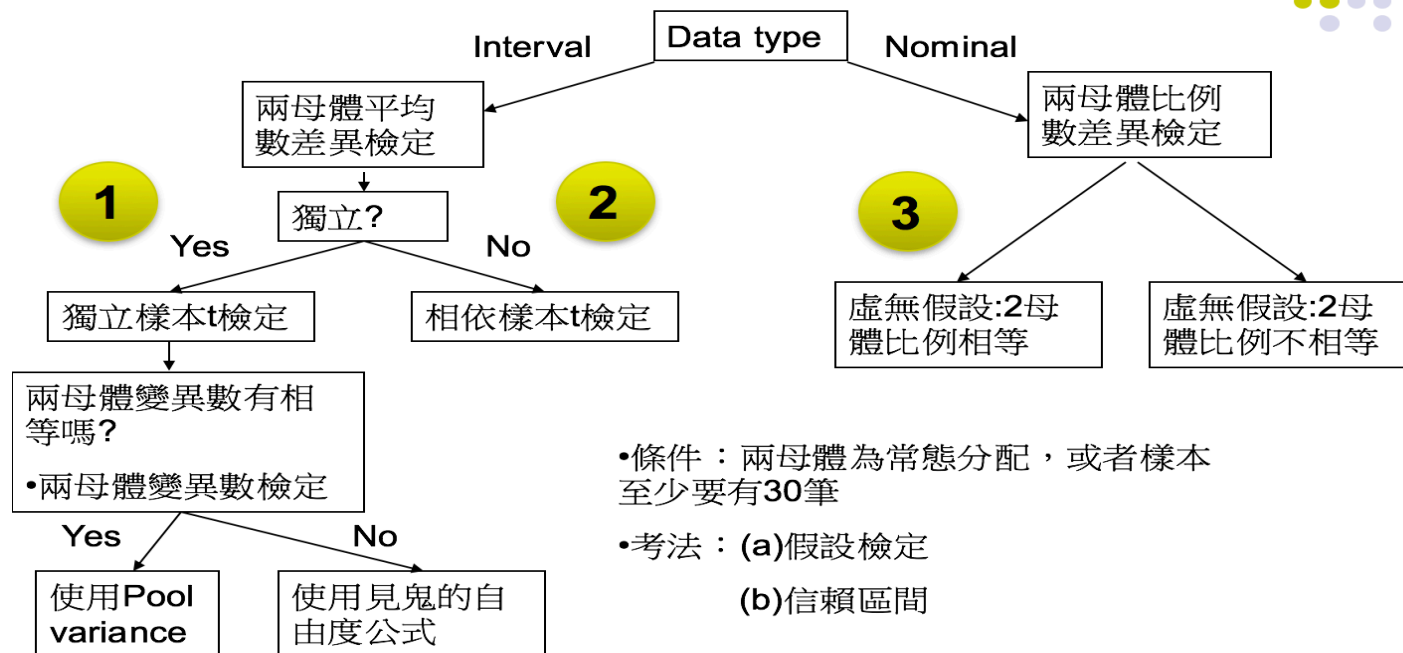
	Male	Female
Less than high school	10	7
	9	13
	12	14
	16	6
	14	11
	17	14
	13	13
	9	11
	11	14
	15	12
High School	12	7
	11	12
	9	6
	14	15
	12	10
	16	13
	10	9
	10	15
	5	12
	11	13
Less than Bachelor's	15	5

推論統計



推論統計 (two population)

Summary



•條件：兩母體為常態分配，或者樣本至少要有30筆

•考法：(a)假設檢定
(b)信賴區間

Hw3 目標

- ▶ 台大工管系歷年來統計課程，都會有三個族群來選修：企組、科組、與外系生。管院院長想知道學生學習的狀況，於是隨機抽選100位同學來做調查。假設各位都是受過專業訓練的統計分析師，請針對下列院長的疑問，提出相關的數據分析（假設資料皆為常態分配，請以 $\alpha=0.05$ 來做相關計算）
- ▶ 請四人一組，下週四繳交（請以紙本繳交，需貼上報表與寫上檢定結果）

EX1:獨立樣本T檢定 (Equal Variance)

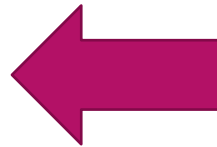
- ▶ Q3: 某家長很討厭自己小孩花時間在社團上，覺得參加社團會影響學業表現。
- ▶ 郭院長告訴家長，參加社團可以增加學生的人脈、團隊合作精神、培養終生學習的技能，對他們未來就業是有幫助的。請使用excel檔案來檢驗，學生有無參加社團，對統計第二次成績是沒有差異的。

