羅吉斯迴歸分析

醫學研究部生統小組

陳俊朋

2023/11/14

Logistic regression

- 主要在探討以自變項 (Independent variable) 的變化來預測或解釋結果變項 (Dependent variable) 的變化。
- 目的要建立變數間的因果關係，以利預測結果
- 依變項為二元分類資料（0 or 1） ⟶ 預測 1 的機率
  - 年齡與有無特定慢性病的關係
  - 吸菸與有無肺癌的關係
  - 嚼檳榔與有無口腔癌的關係
  - 醫師年資與手術成功與否的關係
- 自變項為類別變數或連續變數

Logistic regression

➢ 單變項迴歸表示式

$$\ln\frac{p}{1-p} = \beta_0 + \beta_1 X$$

➢ 多變項迴歸表示式

$$\ln\frac{p}{1-p} = \beta_0 + \beta_1 X + \beta_2 X_2 + \cdots + \beta_k X_k$$

➢ 自變數 (X) 對依變數 (Y) 的影響是以指數的方式做變動，因此不需要常態分配的假設

➢ 使用「最大概似函數估計法 (Maximum Likelihood Estimation)」做參數估計
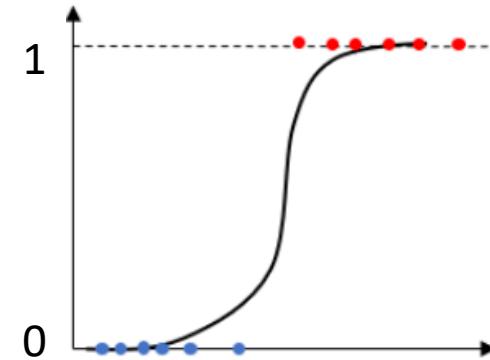
Odds ratio (OR)

Outcome

	Lung cancer	Control
Smoking	a	b
Non Smoking	c	d

Exposure

$$OR = \frac{a/c}{b/d} = \frac{ad}{bc}$$

Odds ratio	意義
OR=1	抽菸與肺癌無關
OR>1	抽菸可能為肺癌的危險因子
OR<1	抽菸可能為肺癌的保護因子

Logistic regression example-1

- 探討角膜生物力學特性預測青光眼的風險

Table 3. Logistic Regression Analyses for Predicting Primary Angle Closure Glaucoma

Measurement	Univariable		Multivariable	
	Odds Ratio (95% CI)	P Value	Odds Ratio (95% CI)	P Value
Age (years)	1.120 (1.070–1.173)	<0.001	1.067 (0.982–1.159)	0.126
Sex (Male)	0.293 (0.148–0.581)	<0.001	0.350 (0.100–1.223)	0.100
CCT (μm)	0.980 (0.970–0.990)	<0.001	0.996 (0.974–1.018)	0.700
IOP (mm Hg)	0.898 (0.814–0.991)	0.032	0.959 (0.796–1.154)	0.656
ACV (mm^3)	0.947 (0.931–0.964)	<0.001	0.950 (0.929–0.971)	<0.001
CBiF	0.021 (0.006–0.079)	<0.001	0.029 (0.003–0.266)	0.002

Logistic regression SPSS dataset

caseno	group	Age	Sex	IOP.nct	CCT	CBiF
1	0	82.0	1.0	11.5	523	.
2	1	77.0	1.0	19.5	532	6.522
3	1	75.0	.0	13.5	641	5.845
4	1	76.0	.0	10.5	523	6.024
5	0	72.0	1.0	14.5	557	6.246
6	0	73.0	.0	18.0	571	6.946
7	1	69.0	.0	12.5	535	5.031
8	0	79.0	.0	14.0	553	6.385
9	1	68.0	1.0	18.5	549	6.441
10	1	68.0	1.0	17.5	542	6.162
11	1	78.0	1.0	14.0	541	6.723
12	0	62.0	.0	14.5	591	6.490
13	0	70.0	1.0	14.5	557	6.283
14	0	79.0	1.0	16.5	534	6.416
15	1	63.0	1.0	18.0	549	6.435

