Diagnostic test

醫學研究部 生物統計小組
徐倩儀
2026/4/28

內容大綱

- Hosmer and Lemeshow test (goodness of fit)
- Bland–Altman plot
- Receiver Operating Characteristic (ROC) Curve Analysis

Hosmer and Lemeshow test

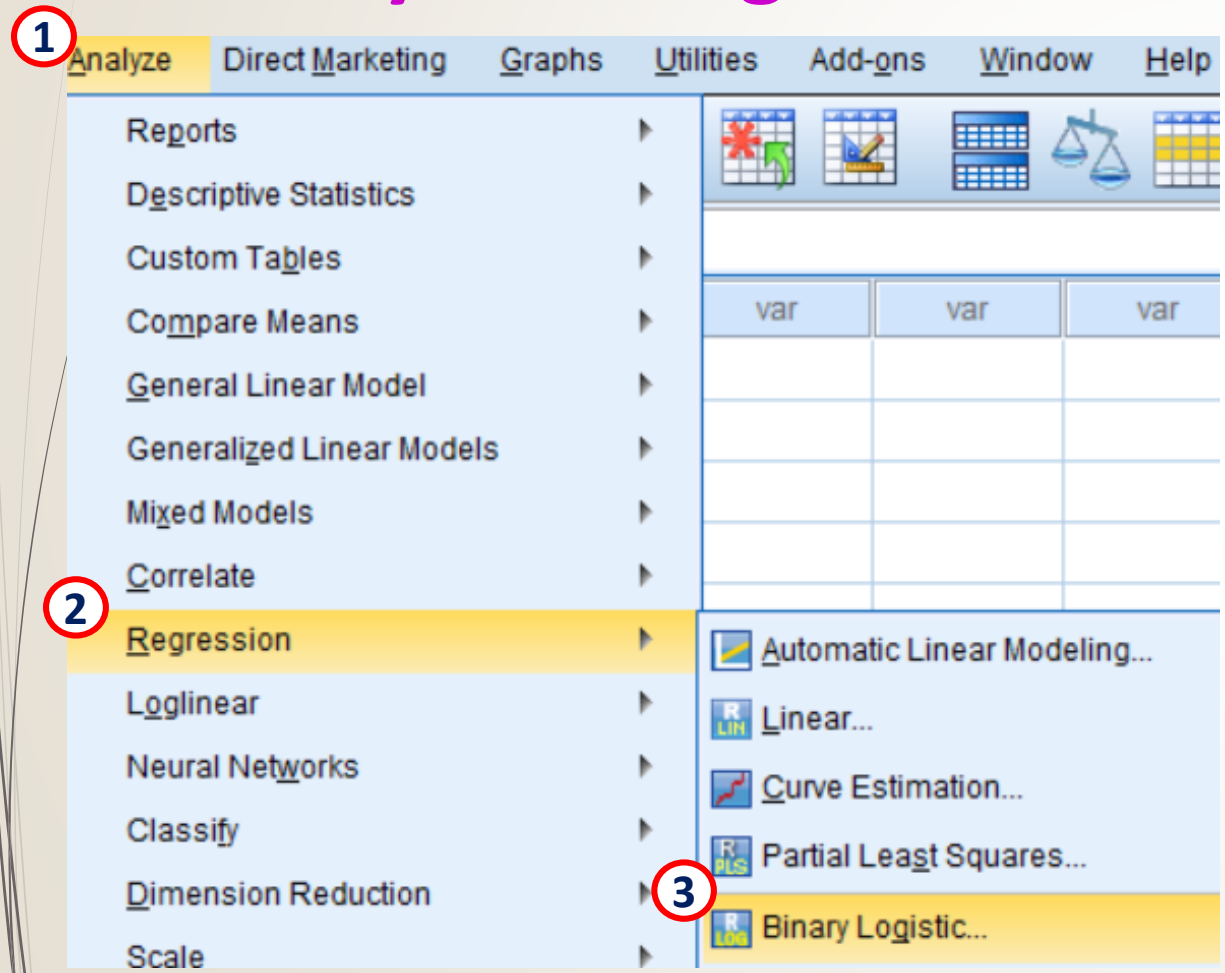
- 對logistic迴歸模型擬合優度的檢定方法
  - 根據預測機率值將資料分成大致相同規模的10個組
  - 將觀測資料按其預測機率做升序排列
    - 第一組機率最小
    - 最後一組估計機率最大

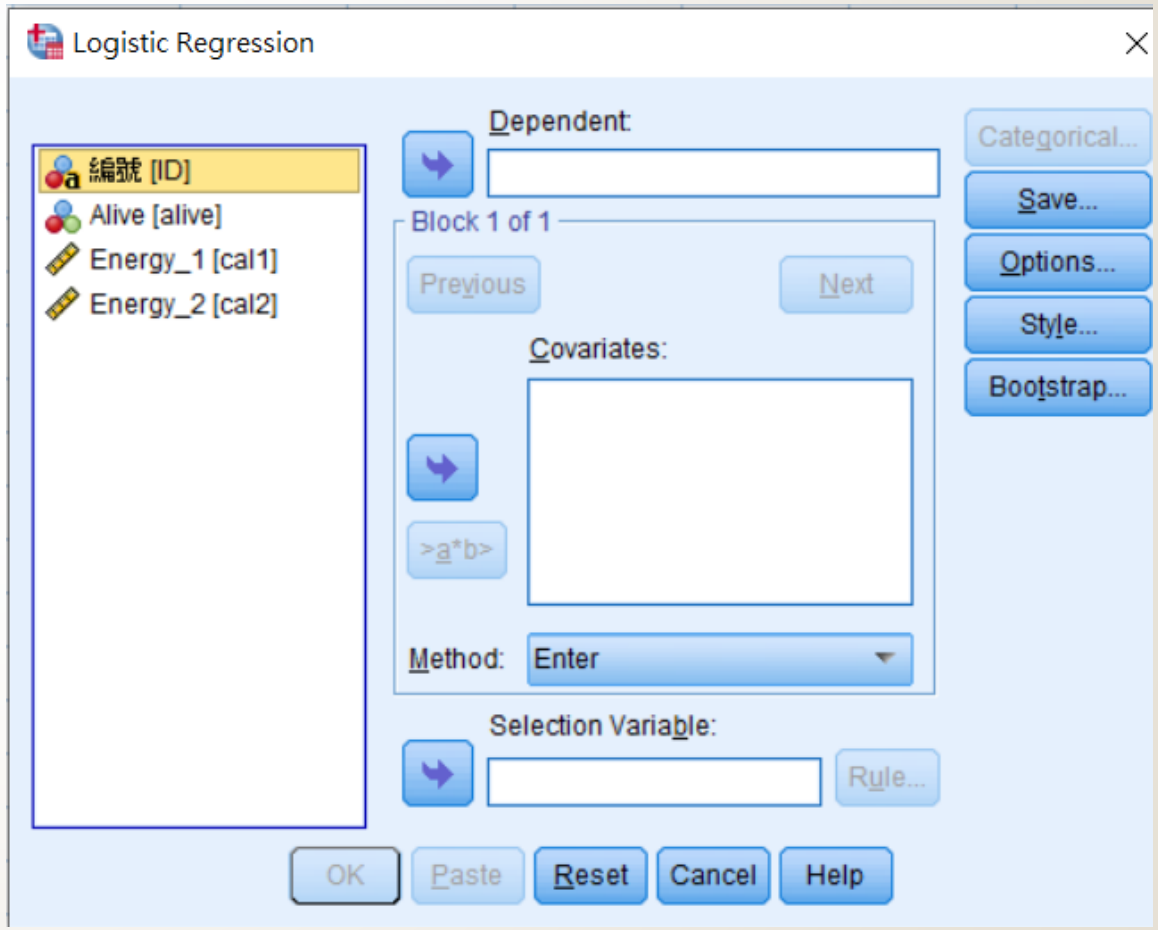
Hosmer and Lemeshow test

- 類似於Pearson Chi-Square統計量的指標
  - 可從觀察值和預測值構成的2*G交互表中求得
  - HL指標與Chi-Square分佈相比較
    - $p > 0.05$ (Acceptable Calibration)

Hosmer and Lemeshow test (SPSS)

➡ Analyze → Regression → Binary Logistic

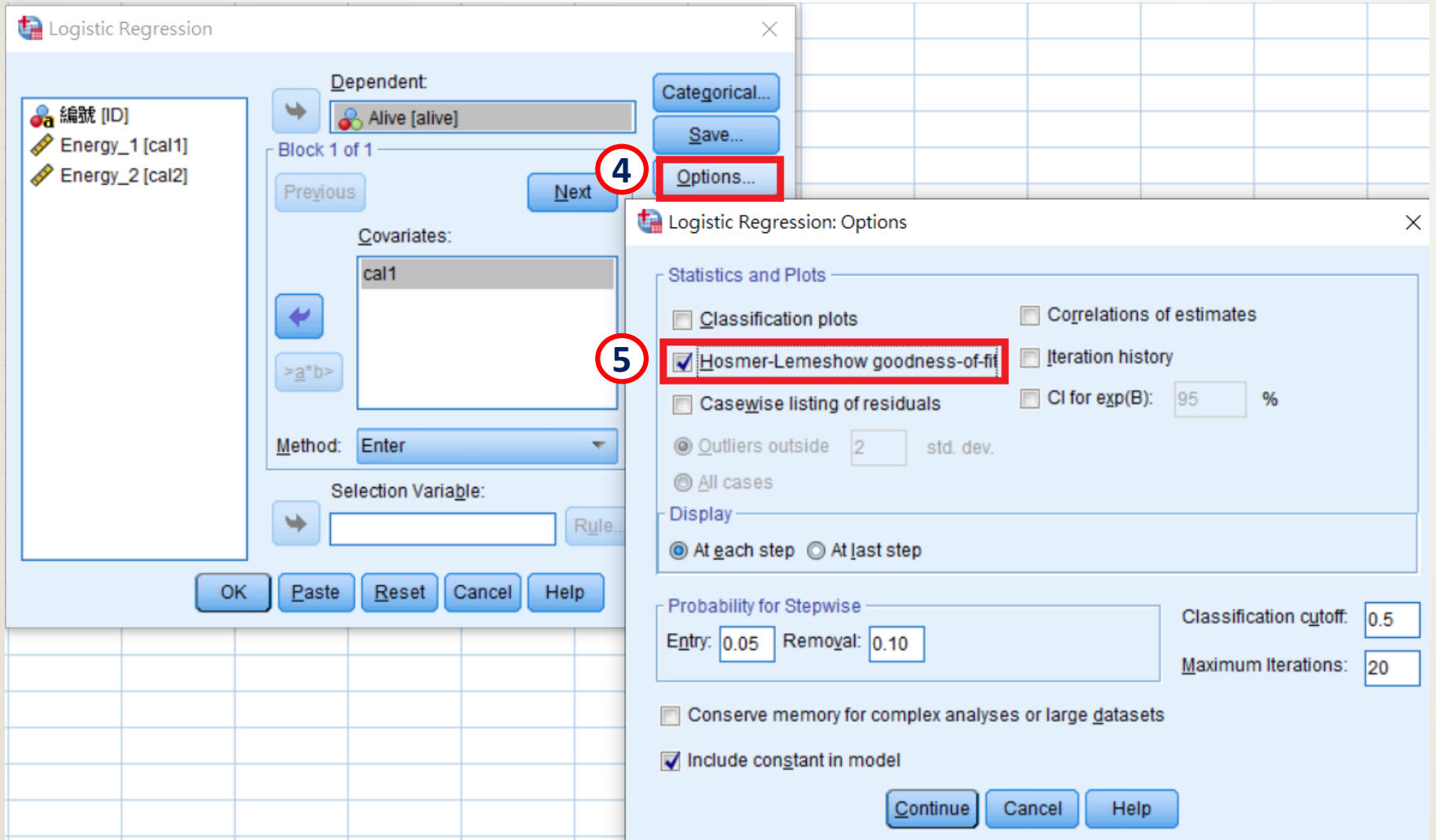




練習範例：ROC ex.sav

Hosmer and Lemeshow test (SPSS)

- Analyze → Regression → Binary Logistic



練習範例：ROC ex.sav

Hosmer and Lemeshow test (SPSS)

- Analyze → Regression → Binary Logistic

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	7.332	7	.395

***p* > 0.05**
Acceptable calibration

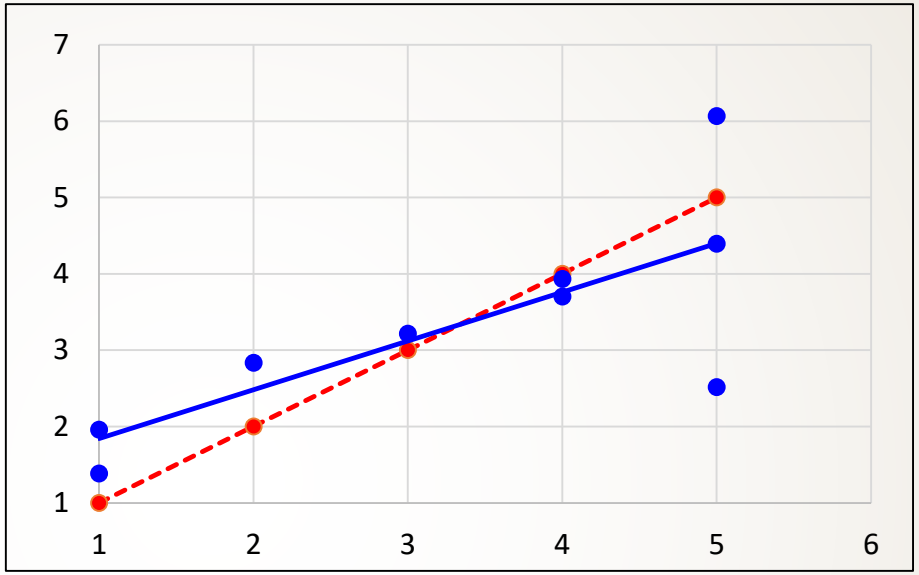
Contingency Table for Hosmer and Lemeshow Test

		Alive = Death		Alive = Alive		Total
		Observed	Expected	Observed	Expected	
Step 1	1	5	4.618	1	1.382	6
	2	5	4.041	1	1.959	6
	3	1	3.483	5	2.517	6
	4	4	3.165	2	2.835	6
	5	3	2.785	3	3.215	6
	6	2	2.298	4	3.702	6
	7	2	2.070	4	3.930	6
	8	1	1.606	5	4.394	6
	9	2	.935	5	6.065	7

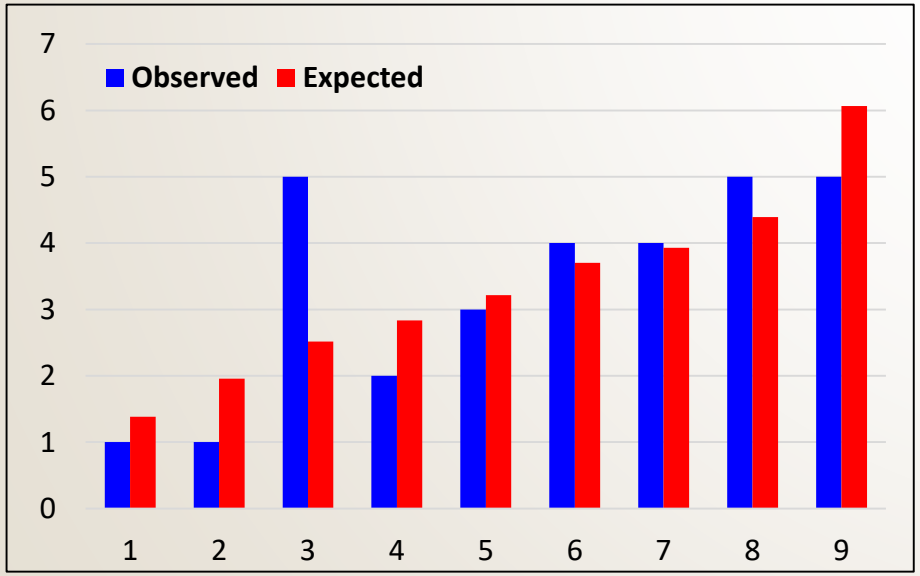
練習範例：ROC ex.sav

Hosmer and Lemeshow test (SPSS & Excel)

- Scatter plot
- Bar chart
- Line chart



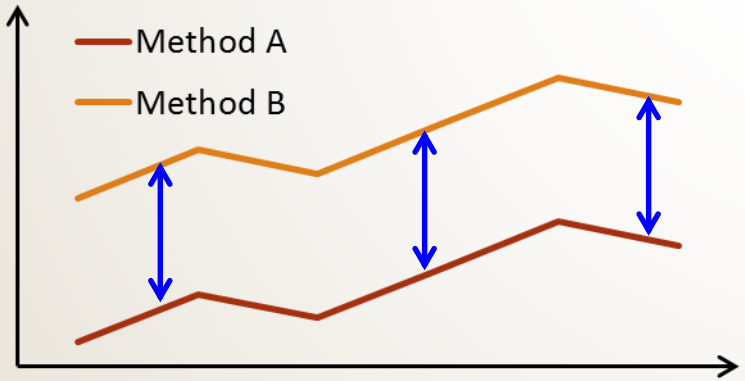
Alive = Alive	
Observed	Expected
1	1.382
1	1.959
5	2.517
2	2.835
3	3.215
4	3.702
4	3.930
5	4.394
5	6.065



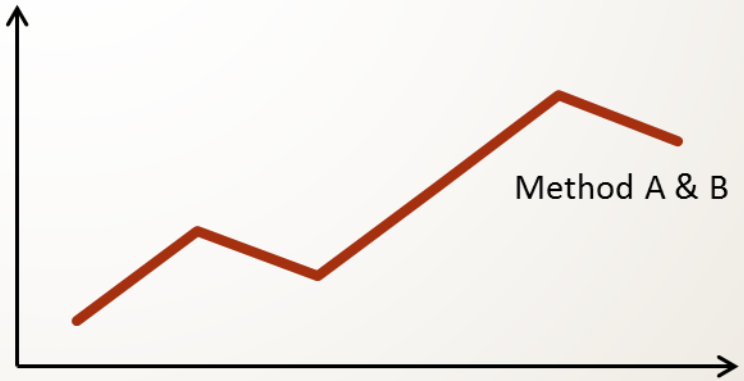
練習範例：ROC ex.sav

Bland–Altman plot

- The agreement between **two quantitative methods** of measurement
  - correlation coefficient (r)
    - the relationship between one variable and another
    - not the differences (not recommended)



r = 1.00



Bland–Altman plot

- The agreement between **two quantitative methods** of measurement
  - Bland–Altman plot
    - In 1983 Altman and Bland (B&A) proposed
    - **Mean difference** and **constructing limits of agreement**

Bland–Altman plot

- Scatter plot
  - Y axis
    - The difference between the two paired measurements (A-B)
    - Normally distributed (Gaussian) - 95% of differences d ± 1.96s
      - not normally distributed - logarithmic transformation
  - X axis
    - The average of these measures ((A+B)/2)

Bland–Altman plot

Y軸：兩種測量的差異(A-B)
顯示兩者之間的誤差大小

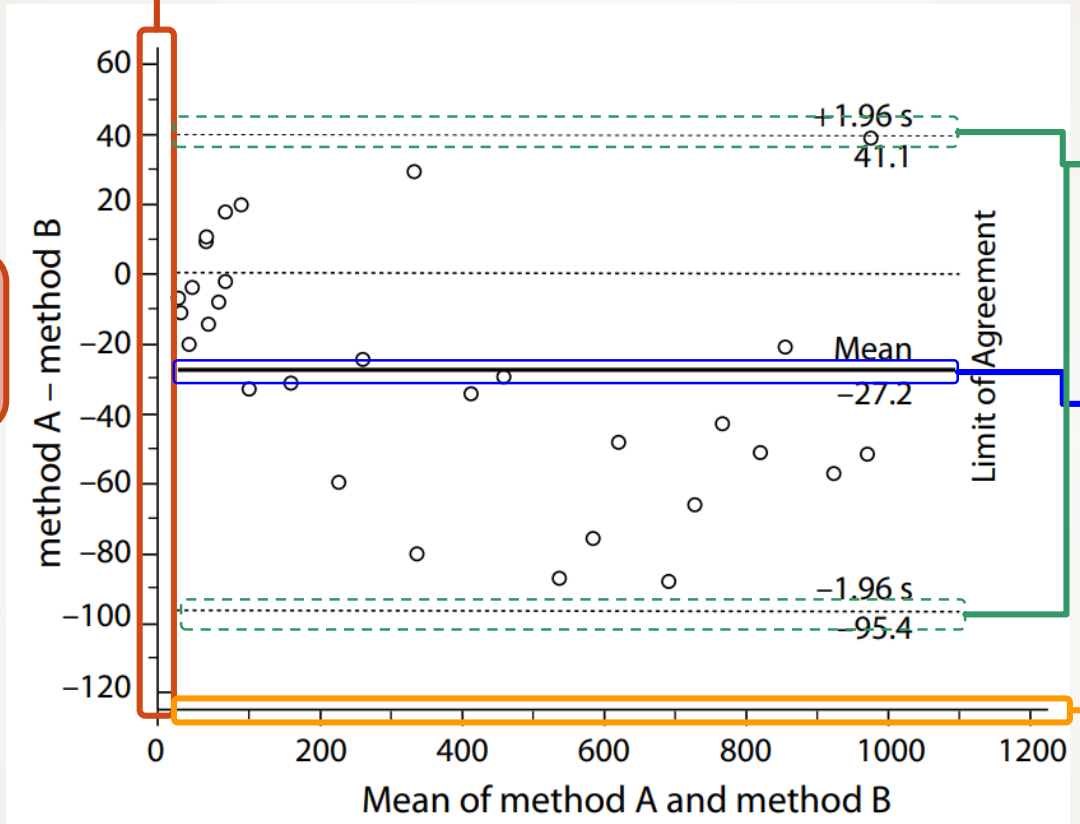


FIGURE 5. Bland and Altman plot for data from the table 1, with the representation of the limits of agreement (doted line), from -1.96*s* to +1.96*s*.