Data analysis: F-test two sample for variance

Data analysis: t-test: Two sample assuming equal variance

Data analysis: t-test: Two sample assuming unequal variance

yes	no
77	85
86	81
65	58
67	41
84	82
75	76
65	76
49	68
72	50
74	85
77	75
73	72
67	72
69	50
80	84
29	74
70	51
77	41
52	50
70	53
64	48
73	78
39	75
62	76
32	73



編號	是否有參加社團	第二次統計
1	No	85
2	No	81
3	No	58
4	No	41
5	No	82
6	No	76
7	No	76
8	No	68
9	Yes	77
10	Yes	86
11	Yes	65
12	Yes	67
13	No	50
14	No	85
15	No	75
16	Yes	84
17	Yes	75
18	No	72
19	No	72
20	Yes	65
21	Yes	49
22	Yes	72
23	No	50
24	No	84
25	Yes	74
26	Yes	77

EX1 (Q3)



$$H_0: \frac{\sigma_1}{\sigma_2} = 0, H_1: \frac{\sigma_1}{\sigma_2} > 0$$

F-Test Two-Sample for Variances

	yes	no
Mean	67.88	66.08
Variance	156.72	217.6669388
Observations	50	50
df	49	49
F	0.7199991	
P(F<=f) one-tail	0.126857026	
F Critical one-tail	0.622165468	

Since P-value > 0.05, don't reject Ho. Hence using equal variance.

$$H_0: u_1 - u_2 = 0, H_1: u_1 - u_2 \neq 0$$

t-Test: Two-Sample Assuming Equal Variances

	yes	no
Mean	67.88	66.08
Variance	156.72	217.6669388
Observations	50	50
Pooled Variance	187.1934694	
Hypothesized Mean Difference	0	
df	98	
t Stat	0.657804989	
P(T<=t) one-tail	0.256102821	
t Critical one-tail	1.660551217	
P(T<=t) two-tail	0.512205642	
t Critical two-tail	1.984467455	

Since two tailed test the P-value > 0.05, Don't reject Ho. 因此在 95%顯著水準下，沒有足夠證據推翻 Ho. 表示統計第二次考試，不受有無參加社團影響。

EX2: 獨立樣本 T 檢定 (Unequal Variance)

- ▶ Q4. 過去微積分老師都告訴同學，如果微積分修不好，將來其他數量課程（例如：統計）也會修不好。
- ▶ 請以excel檔案內微積分成績來做檢測老師的宣稱。請檢驗微積分成績高於 B + 同學（包含B+），其統計第二次成機會比較好。

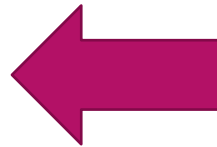


Data analysis: F-test two sample for variance

Data analysis: t-test: Two sample assuming equal variance

Data analysis: t-test: Two sample assuming unequal variance

微積分>B+	微積分<B+
75	41
85	67
82	75
65	29
84	75
75	55
72	39
65	83
72	52
74	60
69	50
68	58
78	67
68	78
68	75
86	76
80	80
60	52
74	63
85	49
81	62
84	13
48	58
70	55
76	50



編號	微積分成績	第二次統計
1	A	85
2	A-	81
3	B-	58
4	B	41
5	A	82
6	A+	76
7	A+	76
8	A+	68
9	B+	77
10	F	86
11	A	65
12	B	67
13	A+	50
14	A+	85
15	B	75
16	A	84
17	A	75
18	A+	72
19	A	72
20	A	65
21	C-	49
22	A	72
23	F	50
24	A-	84
25	A	74
26	A+	77

EX2 (Q4)

$$H_0: \frac{\sigma_1}{\sigma_2} = 0, H_1: \frac{\sigma_1}{\sigma_2} > 0$$

F-Test Two-Sample for Variances

	微積分>B+	微積分<B+
Mean	71.1875	59.5
Variance	103.4880952	250.2571429
Observations	64	36
df	63	35
F	0.413527039	
P(F<=f) one-tail	0.00112764	
F Critical one-tail	0.621464639	

Since P-value < 0.05, Reject Ho. Hence using unequal variance.