Logistic regression SPSS analysis-1

分析>迴歸>二元Logistic

Logistic regression SPSS analysis-2

Logistic regression SPSS output

Variables in the Equation

OR

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	Age	.065	.042	2.341	1	.126	1.067	.982	1.159
	Sex(1)	-1.051	.639	2.706	1	.100	.350	.100	1.223
	IOP.nct	-.042	.095	.198	1	.656	.959	.796	1.154
	CCT	-.004	.011	.149	1	.700	.996	.974	1.018
	CBiF	-3.534	1.127	9.829	1	.002	.029	.003	.266
	ACV	-.052	.011	20.907	1	.000	.950	.929	.971
	Constant	26.788	8.046	11.084	1	.001	4.304E+11		

a. Variable(s) entered on step 1: Age, Sex, IOP.nct, CCT, CBiF, ACV.

經多變項調整後，
CBiF每增加1個單位罹患青光眼的風險降低 0.029 倍
且達統計差異 (p=0.002)。

Logistic regression example-2

- 探討Brain Image對於大腦皮質下區域失智症的影響
  - Adjusted model
    - age, sex and CCI

Frontiers in Neurology 12 (2021)

TABLE 5 | Associations of imaging variables with subcortical vascular dementia, $N = 57$.

	Dementia	
	OR (95%CI)	p
MARS		
Infratentorial	1.00 (0.93–1.07)	0.963
Deep	1.03 (0.99–1.08)	0.161
Lobar	1.00 (0.99–1.02)	0.739
Total	1.00 (0.99–1.01)	0.580
ARWMC		
Infratentorial	0.53 (0.10–2.80)	0.458
Basal ganglia	1.67 (0.74–3.78)	0.221
Subcortical	2.03 (1.24–3.32)	0.005*
Total	1.43 (1.09–1.89)	0.011*
PVSE		
Centrum semiovale	0.72 (0.36–1.42)	0.339
Basal ganglia	1.10 (0.44–2.74)	0.837
Lesion quantity		
ICH	1.00 (0.75–1.34)	1.000
Lacune	1.18 (1.02–1.35)	0.023*
Lesion burden score		
CAA-SVD score	2.33 (1.01–5.40)	0.047*
C1[†]	1.41 (1.09–1.83)	0.009*
C2[‡]	1.38 (1.08–1.76)	0.010*

ICH, intracerebral hemorrhage; MARS, Microbleed anatomical rating scale; ARWMC, Age-related White Matter Change; PVSE, enlargement of the perivascular space; CAA, cerebral amyloid angiopathy; SVD, small vessel disease; OR, odd's ratio.

ORs determined by multivariate logistic regression, adjusted for age, sex, and CCI; *$p < 0.05$.

[†]Sum of total MARS score and total ARWMC scale.

[‡]Sum of total MARS score, total ARWMC scale, BG PVSE ($\geq$20), and lacune ($\geq$5).

Logistic regression SPSS dataset

	caseno	Dementia	CCI	Gender	Age	Infratentor_MARS	Deep_MARS	Lobar_MARS	Total_MARS
1	1	1	4	1	87	6	4	11	21
2	2	0	6	1	99	.	.	.	.
3	3	0	5	1	72	0	2	10	12
4	4	0	2	0	82	14	10	17	41
5	5	0	1	1	80	1	7	5	13
6	6	0	3	1	54	12	21	120	153
7	7	1	2	1	64	0	12	56	68
8	8	1	4	1	61	6	5	1	12
9	9	0	5	1	72	37	43	176	256
10	10	0	3	1	63	1	1	1	3
11	11	0	1	0	89	3	3	5	11
12	12	0	2	1	88	8	19	5	32

Logistic regression SPSS analysis

Logistic regression SPSS output

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	CCI	-.119	.295	.163	1	.687	.888	.498	1.583
	Gender(1)	.287	.824	.122	1	.727	1.333	.265	6.700
	Age	.005	.036	.022	1	.883	1.005	.936	1.079
	Subcortical_ARWMC	.708	.251	7.920	1	.005	2.029	1.240	3.322
	Constant	-5.365	3.040	3.114	1	.078	.005		

a. Variable(s) entered on step 1: CCI, Gender, Age, Subcortical_ARWMC.