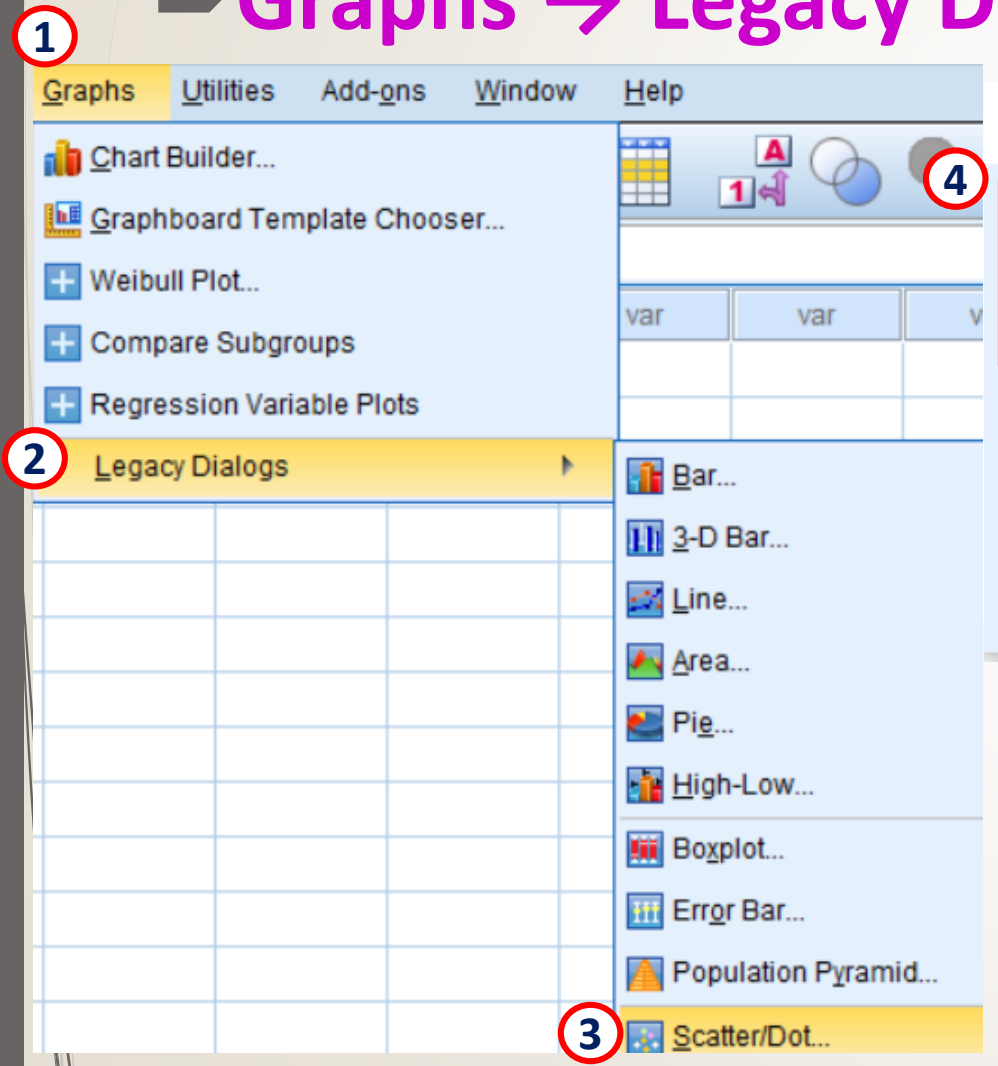
上下虛線：一致性界線
標記95%誤差分佈的範圍

中間實線：平均差值
兩種工具間的系統性偏差

X軸：兩種方法的平均值
作為真實數值的最佳估計

Bland–Altman plot (SPSS)

Graphs → Legacy Dialogs → Scatter/Dot



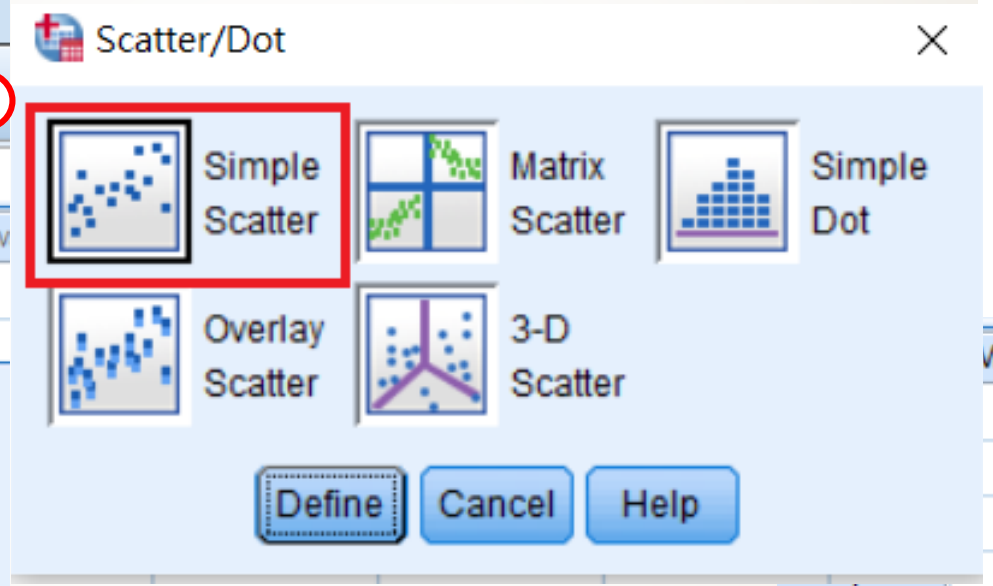
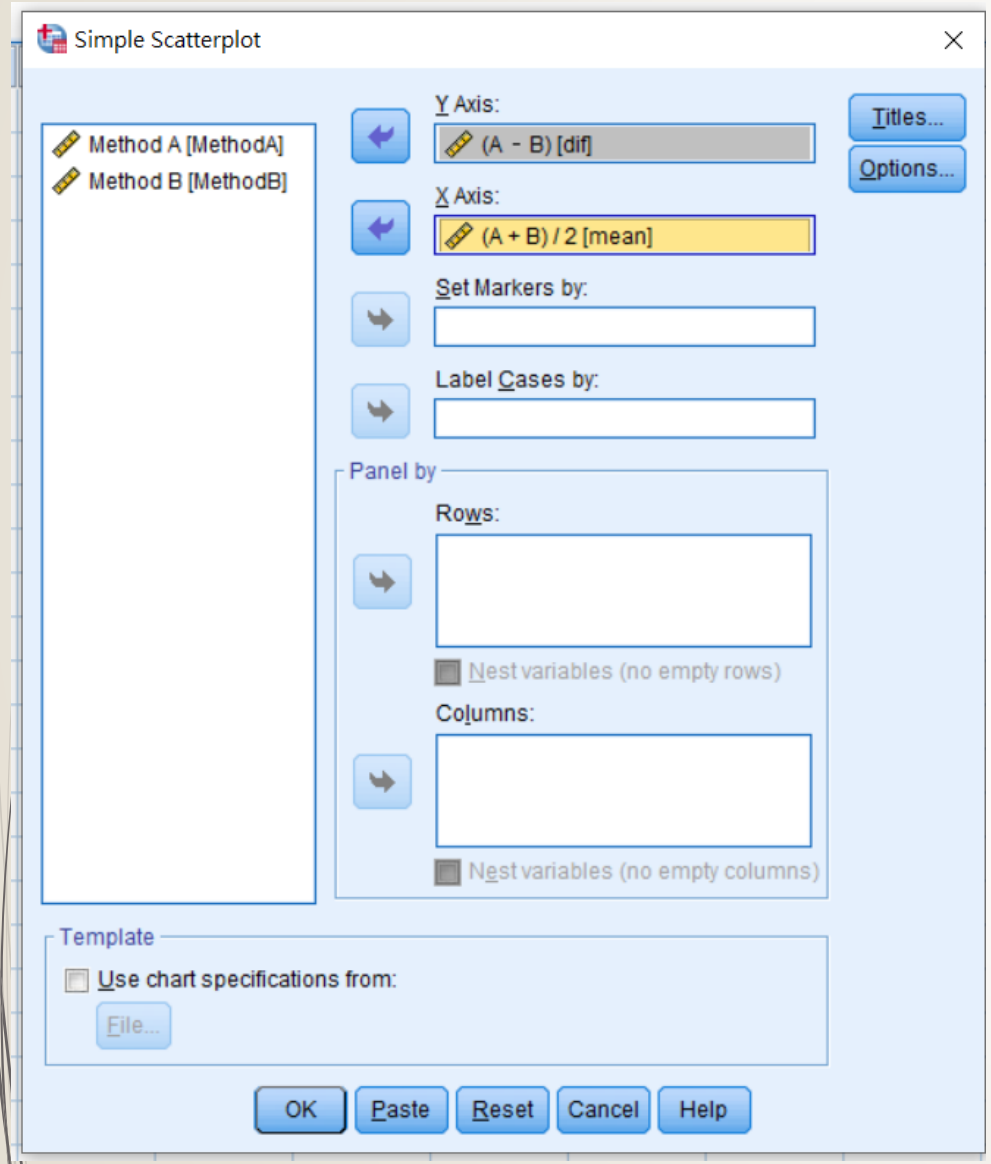


Table 1. Hypothetical data of an agreement between two methods (Method A and B).

Method A (units)	Method B (units)	Mean (A+B)/2 (units)	(A – B) (units)	(A – B)/ Mean (%)
1.0	8.0	4.5	-7.0	-155.6%
5.0	16.0	10.5	-11.0	-104.8%
10.0	30.0	20.0	-20.0	-100.0%
20.0	24.0	22.0	-4.0	-18.2%
50.0	39.0	44.5	11.0	24.7%
40.0	54.0	47.0	-14.0	-29.8%
50.0	40.0	45.0	10.0	22.2%
60.0	68.0	64.0	-8.0	-12.5%
70.0	72.0	71.0	-2.0	-2.8%
80.0	62.0	71.0	18.0	25.4%
90.0	122.0	106.0	-32.0	-30.2%
100.0	80.0	90.0	20.0	22.2%
150.0	181.0	165.5	-31.0	-18.7%
200.0	259.0	229.5	-59.0	-25.7%
250.0	275.0	262.5	-25.0	-9.5%

	MethodA	MethodB	mean	dif
	1	8	5	-7
	5	16	11	-11
	10	30	20	-20
	20	24	22	-4
5	50	39	45	11
6	40	54	47	-14
7	50	40	45	10
8	60	68	64	-8
9	70	72	71	-2
10	80	62	71	18

Bland–Altman plot (SPSS)



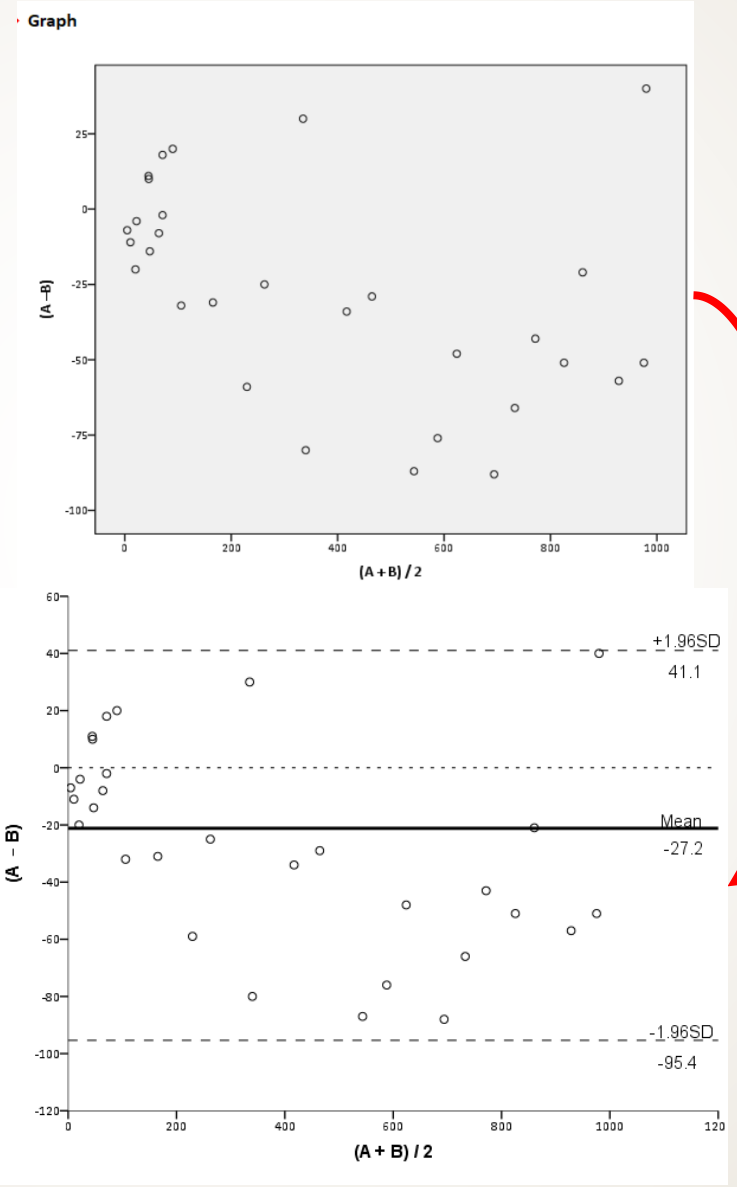
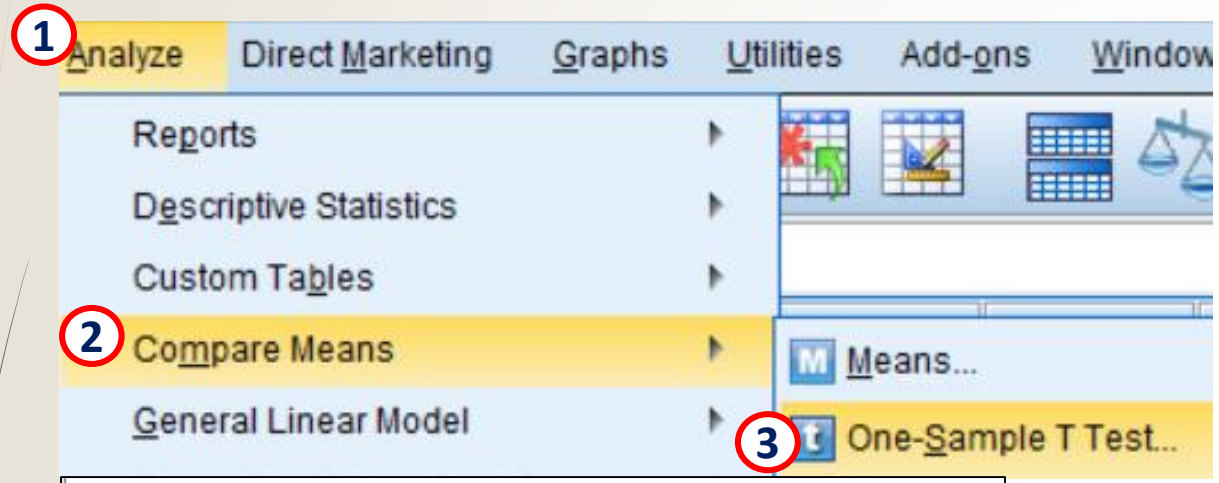


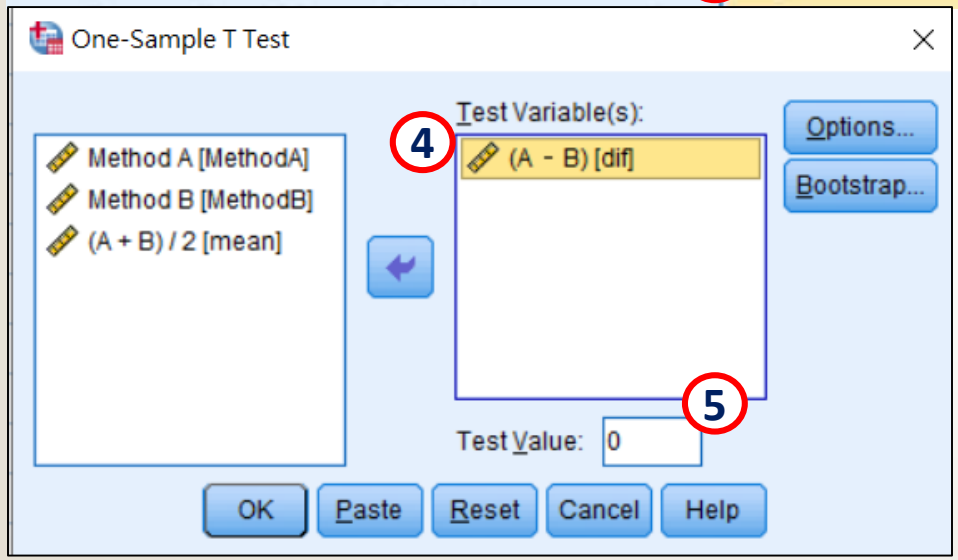
Table 1. Hypothetical data of an agreement between two methods (Method A and B).