$$H_0: u_1 - u_2 = 0, H_1: u_1 - u_2 > 0$$

t-Test: Two-Sample Assuming Unequal Variances

	微積分>B+	微積分<B+
Mean	71.1875	59.5
Variance	103.4880952	250.2571429
Observations	64	36
Hypothesized Mean Difference	0	
df	52	
t Stat	3.992702698	
P(T<=t) one-tail	0.000103028	
t Critical one-tail	1.674689154	
P(T<=t) two-tail	0.000206057	
t Critical two-tail	2.006646805	

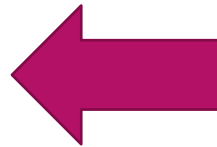
Since one tailed test the P-value < 0.05, Reject Ho. 因此在 95% 顯著水準下，有足夠證據推翻 Ho. 表示微積分成績高於 B+ 的同學，其相對統計成績高於 B+ 以下的同學。

EX3: 相依樣本 T 檢定 (Paired T test)

- ▶ Q8: 郭院長認為，因為同學是第一次修課，所以有很多細節與訣竅尚未熟悉。
- ▶ 院長認為學生經過三個月的課程後，應該比較能適應老師的教法。請檢測女生考試成績有明顯進步

Data analysis: t-test: Two sample assuming equal variance

第一次統計	第二次統計
58	58
86	82
70	68
73	86
98	67
93	75
96	75
92	72
92	72
99	65
20	49
98	72
98	77
95	51
93	67
74	69
55	53
97	78
74	80
84	29
96	70
99	77
51	52
90	76
64	73
92	73
79	55
77	63
83	62
56	32
72	61
54	75



編號	性別	組別	第一次統計	第二次統計
1	男	企組	80	85
2	男	生科	66	81
3	女	企組	58	58
4	男	企組	67	41
5	女	資工四	86	82
6	男	外文系	77	76
7	男	外文系	77	76
8	女	國企	70	68
9	男	企組	92	77
10	女	政治公共行政組	73	86
11	男	森林系	76	65
12	女	企組	98	67
13	男	企組	38	50
14	男	企組	88	85
15	女	企組	93	75
16	男	企組	76	84
17	女	企組	96	75
18	女	企組	92	72
19	女	企組	92	72
20	女	企組	99	65
21	女	企組	20	49
22	女	企組	98	72
23	男	企組	28	50
24	男	企組	99	84
25	男	企組	84	74
26	女	企組	98	77
27	男	企組	98	74
28	女	企組	95	51
29	男	企組	70	41
30	男	企組	83	73
31	男	企組	95	50
32	女	企組	93	67
33	女	企組	74	69
34	女	企組	55	53

EX3 (Q8)

8-b

女生

$$H_0: u_d = 0, H_1: u_d < 0$$

t-Test: Paired Two Sample for Means

	第一次統計	第二次統計
Mean	82.54545455	67.69090909
Variance	287.6969697	150.810101
Observations	55	55
Pearson Correlation	0.485709124	
Hypothesized Mean Difference	0	
df	54	
t Stat	7.168597528	
P(T<=t) one-tail	1.09657E-09	
t Critical one-tail	1.673564906	
P(T<=t) two-tail	2.19314E-09	
t Critical two-tail	2.004879288	

Since one tailed test the P-value < 0.05, Reject H_0 . 因此在 95% 顯著水準下，有足夠證據推翻 H_0 。但是並非像院長說的，因為女生第二次成績明顯是退步的。

EX4: 比例樣本 Z 檢定 (Q9)

- ▶ 今天富邦總經理來學校與郭院長聊天，希望院長能推薦相關學生，提前進入富邦學習，以促進雙方實習的機會。除了成績考量外，院長覺得學生的性向與人生目標比成績更重要。
- ▶ 過去發現，企組畢業的學生，後來想成為企業創辦人（企管相關領域）的比例相對較高。院長想知道是否企組同學未來想成為企業創辦人（企管相關領域）的比例，明顯高於非企組比例？

EX4 (Q9)

$$H_0: P_1 - P_2 = 0, H_1: P - P_2 > 0$$



z-Test: Two Proportions

	非企組	企組
Sample Proportions	0.5227	0.4464
Observations	44	56
Hypothesized Difference	0	
z Stat	0.7581	
P(Z<=z) one tail	0.2242	
z Critical one-tail	1.6449	
P(Z<=z) two-tail	0.4484	
z Critical two-tail	1.96	

□

Since one tailed test the P-value >0.05, Don't reject Ho. 因此在 95%顯著水準下，沒有足夠證據推翻 Ho.。表示企組與非企組學生，未來想走企管領域行業比例，並無差異。