Method A (units)	Method B (units)	Mean (A+B)/2 (units)	(A – B) (units)	(A – B)/ Mean (%)
1.0	8.0	4.5	-7.0	-155.6%
5.0	16.0	10.5	-11.0	-104.8%
10.0	30.0	20.0	-20.0	-100.0%
20.0	24.0	22.0	-4.0	-18.2%
50.0	39.0	44.5	11.0	24.7%
40.0	54.0	47.0	-14.0	-29.8%
50.0	40.0	45.0	10.0	22.2%
60.0	68.0	64.0	-8.0	-12.5%
70.0	72.0	71.0	-2.0	-2.8%
80.0	62.0	71.0	18.0	25.4%
90.0	122.0	106.0	-32.0	-30.2%
100.0	80.0	90.0	20.0	22.2%
150.0	181.0	165.5	-31.0	-18.7%
200.0	259.0	229.5	-59.0	-25.7%
250.0	275.0	262.5	-25.0	-9.5%
300.0	380.0	340.0	-80.0	-23.5%
350.0	320.0	335.0	30.0	9.0%
400.0	434.0	417.0	-34.0	-8.2%
450.0	479.0	464.5	-29.0	-6.2%
500.0	587.0	543.5	-87.0	-16.0%
550.0	626.0	588.0	-76.0	-12.9%
600.0	648.0	624.0	-48.0	-7.7%
650.0	738.0	694.0	-88.0	-12.7%
700.0	766.0	733.0	-66.0	-9.0%
750.0	793.0	771.5	-43.0	-5.6%
800.0	851.0	825.5	-51.0	-6.2%
850.0	871.0	860.5	-21.0	-2.4%
900.0	957.0	928.5	-57.0	-6.1%
950.0	1001.0	975.5	-51.0	-5.2%
1000.0	960.0	980.0	40.0	4.1%

Bland–Altman plot (SPSS)

Analyze → Compare Means → One-Sample T test





T-Test

One-Sample Statistics

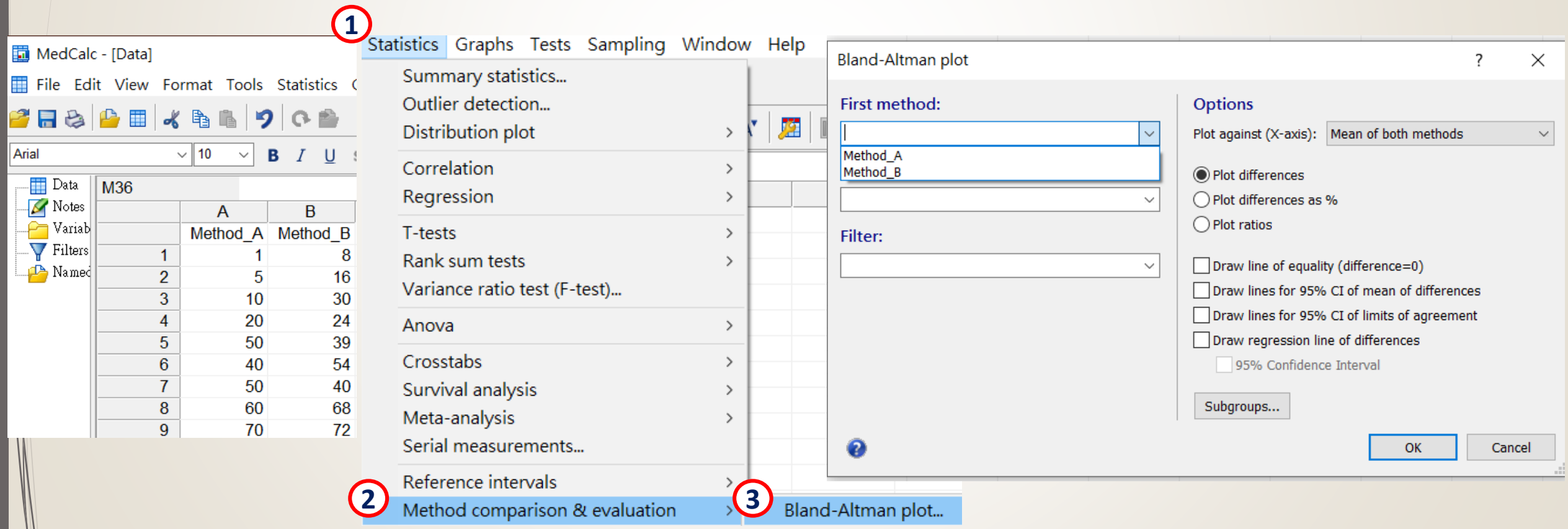
	N	Mean	Std. Deviation	Std. Error Mean
(A - B)	30	-27.17	34.806	6.355

One-Sample Test

	Test Value = 0					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
(A - B)	-4.275	29	.000	-27.167	-40.16	-14.17

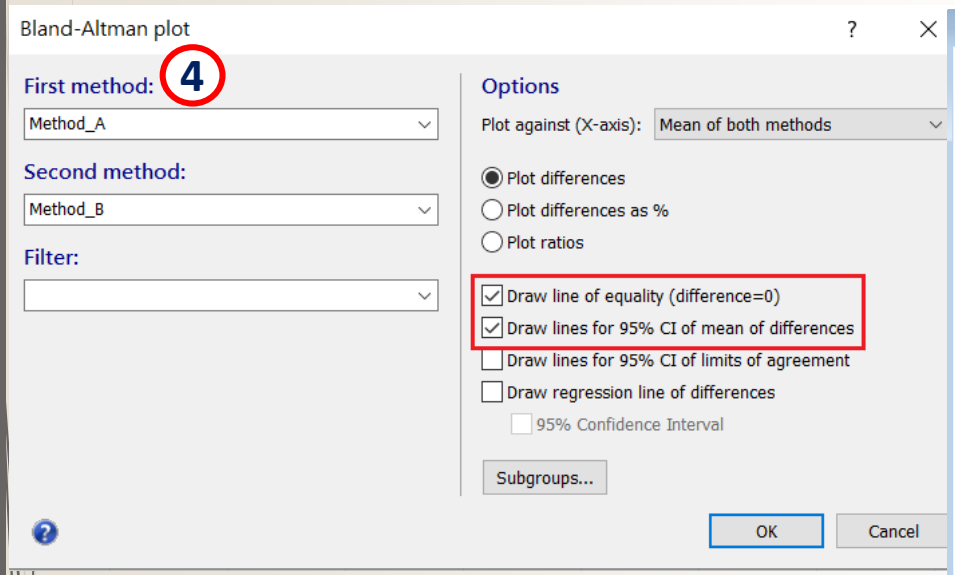
Bland–Altman plot (MedCalc)

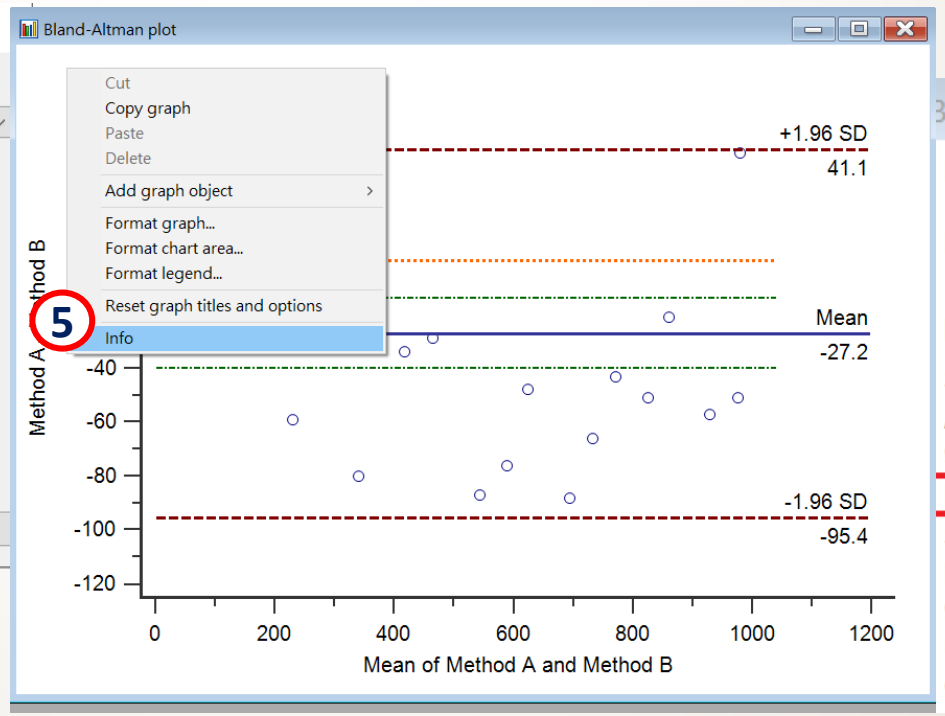
- Statistics → Method comparison & evaluation → Bland–Altman plot

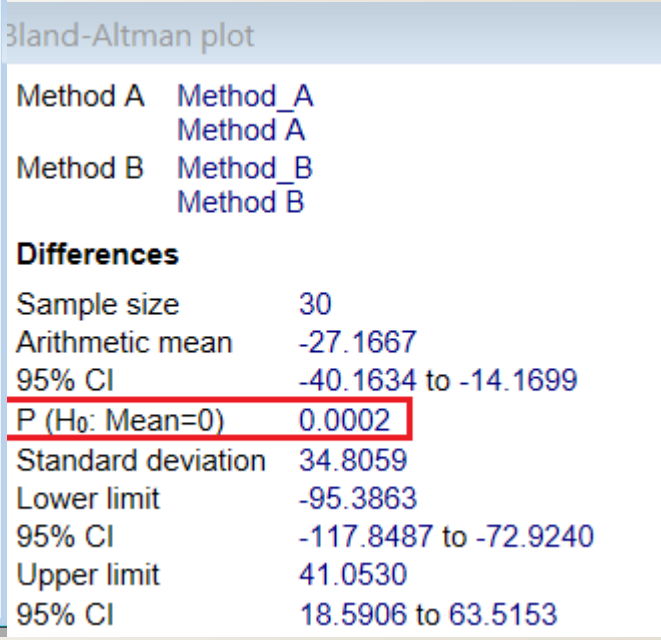


Bland–Altman plot (MedCalc)

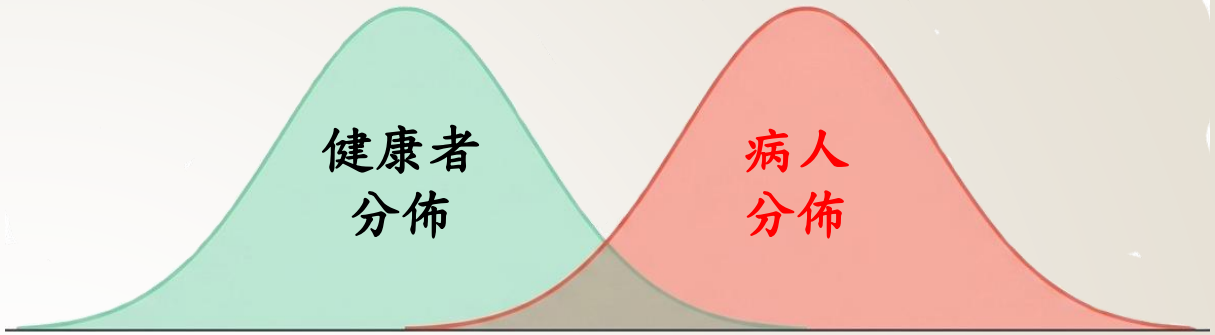
- Statistics → Method comparison & evaluation → Bland–Altman plot

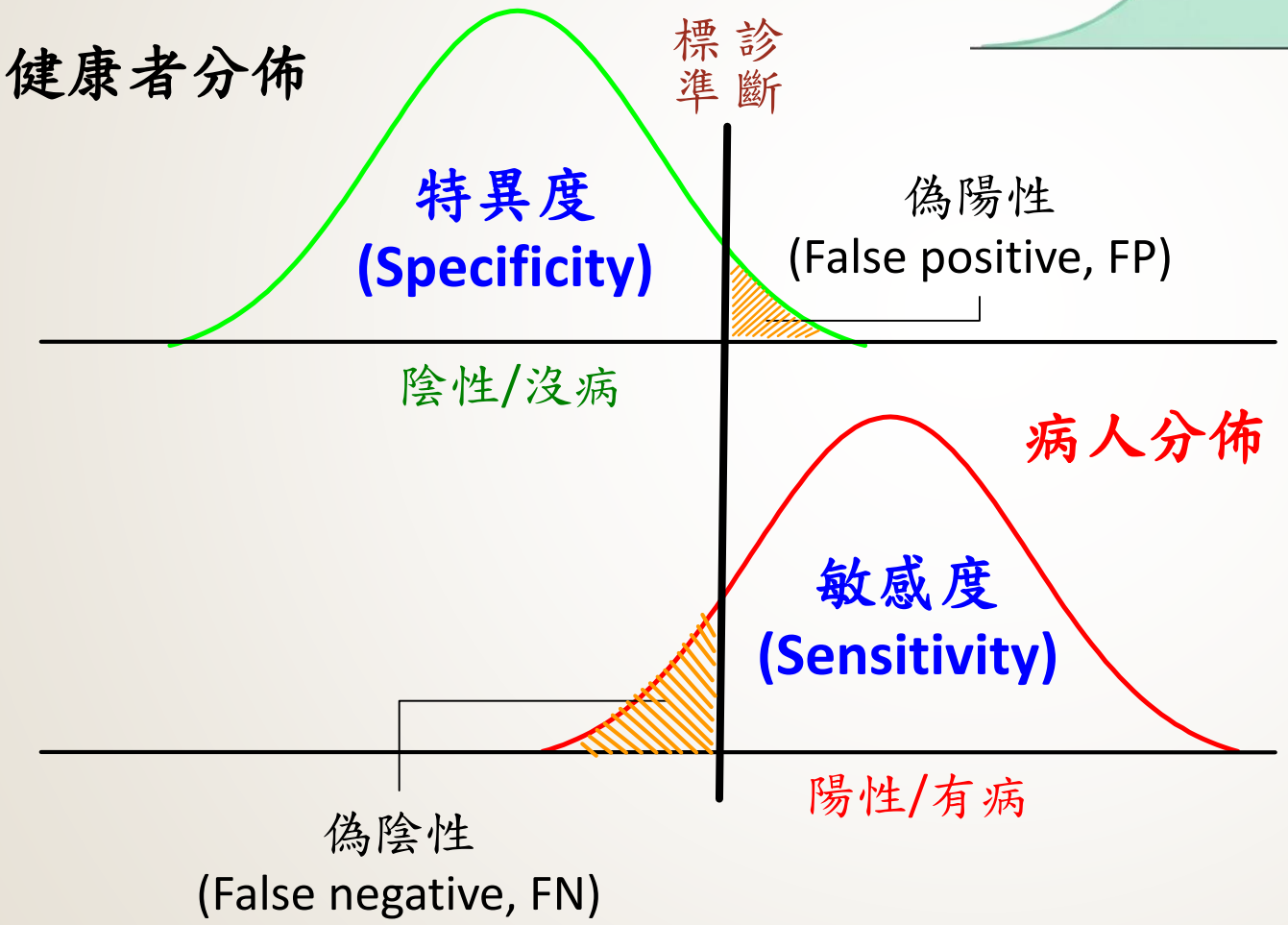




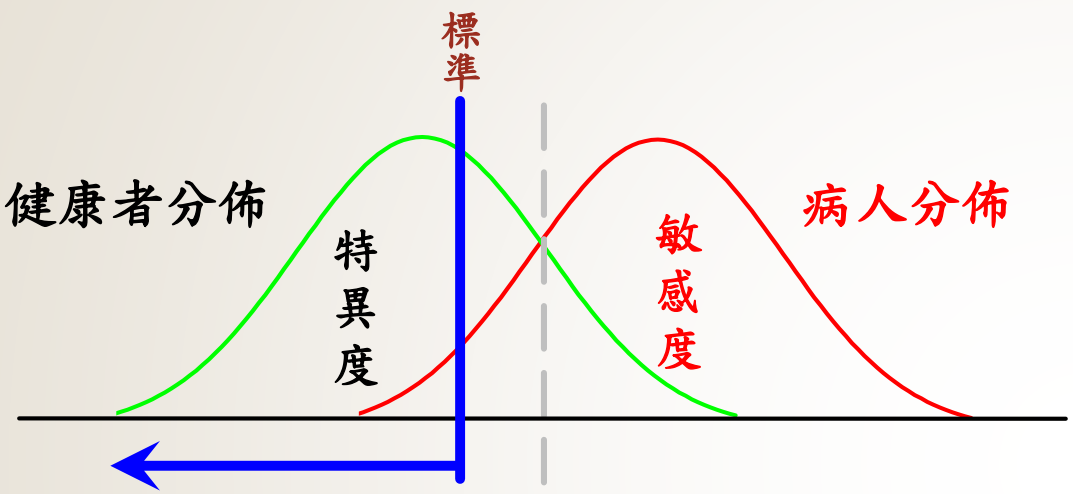


Diagnostic test





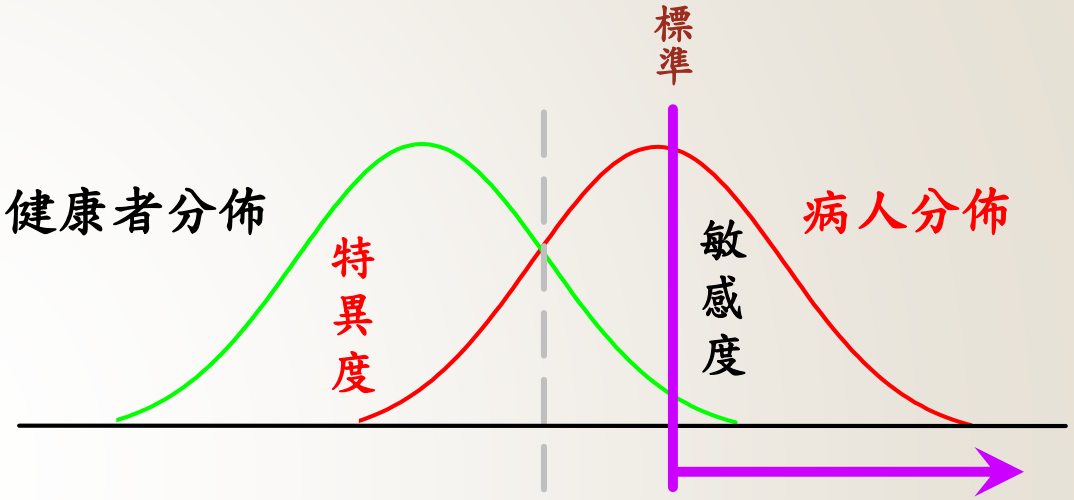
Diagnostic test



寬鬆

敏感度 (Sensitivity) 上升

特異度 (Specificity) 下降

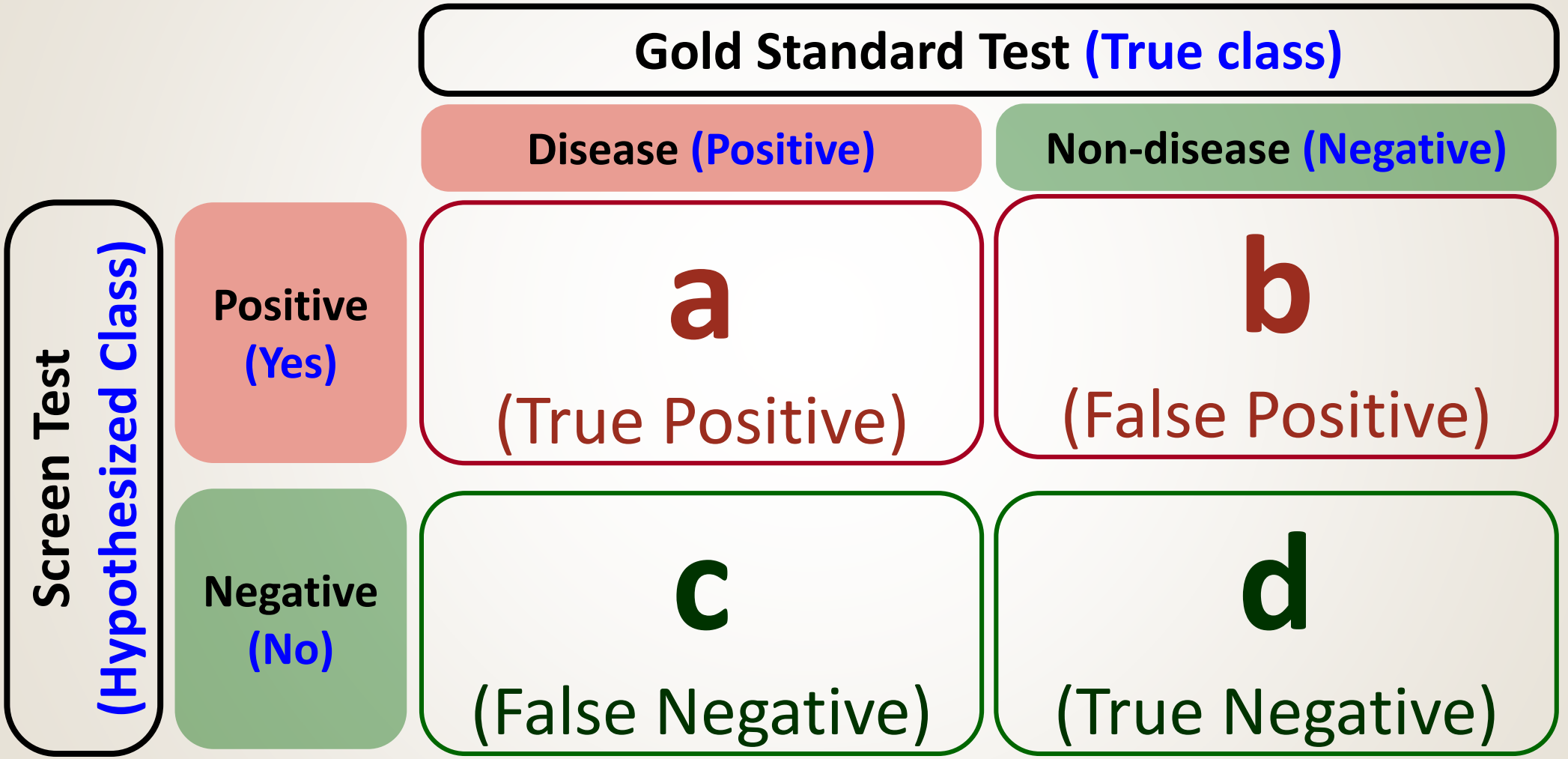


嚴格

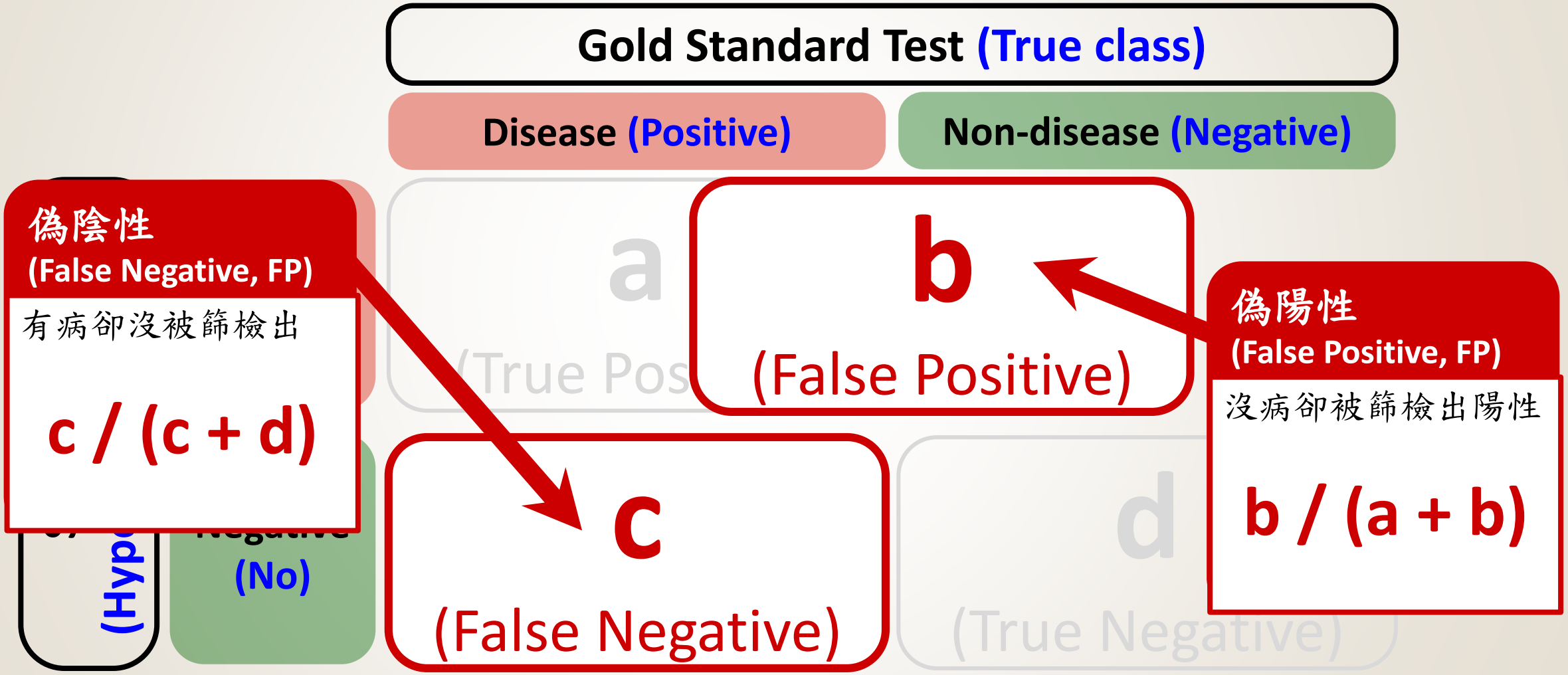
敏感度 (Sensitivity) 下降

特異度 (Specificity) 上升

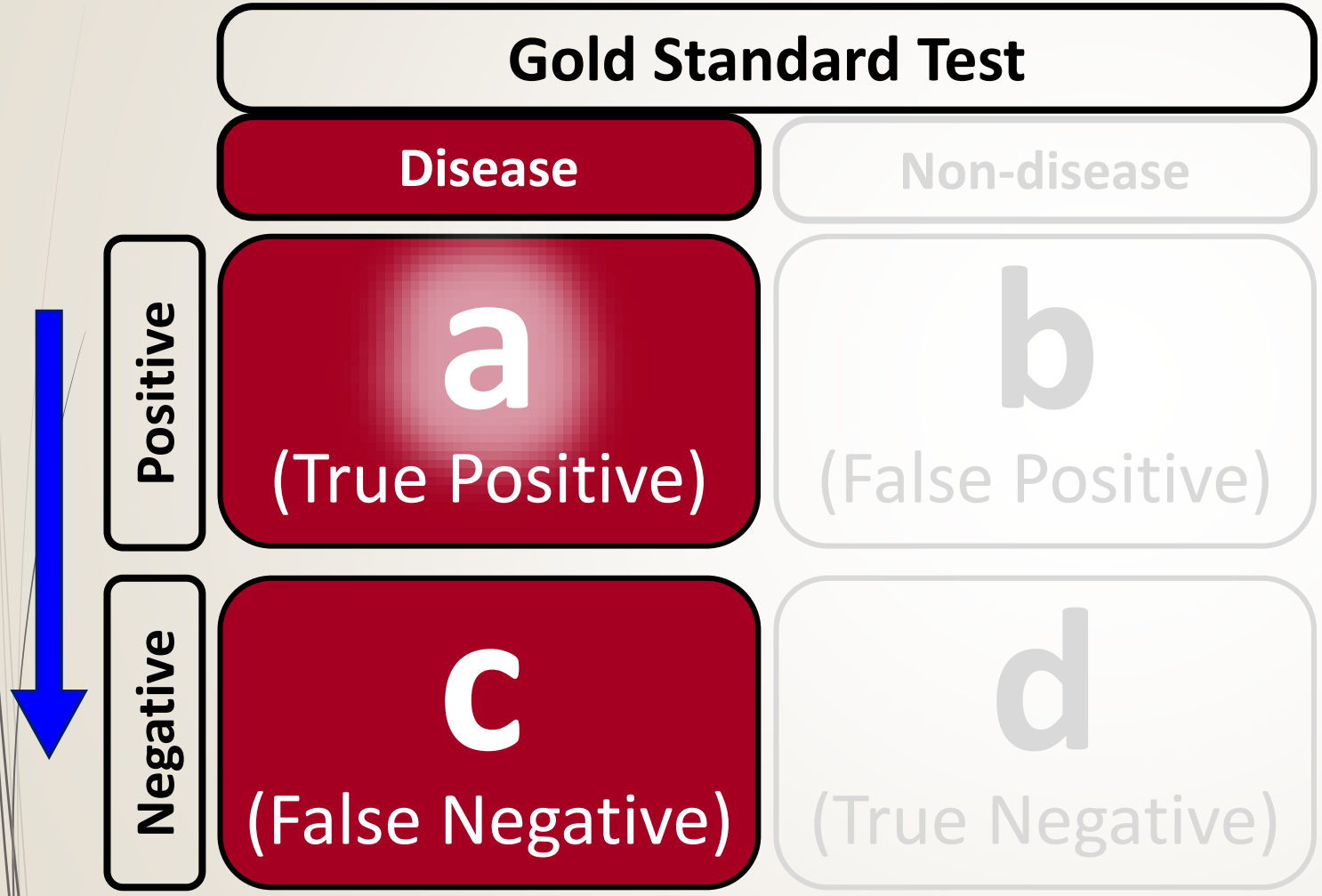
Diagnostic test



Diagnostic test

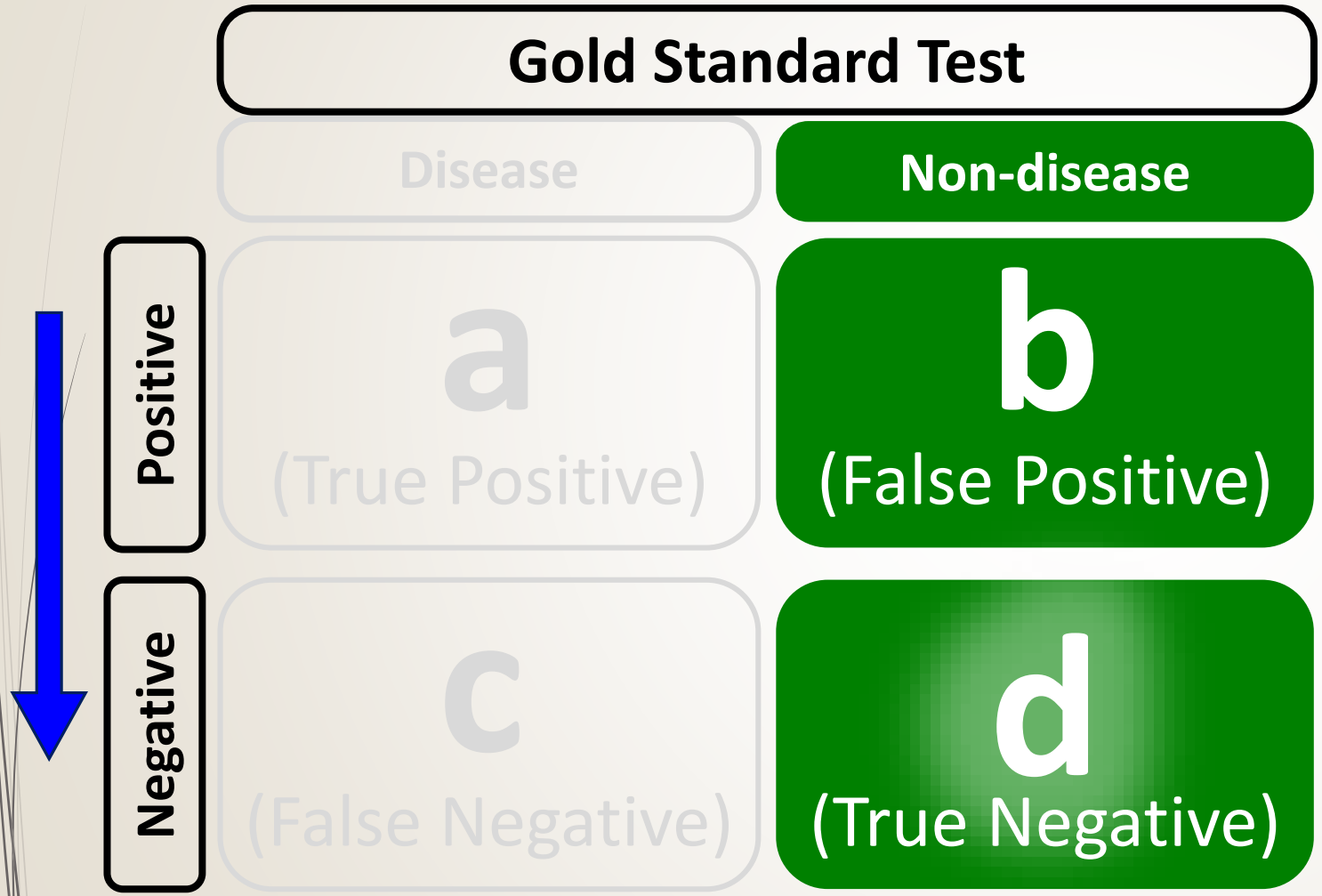


Diagnostic test



$$\text{Sensitivity (敏感度)} = \frac{a}{(a+c)}$$

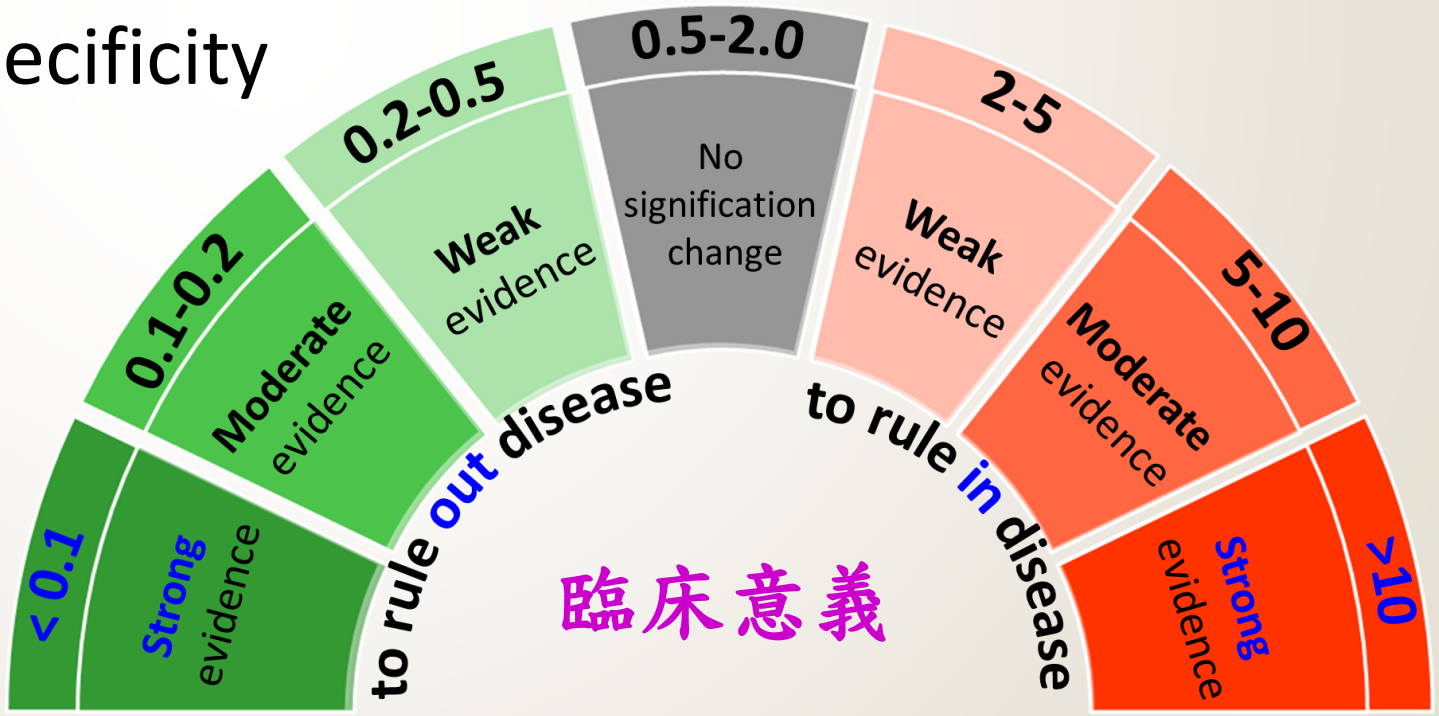
Diagnostic test



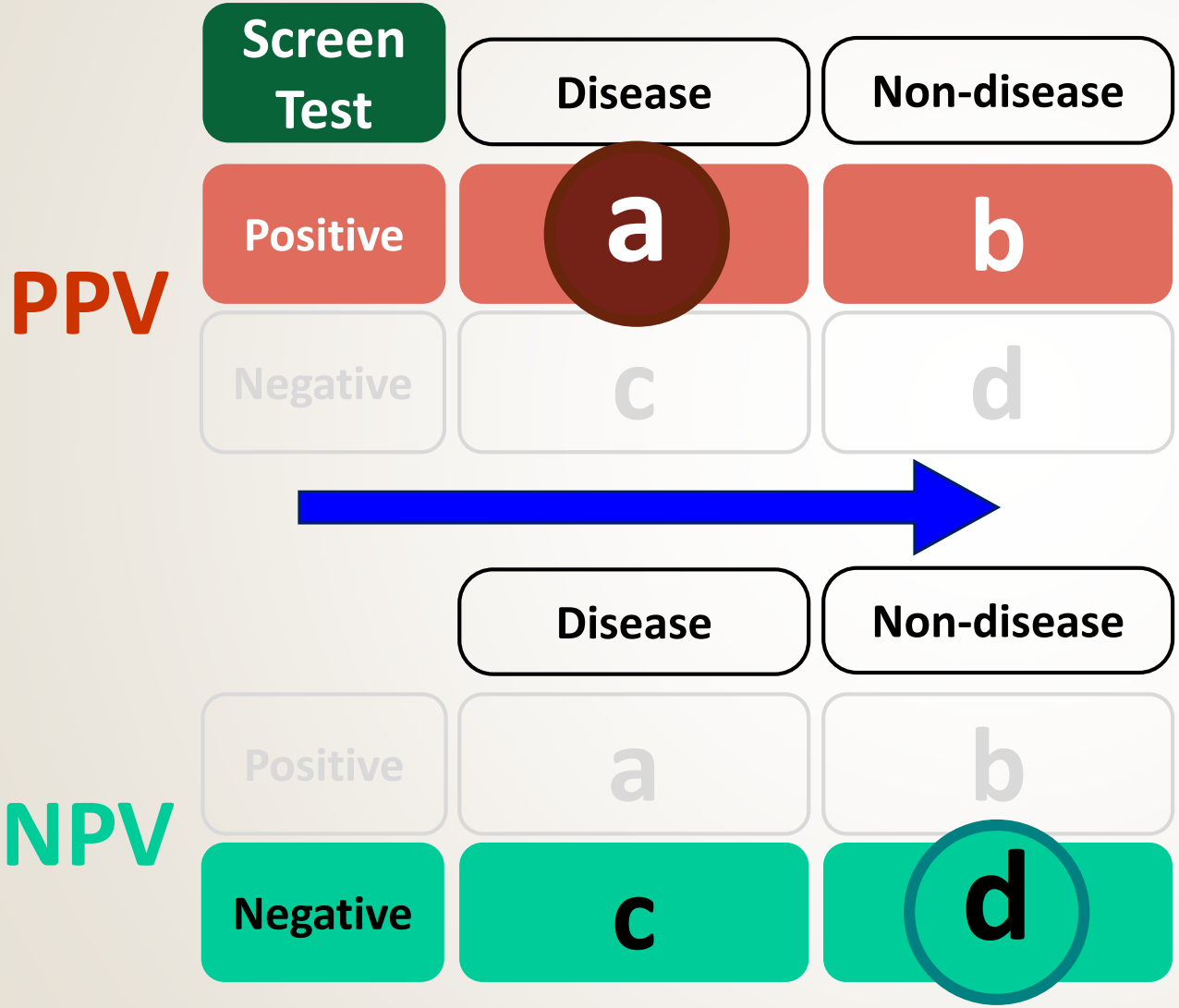
$$\text{Specificity (特異度)} = \frac{d}{(b+d)}$$

Diagnostic test

- Likelihood ratio positive (LR+)
  - Sensitivity / (1 – Specificity)
- Likelihood ratio negative (LR-)
  - (1 – Sensitivity) / Specificity



Diagnostic test



$$\text{Positive predictive value (PPV)} = \frac{a}{(a+b)}$$

$$\text{Negative predictive value (NPV)} = \frac{d}{(c+d)}$$

Diagnostic test

	Disease	Health
Positive	true positive (sen)	false positive
Negative	false negative	true negative (spe)

$$PPV = \frac{\text{true positive}}{\text{all positive}} = \frac{\text{true positive}}{\text{true positive} + \text{false positive}}$$

$$NPV = \frac{\text{true negative}}{\text{all negative}} = \frac{\text{true negative}}{\text{true negative} + \text{false negative}}$$

true positive
= all cases * sensitivity

false positive
= all health * (1 – specificity)

all cases = total * prevalence

all health = total * (1 – prevalence)

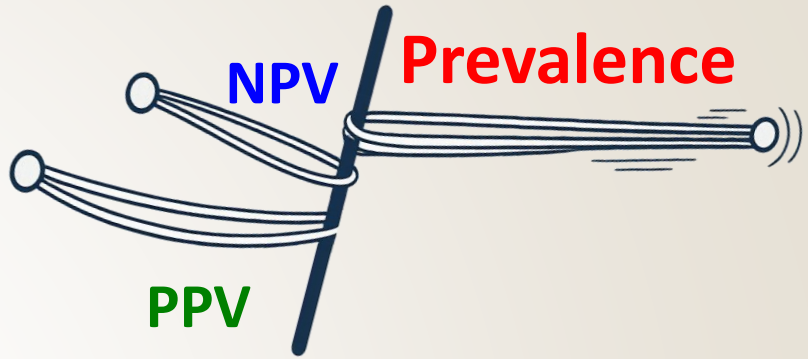
$$PPV = \frac{(\text{total} * \text{prevalence}) * \text{sensitivity}}{(\text{total} * \text{prevalence}) * \text{sensitivity} + (\text{total} * (1 - \text{prevalence}) * (1 - \text{specificity}))}$$

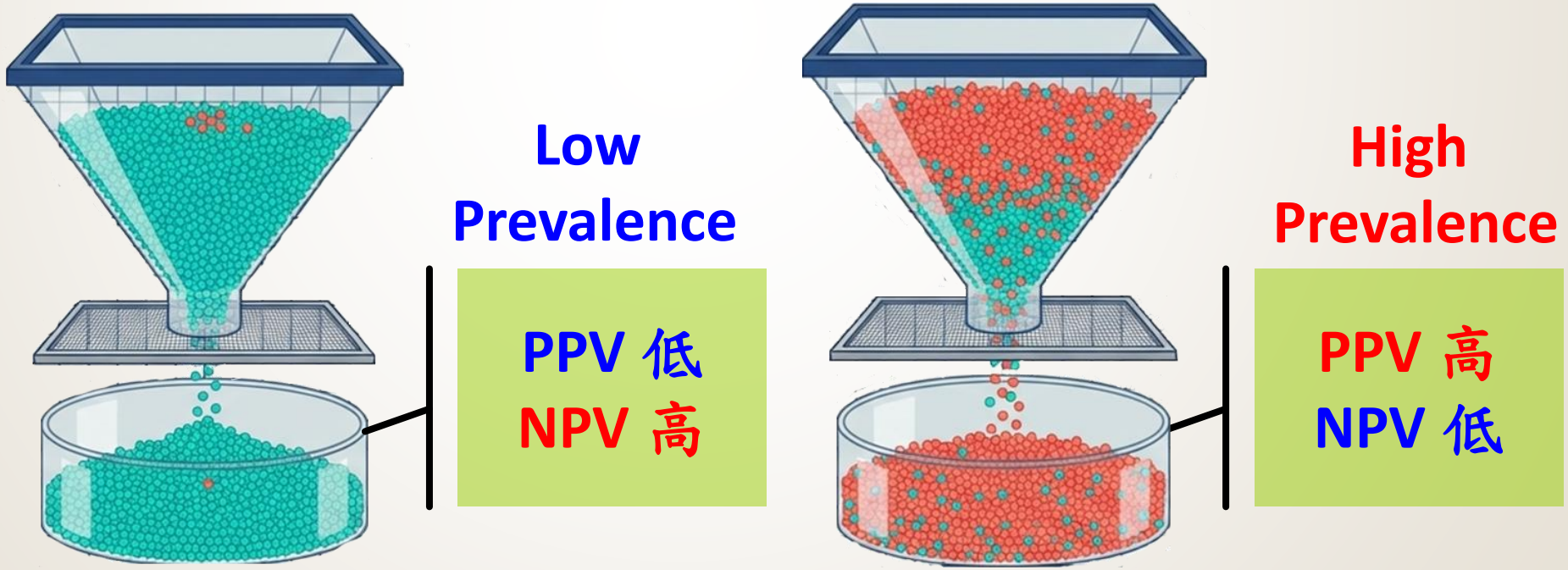
$$= \frac{\text{prevalence} * \text{sensitivity}}{(\text{prevalence} * \text{sensitivity}) + (1 - \text{prevalence}) * (1 - \text{specificity})}$$

$$NPV = \frac{(1 - \text{prevalence}) * \text{specificity}}{(1 - \text{prevalence}) * \text{specificity} + \text{prevalence} * (1 - \text{sensitivity})}$$

Diagnostic test

- Prevalence
  - PPV / NPV





Diagnostic test

評估患者風險 (PPV & NPV) → 受「盛行率」影響

評估篩檢工具 (敏感度&特異度)

真陽性(True Positive)

有病且篩檢陽性

$$\text{Sensitivity} = \frac{TP}{(TP + FN)}$$

偽陽性(False Positive)

沒病但篩檢陽性

$$(1 - \text{Specificity}) = \frac{FP}{(FP + TN)}$$

偽陰性(False Negative)

有病但篩檢陰性

$$(1 - \text{Sensitivity}) = \frac{FN}{(TP + FN)}$$

真陰性(True Negative)

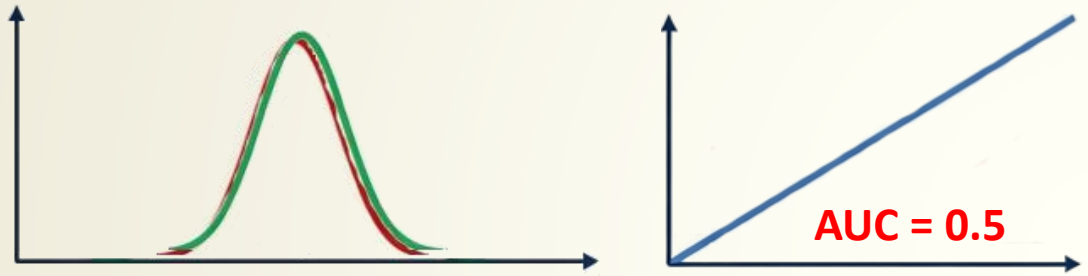
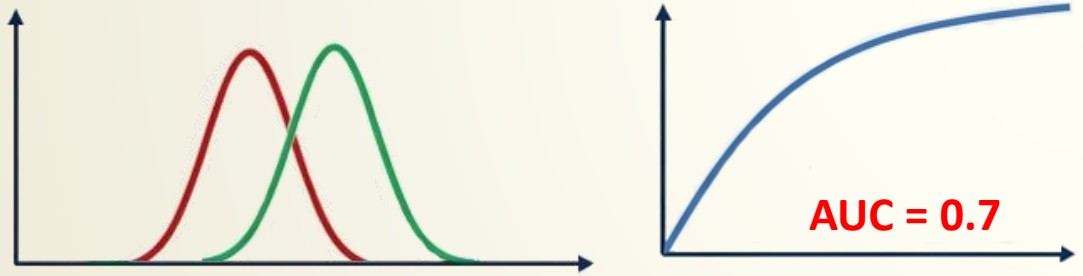
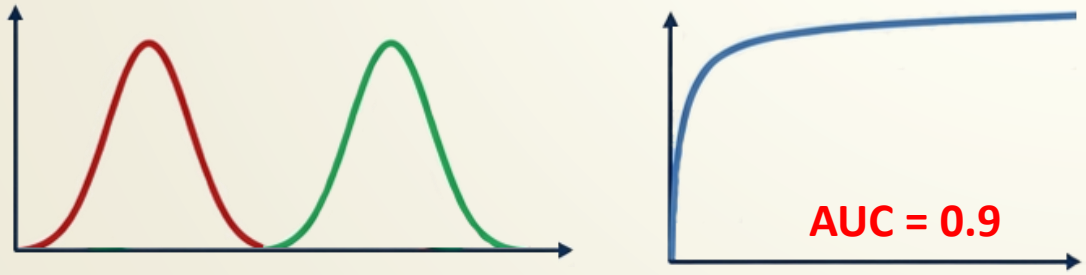
沒病且篩檢陰性

$$\text{Specificity} = \frac{TN}{(FP + TN)}$$

ROC Curve Analysis

- 使用時機
  - 當開發新的檢驗方法，無法決定臨界值 (Cut-off value)
  - 利用連續數值預測結果 (二元分類)
- 目的
  - 將連續數值決定臨界值
  - 用來比較不同工具的好壞

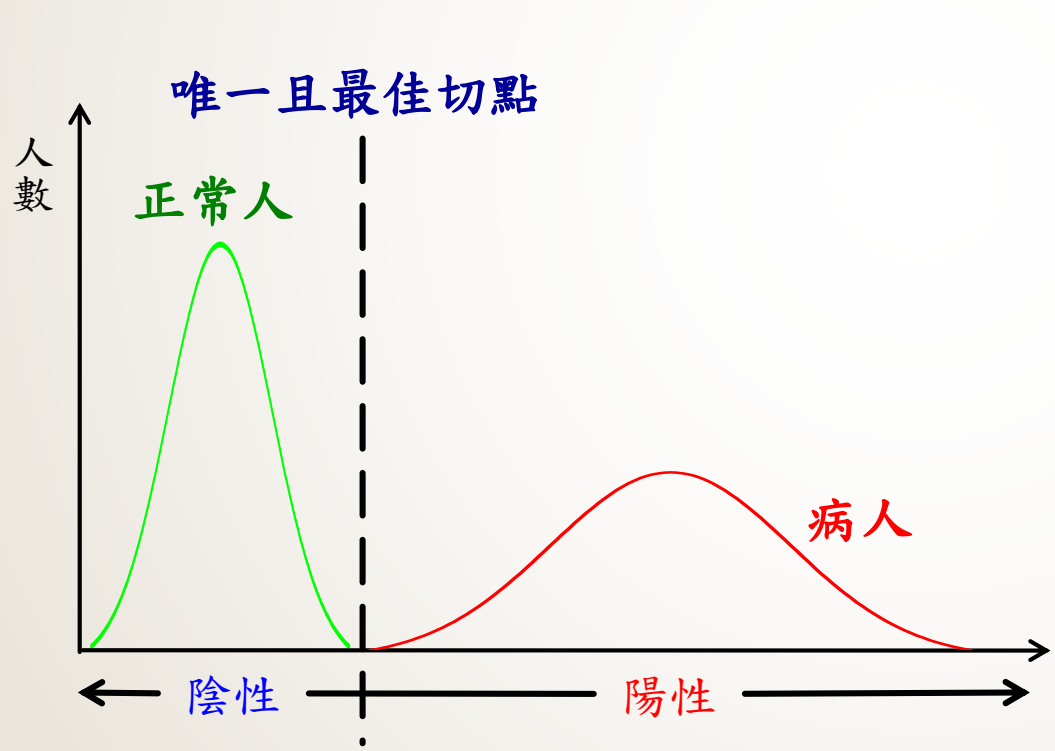
ROC Curve Analysis (AUC 曲線下面積)

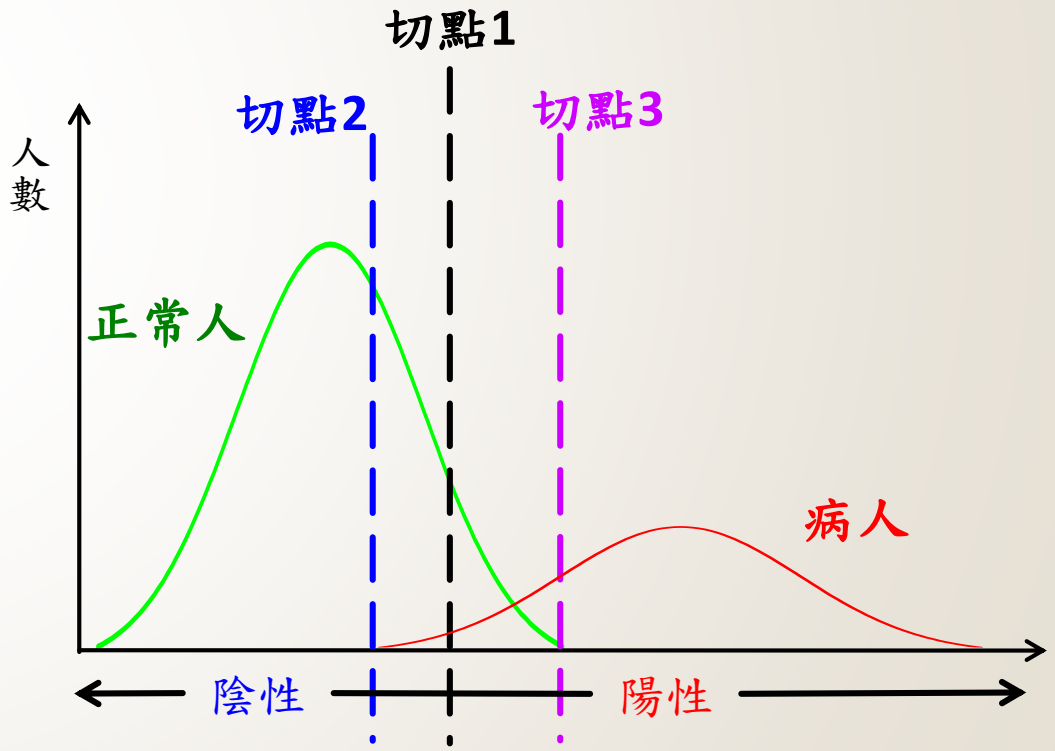
	AUC	Discrimination
	0.5	No discrimination
	0.7-0.8 0.8-0.9	Acceptable discrimination Excellent discrimination
	0.9-1.0	Outstanding discrimination

Ref: https://reurl.cc/288Xr9

ROC Curve Analysis

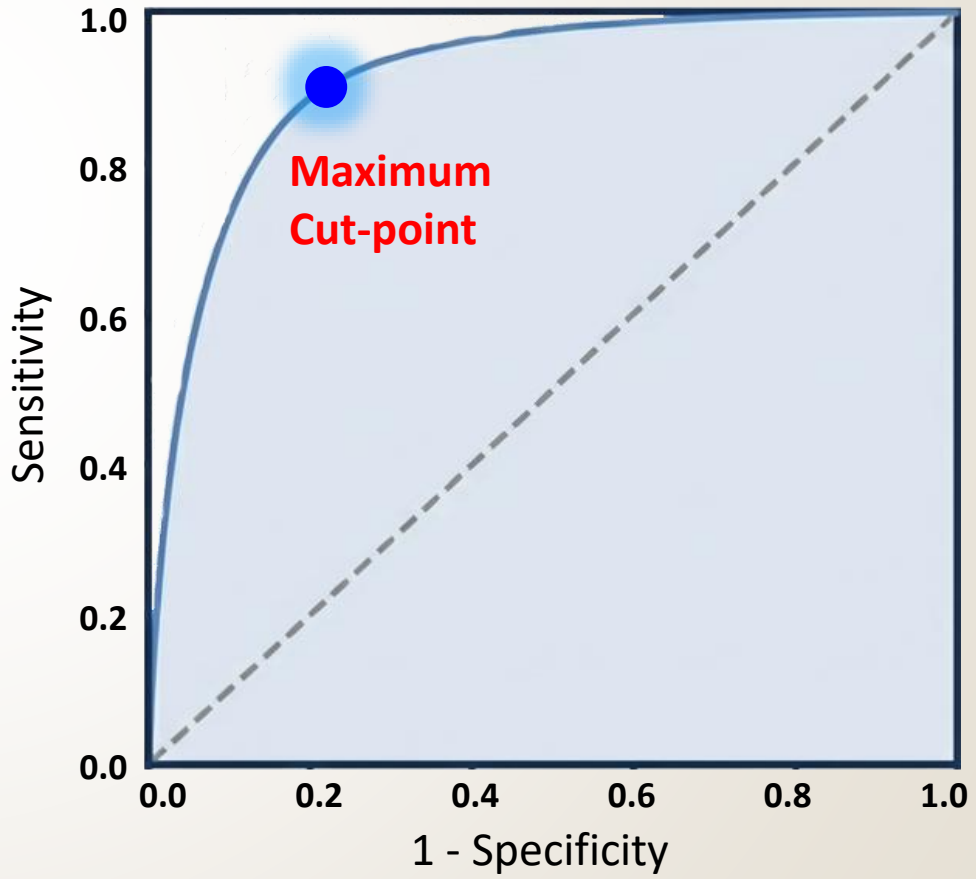
- Cut-point





ROC Curve Analysis

- Youden's index
  - 反映在有疾病和沒有疾病的陽性結果可能機率
  - 公式
    - Sensitivity + Specificity - 1
    - Sensitivity - False positive
  - Range 0-1
  - 不受盛行率影響
  - Maximum
    - Cut-point

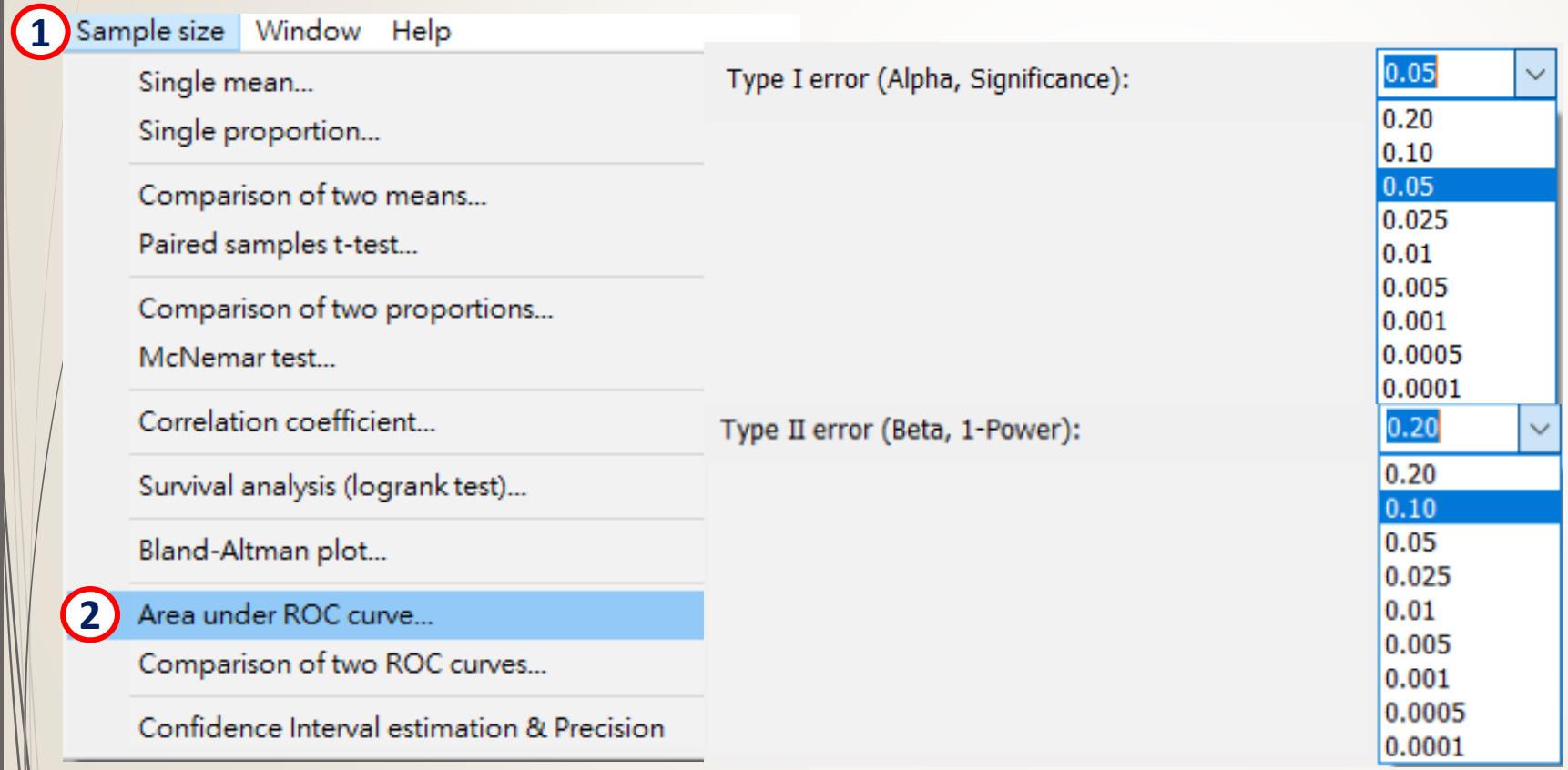


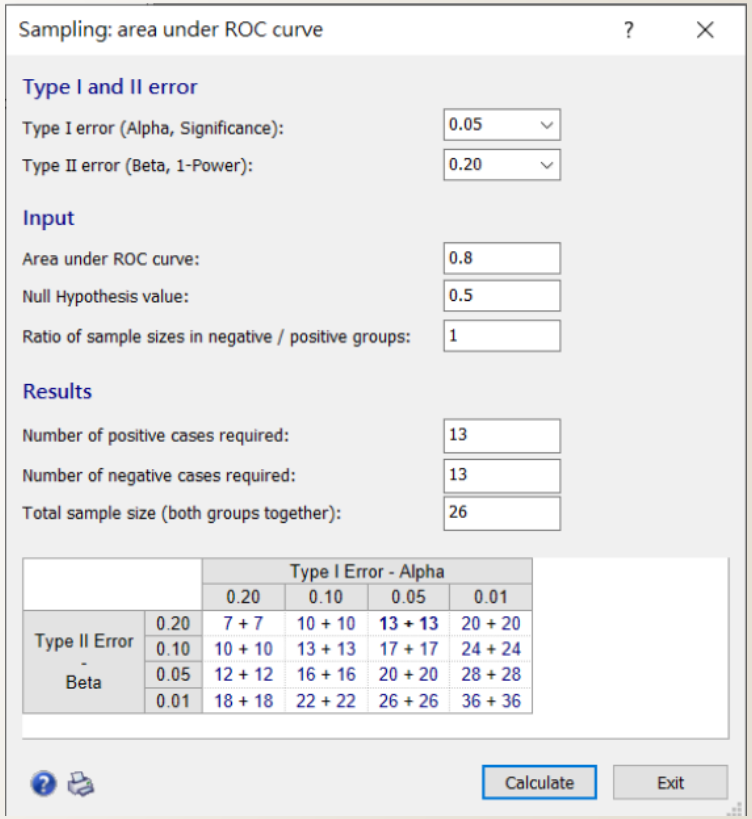
ROC Curve Analysis

- 樣本量-分析結果推論的精確度
  - 假設：AUC = 0.8 、 α = 0.05、 β = 0.2
  - Medcalc計算樣本數
    - **Sampling → Area under ROC**
      - Type I error (Alpha, Significance) = 0.05
      - Type II error (Beta, 1-Power) = 0.20
      - Area under ROC curve = 0.80
      - Ratio of sample sizes in negative/positive groups = 1

ROC Curve Analysis (MedCalc 計算樣本數)

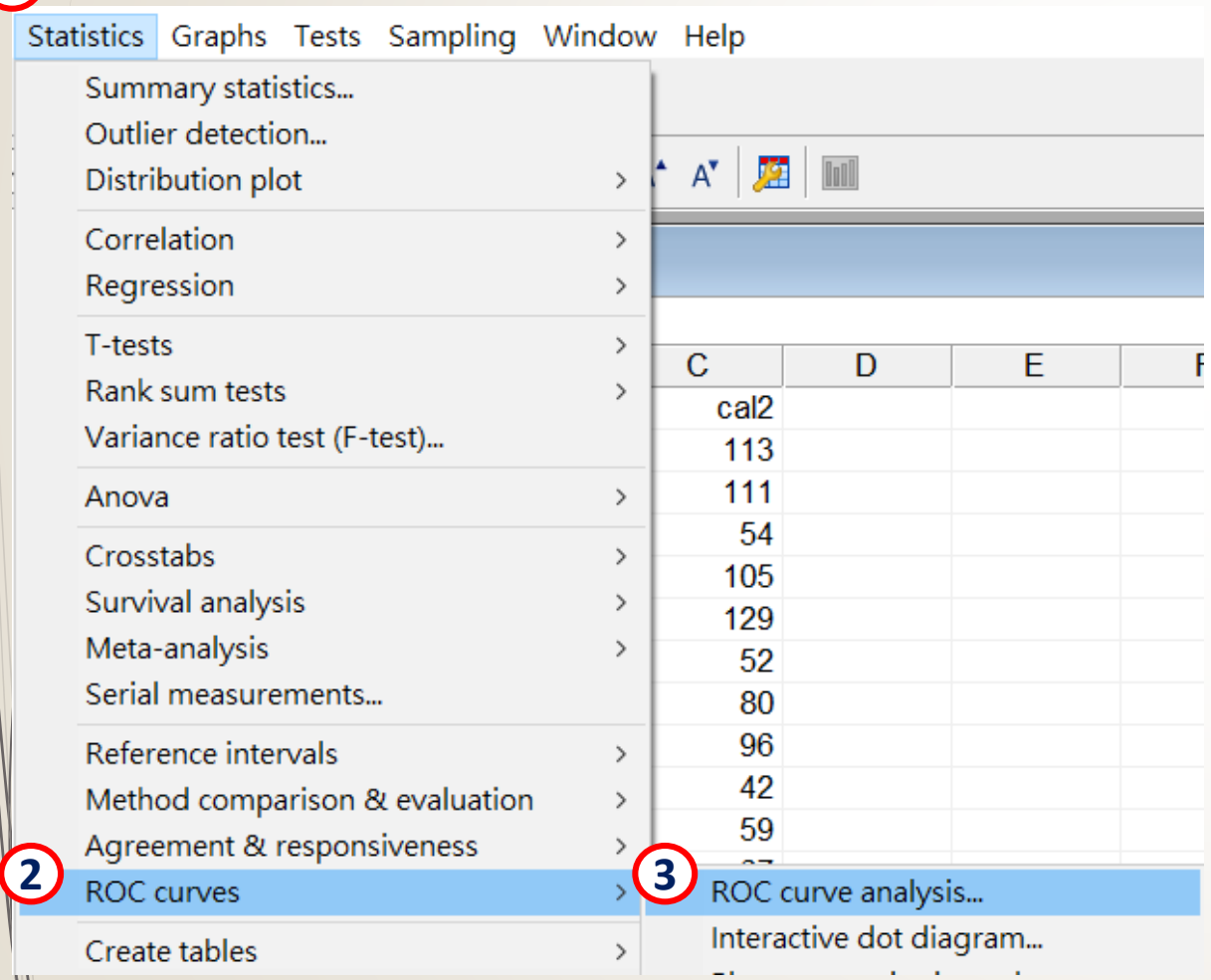
➡ Sample size → Area under ROC curve

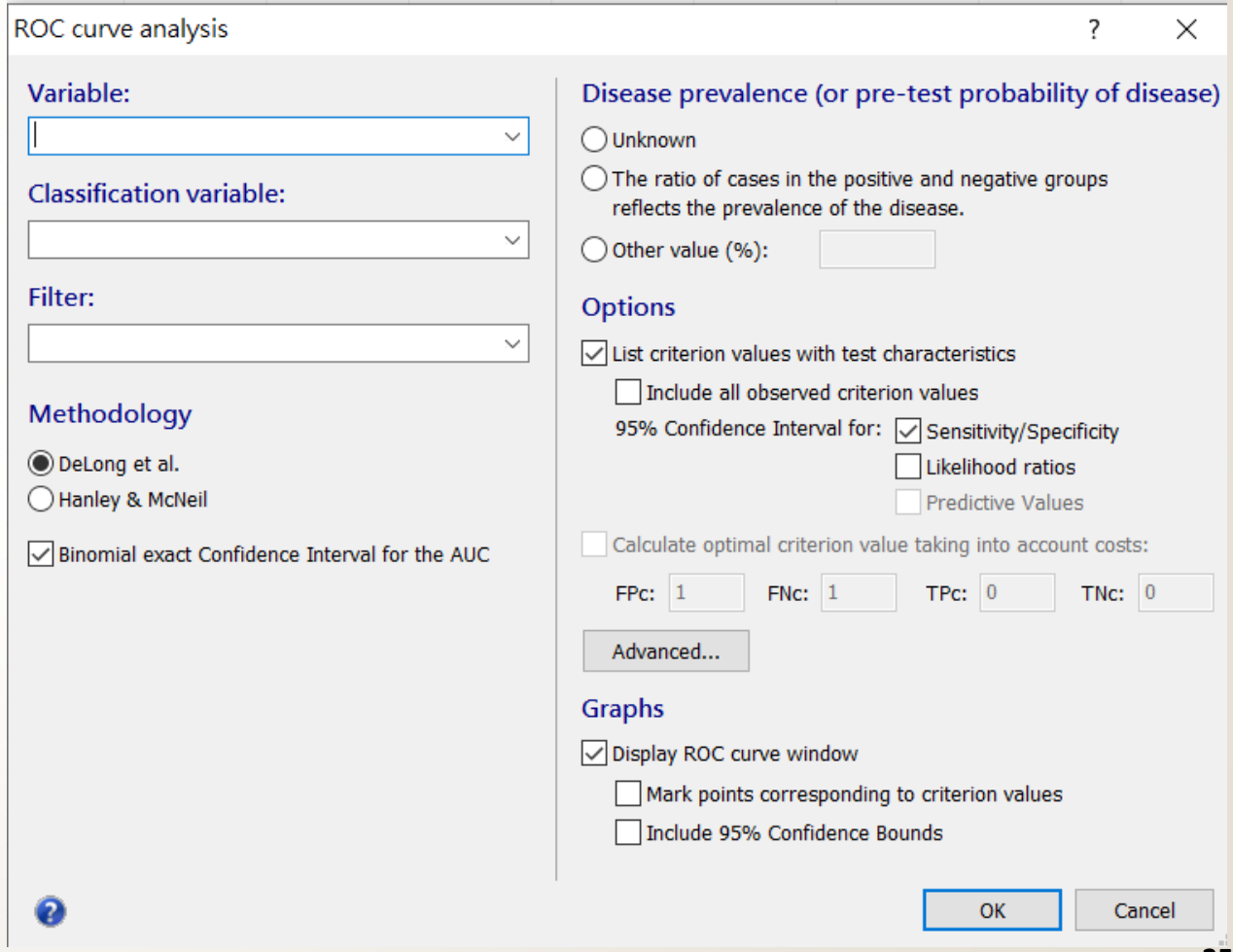




ROC Curve Analysis (MedCalc)

➡ Statistics → ROC curves → ROC Curve Analysis

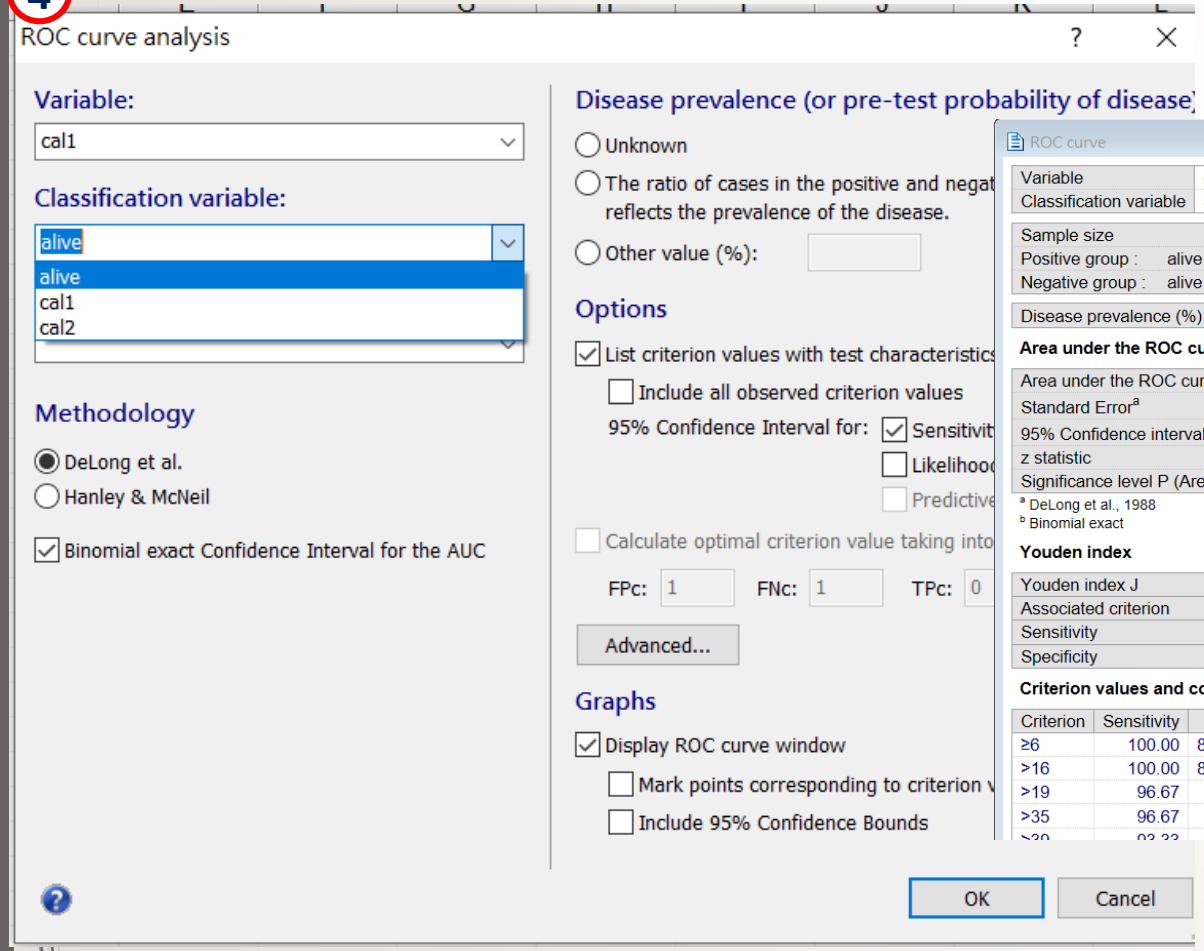


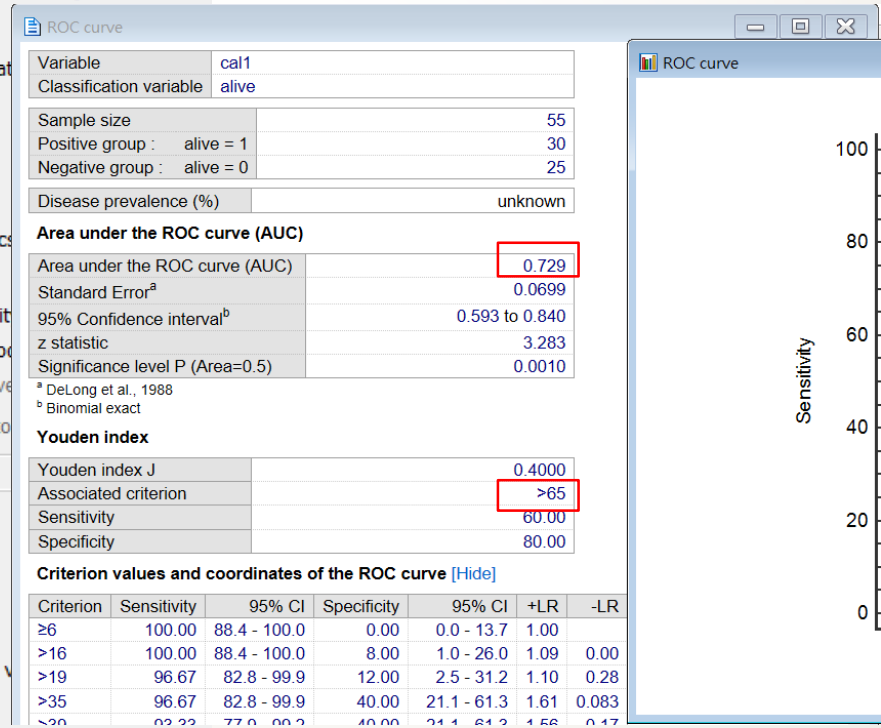


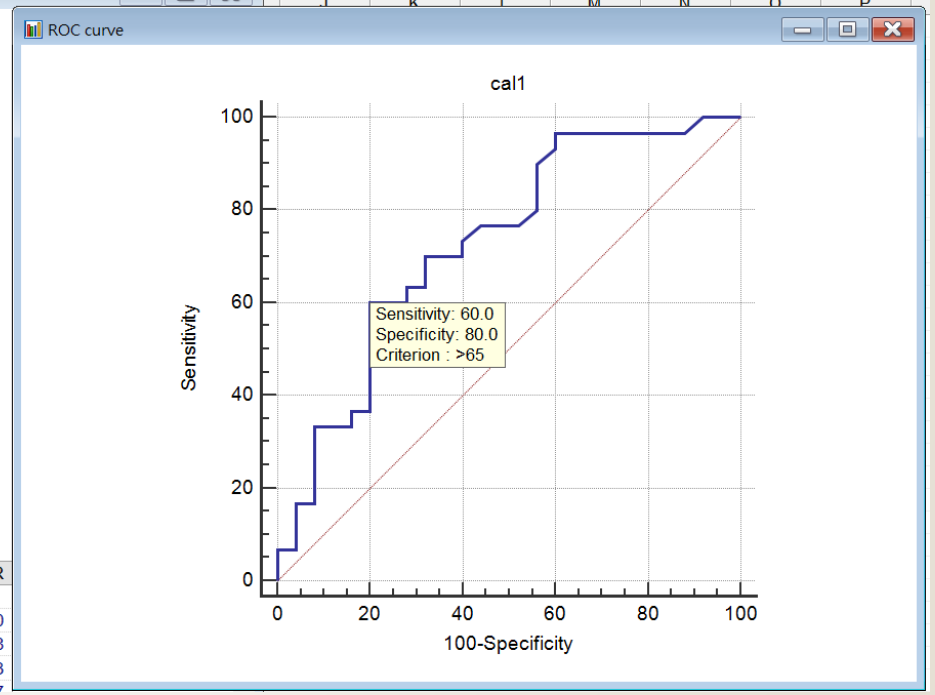
ROC Curve Analysis (MedCalc)

➡ Statistics → ROC curves → ROC Curve Analysis

4







ROC Curve Analysis (MedCalc)

- Statistics → ROC curves → ROC Curve Analysis

ROC curve

Variable	test3
Classification variable	Dis

Sample size	100
Positive group : Dis = 1	20
Negative group : Dis = 0	80

Disease prevalence (%)	unknown

Area under the ROC curve (AUC)

Area under the ROC curve (AUC)	0.739
Standard Error[a]	0.0512
95% Confidence interval[b]	0.641 to 0.822
z statistic	4.660
Significance level P (Area=0.5)	<0.0001

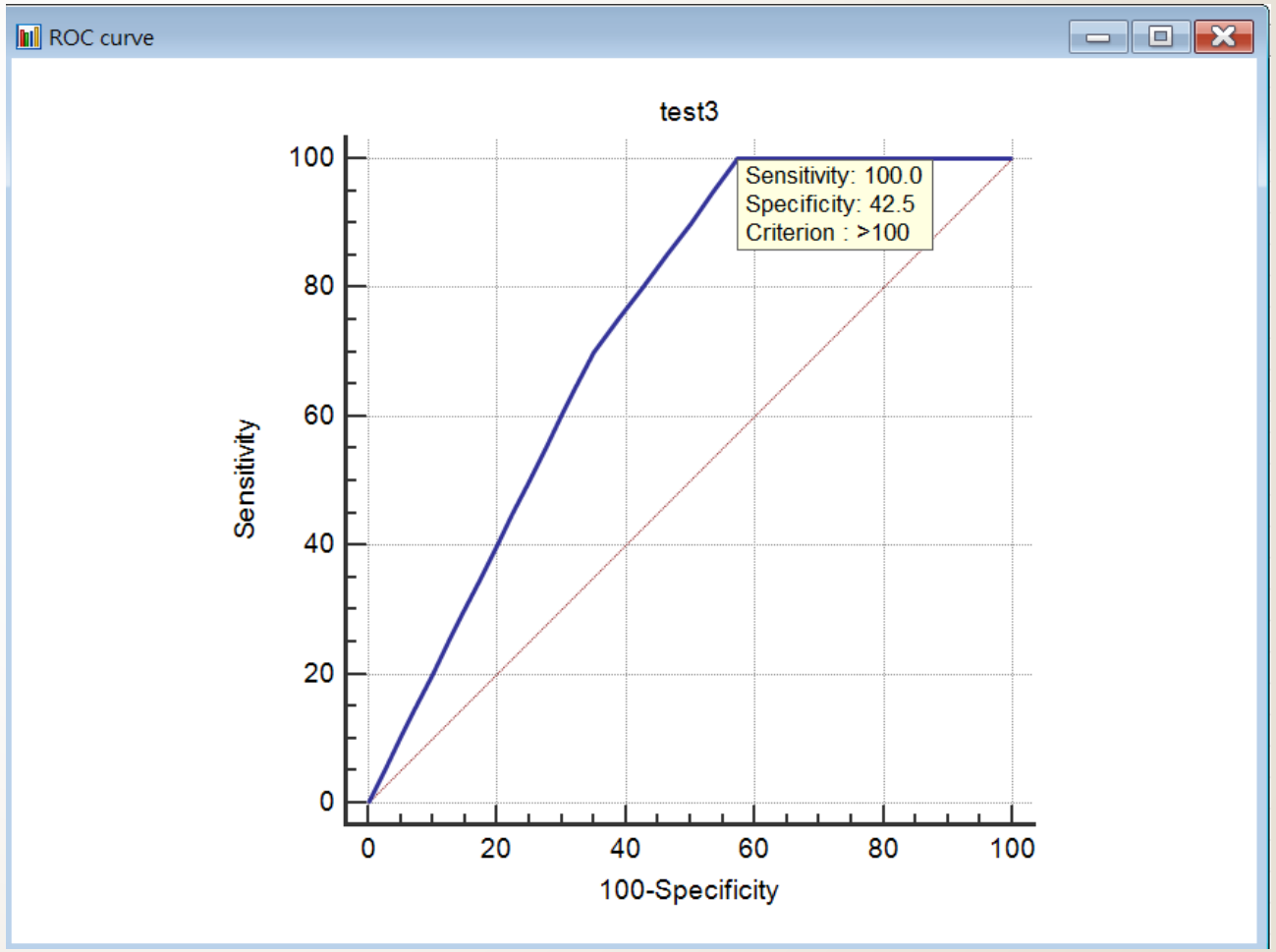
[a] DeLong et al., 1988
[b] Binomial exact

Youden index

Youden index J	0.4250
Associated criterion	>100
Sensitivity	100.00
Specificity	42.50

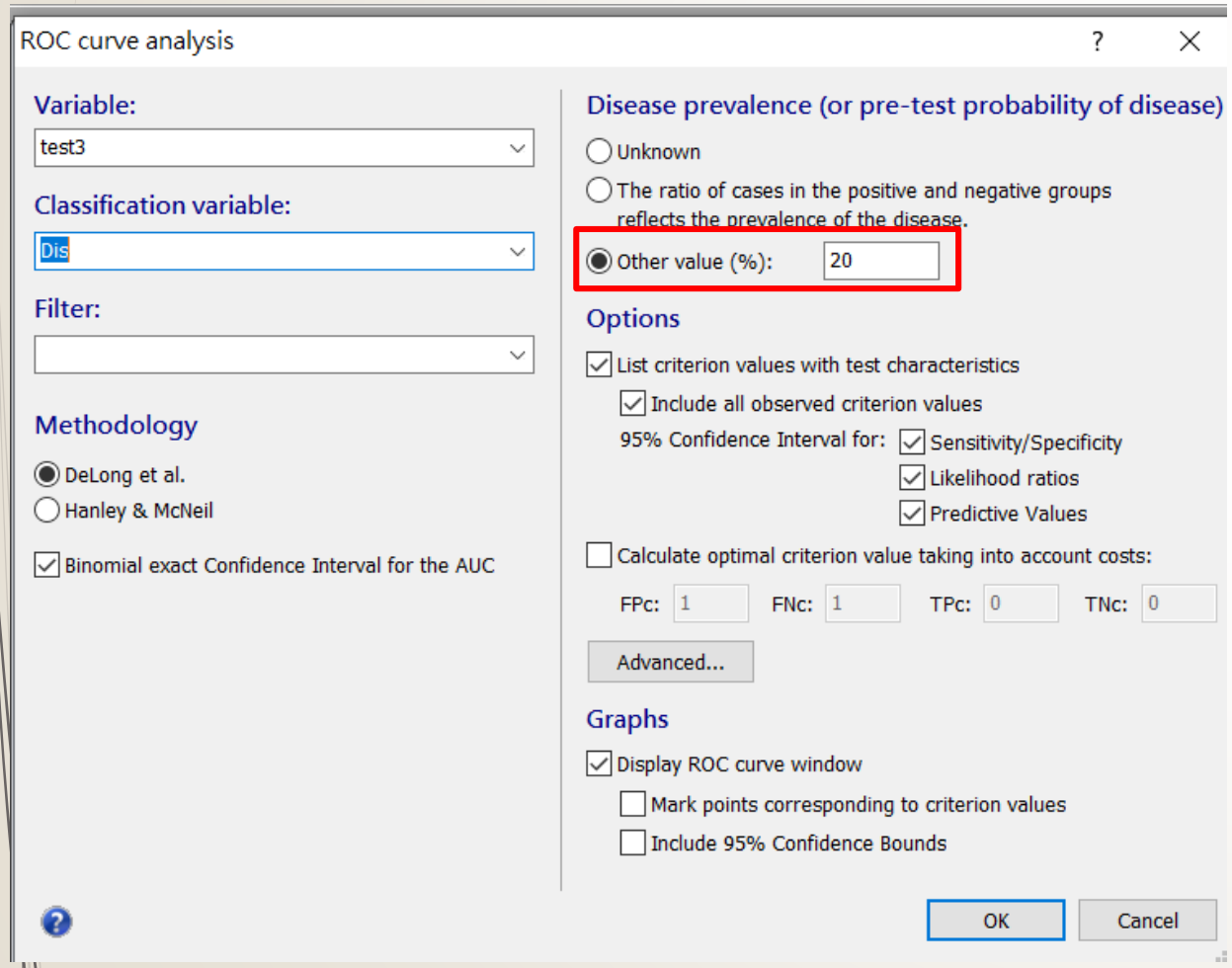
Criterion values and coordinates of the ROC curve [Hide]

Criterion	Sensitivity	95% CI	Specificity	95% CI	+LR	-LR
≥80	100.00	83.2 - 100.0	0.00	0.0 - 4.5	1.00	
>100	100.00	83.2 - 100.0	42.50	31.5 - 54.1	1.74	0.00
>110	95.00	75.1 - 99.9	46.25	35.0 - 57.8	1.77	0.11



ROC Curve Analysis (MedCalc)

➡ Statistics → ROC curves → ROC Curve Analysis



ROC curve

Sample size		100
Positive group :	Dis = 1	20
Negative group :	Dis = 0	80

Disease prevalence (%)	20

Area under the ROC curve (AUC)

Area under the ROC curve (AUC)	0.739
Standard Error[a]	0.0512
95% Confidence interval[b]	0.641 to 0.822
z statistic	4.660
Significance level P (Area=0.5)	<0.0001

[a] DeLong et al., 1988
[b] Binomial exact

Youden index

Youden index J	0.4250
Associated criterion	>100
Sensitivity	100.00
Specificity	42.50

ROC Curve Analysis (MedCalc)

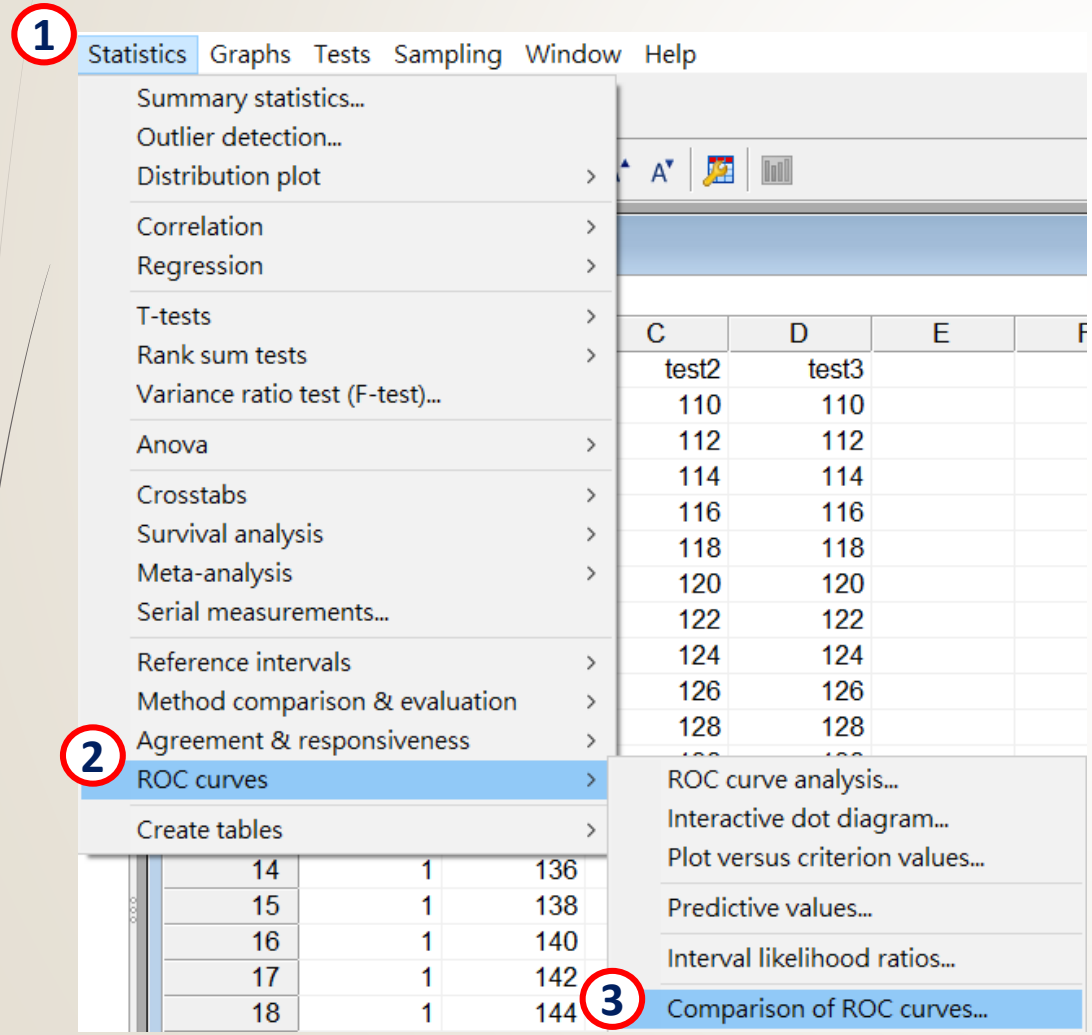
- Statistics → ROC curves → ROC Curve Analysis

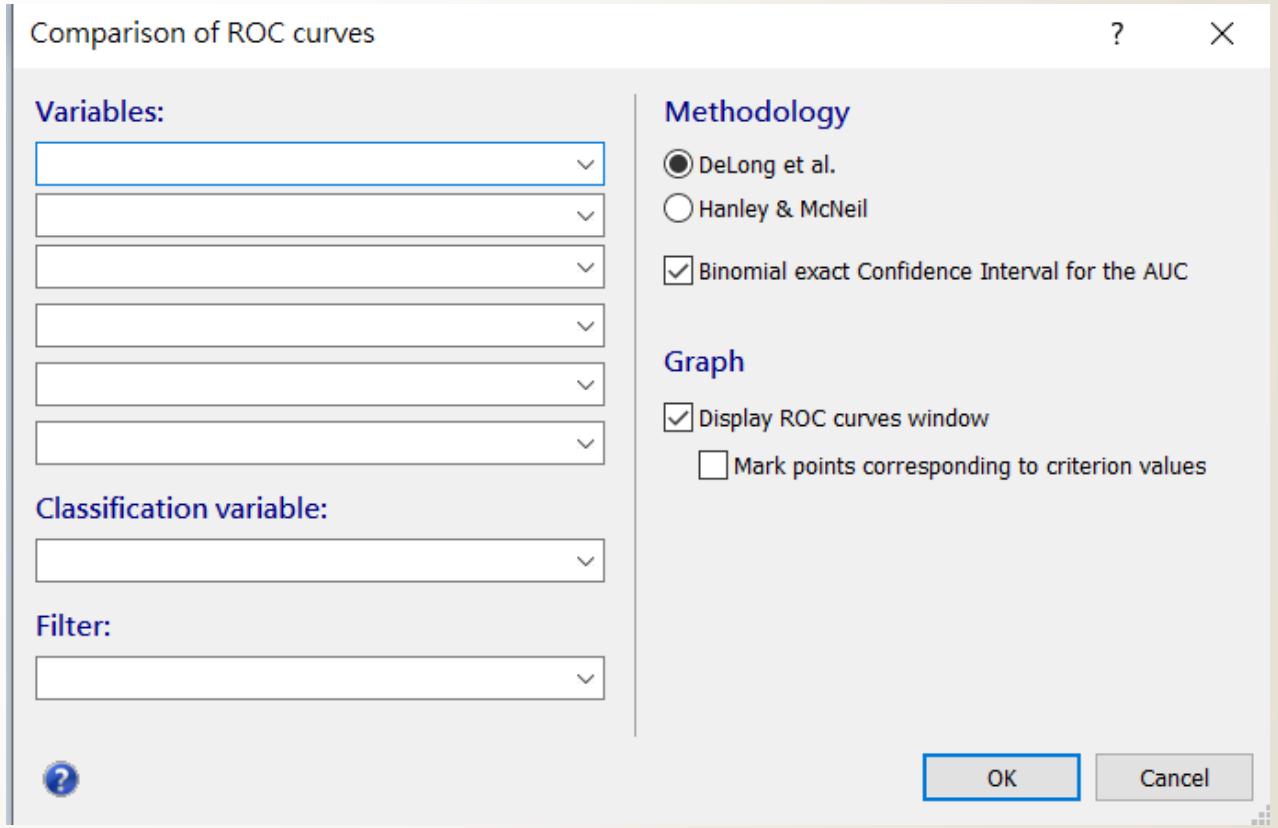
Criterion values and coordinates of the ROC curve [Hide]

Criterion	Sensitivity	95% CI	Specificity	95% CI	+LR	95% CI	-LR	95% CI	+PV	95% CI	-PV	95% CI
≥80	100.00	83.2 - 100.0	0.00	0.0 - 4.5	1.00	1.0 - 1.0			20.0	12.7 - 29.2		
>80	100.00	83.2 - 100.0	2.50	0.3 - 8.7	1.03	1.0 - 1.1	0.00		20.4	12.9 - 29.7	100.0	15.8 - 100.0
>82	100.00	83.2 - 100.0	5.00	1.4 - 12.3	1.05	1.0 - 1.1	0.00		20.8	13.2 - 30.3	100.0	39.8 - 100.0
>84	100.00	83.2 - 100.0	7.50	2.8 - 15.6	1.08	1.0 - 1.2	0.00		21.3	13.5 - 30.9	100.0	54.1 - 100.0
>86	100.00	83.2 - 100.0	10.00	4.4 - 18.8	1.11	1.0 - 1.2	0.00		21.7	13.8 - 31.6	100.0	63.1 - 100.0
>88	100.00	83.2 - 100.0	12.50	6.2 - 21.8	1.14	1.1 - 1.2	0.00		22.2	14.1 - 32.2	100.0	69.2 - 100.0
>90	100.00	83.2 - 100.0	17.50	9.9 - 27.6	1.21	1.1 - 1.3	0.00		23.3	14.8 - 33.6	100.0	76.8 - 100.0
>92	100.00	83.2 - 100.0	22.50	13.9 - 33.2	1.29	1.1 - 1.5	0.00		24.4	15.6 - 35.1	100.0	81.5 - 100.0
>94	100.00	83.2 - 100.0	27.50	18.1 - 38.6	1.38	1.2 - 1.6	0.00		25.6	16.4 - 36.8	100.0	84.6 - 100.0
>96	100.00	83.2 - 100.0	32.50	22.4 - 43.9	1.48	1.3 - 1.7	0.00		27.0	17.4 - 38.6	100.0	86.8 - 100.0
>98	100.00	83.2 - 100.0	37.50	26.9 - 49.0	1.60	1.4 - 1.9	0.00		28.6	18.4 - 40.6	100.0	88.4 - 100.0
>100	100.00	83.2 - 100.0	42.50	31.5 - 54.1	1.74	1.4 - 2.1	0.00		30.3	19.6 - 42.9	100.0	89.7 - 100.0
>110	95.00	75.1 - 99.9	46.25	35.0 - 57.8	1.77	1.4 - 2.2	0.11	0.02 - 0.7	30.6	19.6 - 43.7	97.4	86.2 - 99.9
>112	90.00	68.3 - 98.8	50.00	38.6 - 61.4	1.80	1.4 - 2.3	0.20	0.05 - 0.8	31.0	19.5 - 44.5	95.2	83.8 - 99.4

ROC Curve Analysis (MedCalc)

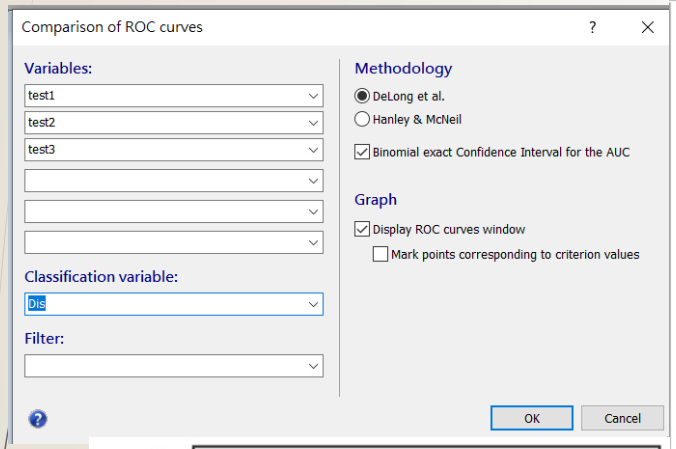
Statistics → ROC curves → Comparison of ROC Curves

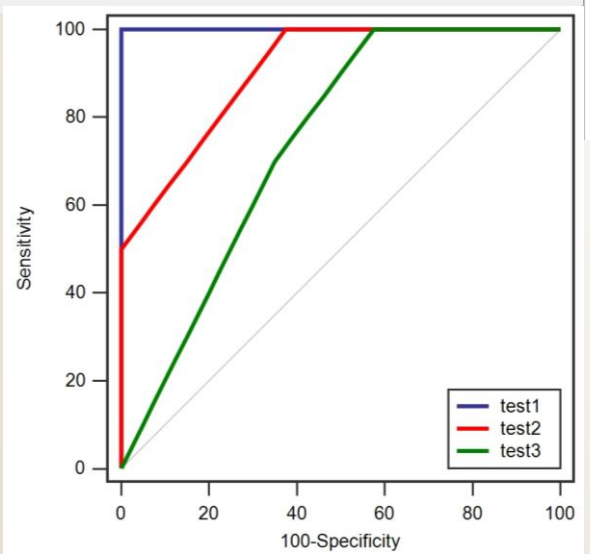




ROC Curve Analysis (MedCalc)

➡ Statistics → ROC curves → Comparison of ROC Curves





Variable 1	test1
Variable 2	test2
Variable 3	test3
Classification variable	Dis

Sample size	100
Positive group : Dis = 1	20
Negative group : Dis = 0	80

	AUC	SE [a]	95% CI [b]
test1	1.000	0.000	0.964 to 1.000
test2	0.906	0.0324	0.831 to 0.955
test3	0.739	0.0512	0.641 to 0.822

[a] DeLong et al., 1988
[b] Binomial exact

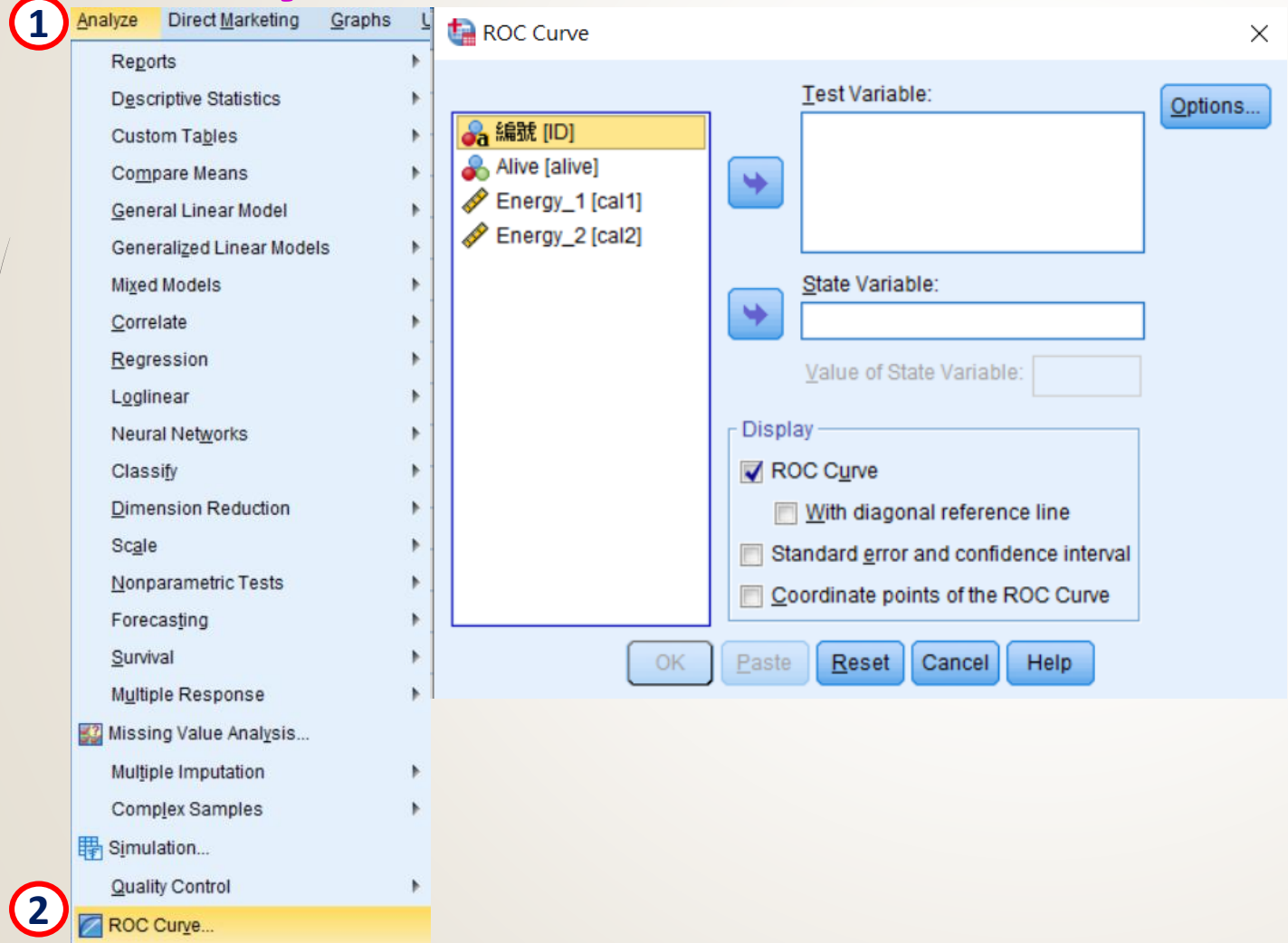
Pairwise comparison of ROC curves

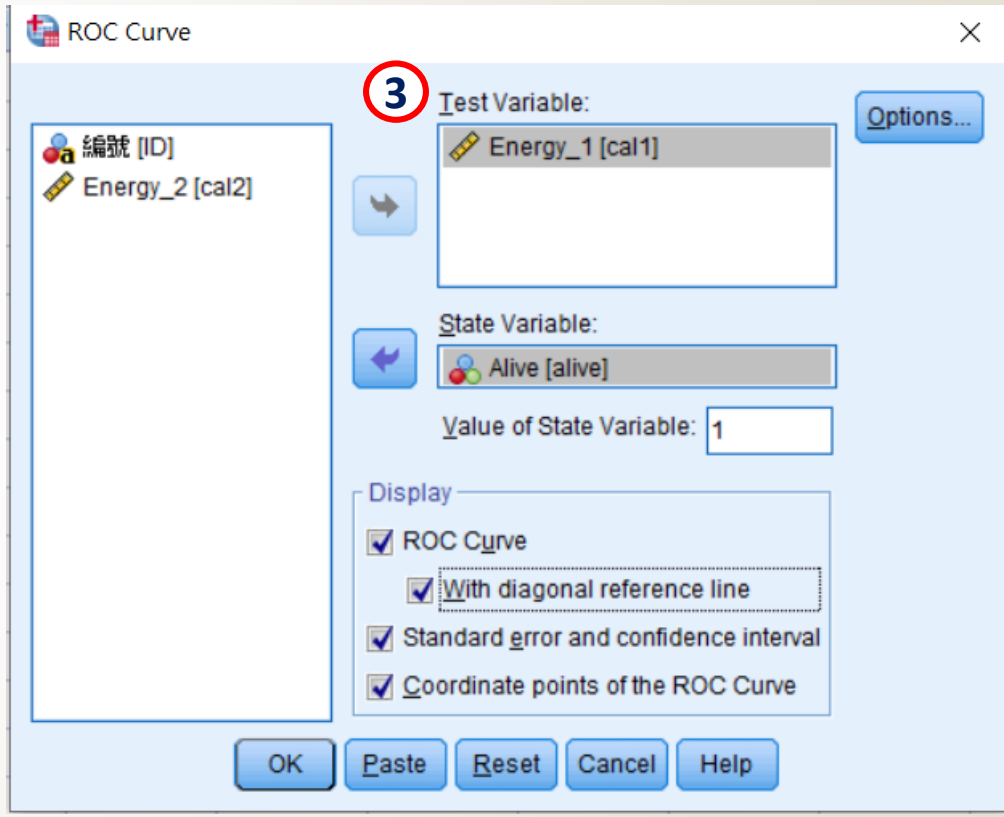
test1 ~ test2	
Difference between areas	0.0937
Standard Error[c]	0.0324
95% Confidence Interval	0.0302 to 0.157
z statistic	2.892
Significance level	P = 0.0038
test1 ~ test3	
Difference between areas	0.261
Standard Error[c]	0.0512
95% Confidence Interval	0.161 to 0.362
z statistic	5.099
Significance level	P < 0.0001
test2 ~ test3	
Difference between areas	0.168
Standard Error[c]	0.0349
95% Confidence Interval	0.0991 to 0.236
z statistic	4.797
Significance level	P < 0.0001

[c] DeLong et al., 1988

ROC Curve Analysis (SPSS)

Analyze → ROC curve





練習範例：ROC ex.sav

ROC Curve Analysis (SPSS)

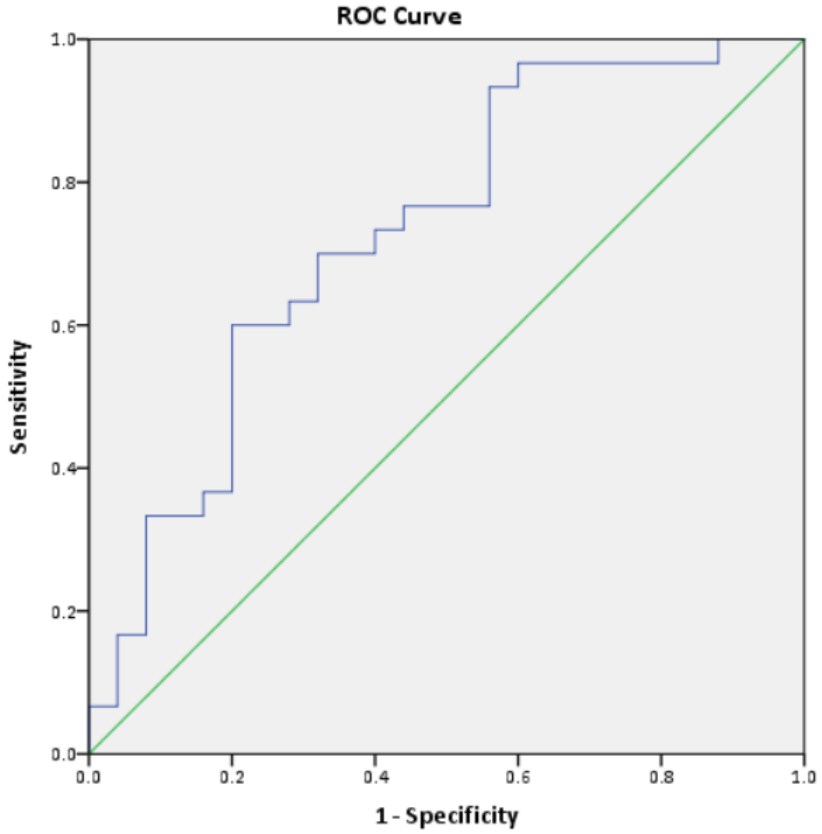
ROC Curve

Case Processing Summary

Alive	Valid N (listwise)
Positive[a]	30
Negative	25

Larger values of the test result variable(s) indicate stronger evidence for a positive actual state.

a. The positive actual state is Alive.



Area Under the Curve

Test Result Variable(s): Energy_1

Area	Std. Error[a]	Asymptotic Sig.[b]	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.729	.069	.004	.594	.865

a. Under the n

b. Null hypothes

Co

Test Result Va

Positive if Greater Than or Equal To[a]	Sensitivity	1 - Specificity
4.82	1.000	1.000
10.68	1.000	.960
17.07	1.000	.920
18.76	1.000	.880
21.61	.967	.880
24.43	.967	.840
25.66	.967	.800

AUC = 0.73
(0.59-0.87)
Acceptable discrimination

練習範例：ROC ex.sav

ROC Curve Analysis (SPSS & Excel)

- Cut-point
  - Youden's index (Maximum)
    - Sensitivity + Specificity - 1

IF | =B4+D4-1

	A	B	C	D	E
1	Coordinates of the Curve				
2	Test Result Variable(s):				
3	Positive if Greater Than or Equal To[a]	Sensitivity	1 - Specificity	Specificity	Youden's index
4	65.22	.600	.200	0.800	=B4+D4-1
5	54.90	.700	.320	0.680	0.380
6	40.99	.933	.560	0.440	0.373
7	36.66	.967	.600	0.400	0.367
8	66.28	.567	.200	0.800	0.367
9	64.15	.600	.240	0.760	0.360
10	60.46	.633	.280	0.720	0.353
11	55.91	.667	.320	0.680	0.347
12	41.64	.900	.560	0.440	0.340
13	53.00	.700	.360	0.640	0.340
14	39.65	.933	.600	0.400	0.333
15	66.95	.533	.200	0.800	0.333

練習範例：ROC ex.sav

Reference

Supplementary Materials

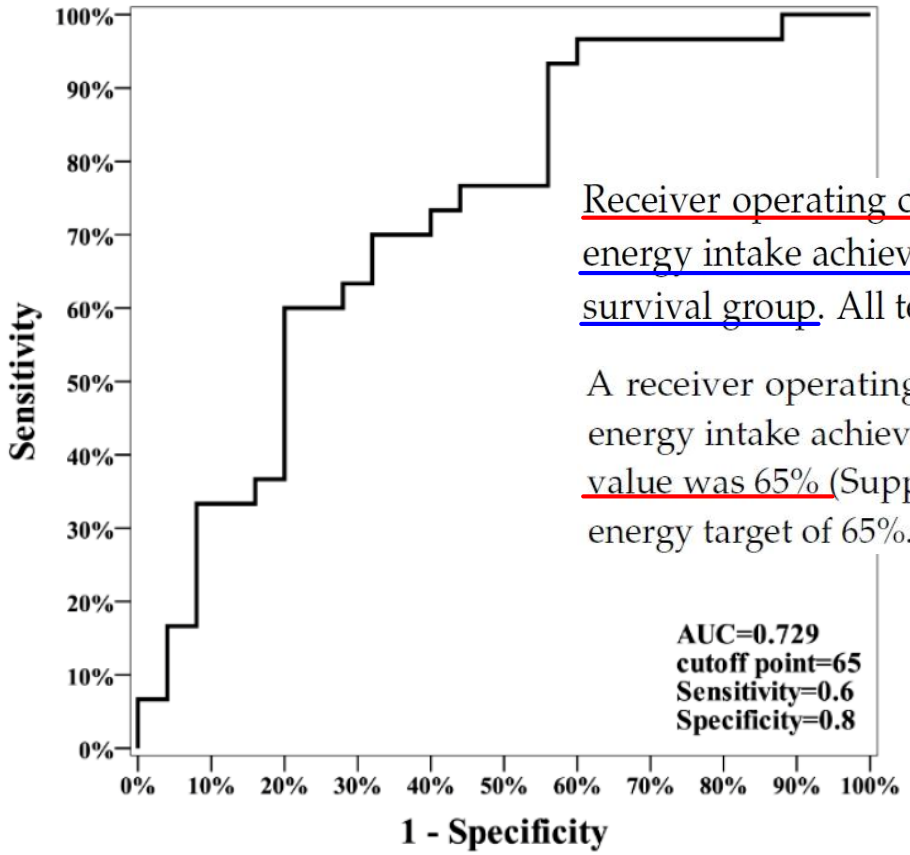


Figure 1. Receiver operating characteristic (ROC) curve to determine the cutoff point for the feeding target between surviving and non-surviving malnourished patients administered SBEN.

Receiver operating characteristic (ROC) curves were used to evaluate the discriminative ability of energy intake achievement rates on day 3 after the initiation of small bowel feeding to identify the survival group. All tests were two-sided, with $p < 0.05$ considered significant.

A receiver operating characteristic (ROC) curve was used to evaluate the discriminative ability of energy intake achievement rates 3 days after SBEN initiation to identify the survival group; the cutoff value was 65% (Supplement Figure S1). In the survival group, two-thirds of the patients achieved the energy target of 65%.

Nutrients 2019, 11, 645; Optimal Time and Target for Evaluating Energy Delivery after Adjuvant Feeding with Small Bowel Enteral Nutrition in Critically Ill Patients at High Nutrition Risk

Reference

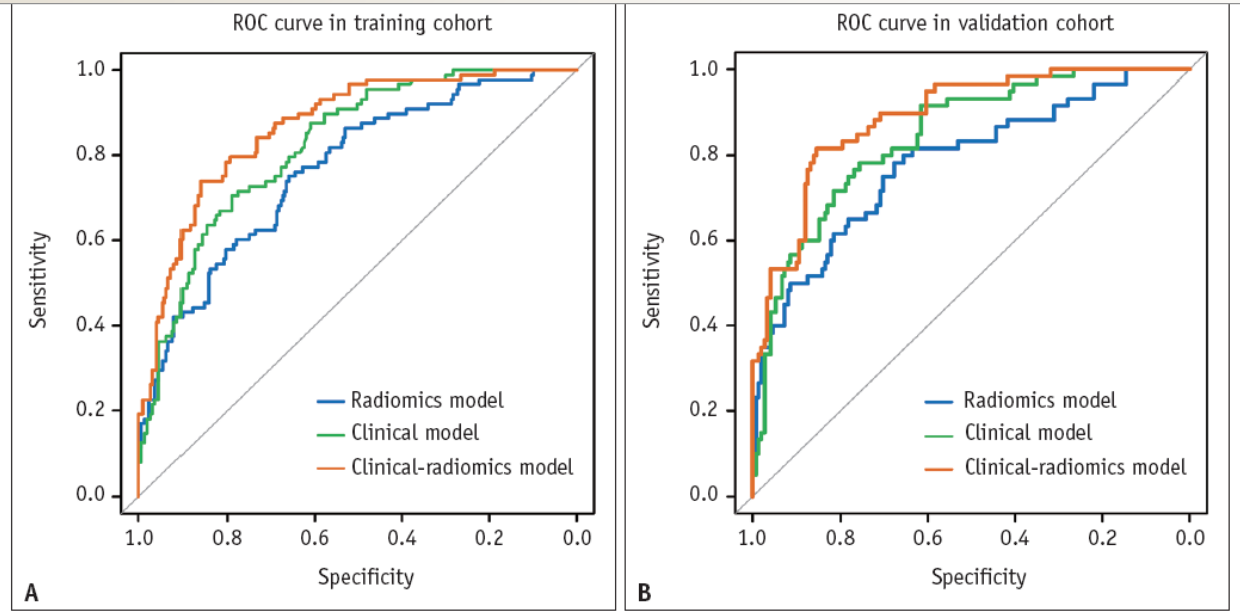


Fig. 2. ROC curves of the radiomics model, clinical model, and clinical-radiomics model in the training (A) and validation (B) cohorts. ROC = receiver operating characteristic

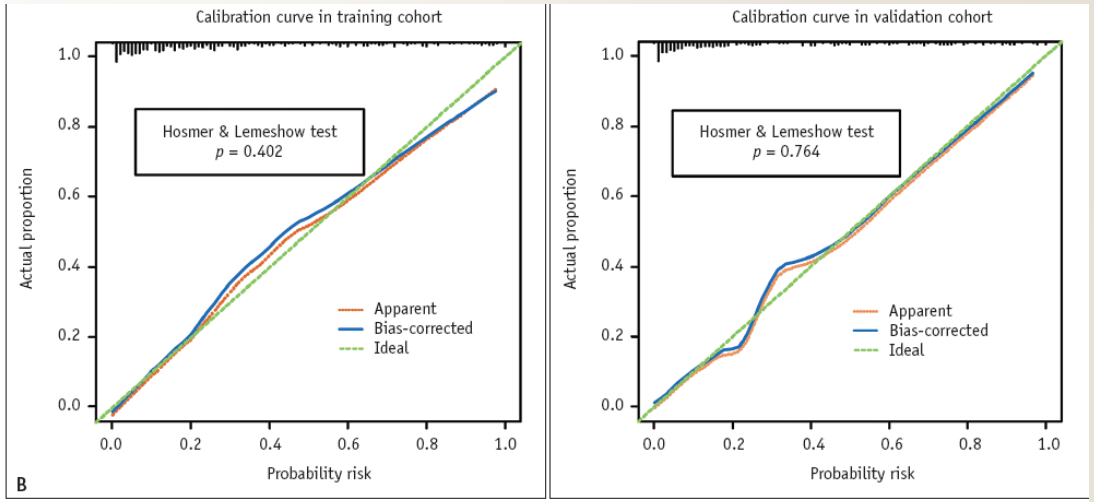


Fig. 3. The clinical-radiomics nomogram for predicting acute ischemic stroke outcomes.
A. The developed nomogram based on the clinical-radiomics prediction model to predict the risk of poor stroke outcome. Diabetes: 0, no diabetes; 1, diabetes. Sex: 0, female; 1, male. Stroke history: 0, no stroke history; 1, stroke history; $mRS_{baseline}$: 0, ≤ 2; 1, > 2. **B.** Calibration curves for the nomogram in the training and validation cohorts. The green dashed line represents the ideal prediction and the red dashed line represents the predictive ability of the nomogram. The closer the dashed red line fit to the dashed green line, the greater the prediction accuracy of the nomogram. **C.** Decision curve analysis for the nomogram. The black line represents the net benefit of assuming no stroke patients have

The predictive performance of the radiomics, clinical, and clinical-radiomics models was evaluated using **receiver operating characteristic (ROC) curves**. The area under the ROC curve **(AUC)** and balanced **sensitivity** and **specificity** at the cutoff yielding the largest **Youden index** value were calculated. The performance of the three models was tested in the training and validation cohorts. The Delong test was used to compare the AUC between the models.

The calibration curve and **Hosmer–Lemeshow test** were used to assess the calibration performance of the clinicalradiomics nomogram.

Korean J Radiol. 2022 Aug;23(8):811-820. Feasibility of a Clinical-Radiomics Model to Predict the Outcomes of Acute Ischemic Stroke. Yiran Zhou, Di Wu, Su Yan, Yan Xie, Shun Zhang , Wenzhi Lv, Yuanyuan Qin, Yufei Liu, Chengxia Liu, Jun Lu, Jia Li, Hongquan Zhu, Weiyin Vivian Liu, Huan Liu, Guiling Zhang, Wenzhen Zhu

總結

Calibration

X

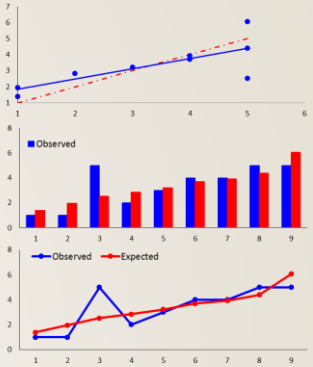
(模型)

預測

Y

(事件機率)

Hosmer and Lemeshow test

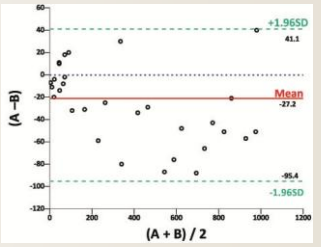


Agreement

quantitative Method A

quantitative Method B

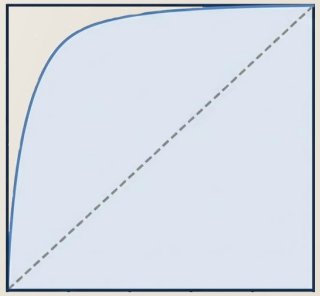
Bland–Altman plot



Discrimination

檢驗方法
(診斷工具)

ROC curve



生統小組
統計方法教育訓練

Thank you

問卷調查

4/28問卷

For your attention!!