經性別、年齡和**CCI**調整後，
Subcortical_ARWMC每增加**1**個單位罹患失智症的風險增加 **2.029** 倍
且達統計差異 **(p=0.005)**。

Multinomial Logistic Regression SPSS dataset

- 依變項為類別變數，而且超過兩種以上
- 自變項為類別變數或連續變數
- 探討基因突變對不同組織類型的肺癌發生風險的關係

Value Labels — type

Value: ___
Label: ___

0 = "Control"
1 = "Adenocarcinoma"
2 = "SCC"
3 = "LCC"

Add | Change | Remove

OK | Cancel | Help

Value Labels — gene

Value: ___
Label: ___

0 = "Positive"
1 = "Negative"

Add | Change | Remove

OK | Cancel | Help
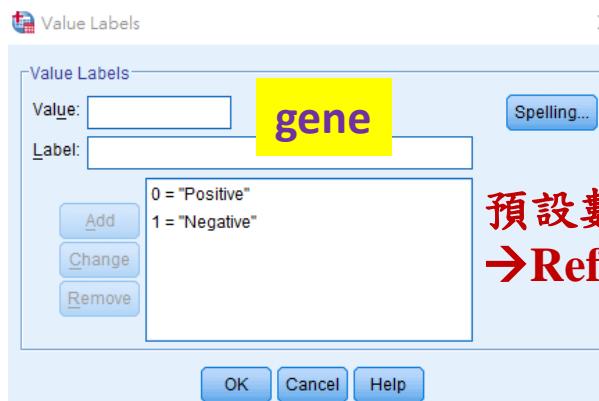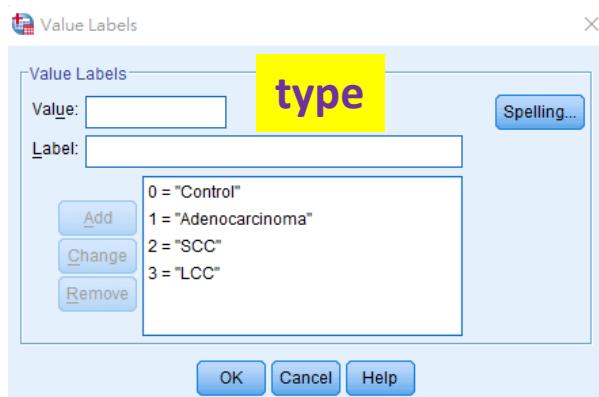
預設數值大
→Reference

	ID	type	gene
1	1	1	0
2	2	1	0
3	3	1	0
4	4	1	0
5	5	1	0
6	6	1	0
7	7	1	0
8	8	1	0
9	9	1	0
10	10	1	0
11	11	1	0
12	12	1	0
13	13	1	0
14	14	1	0
15	15	1	0

Multinomial Logistic Regression SPSS analysis-1

分析>迴歸>多項式**Logistic**

Multinomial Logistic Regression SPSS analysis-2

Multinomial Logistic Regression SPSS output

Parameter Estimates

OR

type[a]		B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp (B)	
								Lower Bound	Upper Bound
Adenocarcinoma	Intercept	-1.769	.279	40.117	1	.000			
	[gene=0]	2.686	.441	37.044	1	.000	14.667	6.177	34.827
	[gene=1]	0[b]	.	.	0	.	.	.	.
SCC	Intercept	-.788	.191	17.096	1	.000			
	[gene=0]	1.299	.412	9.948	1	.002	3.667	1.635	8.221
	[gene=1]	0[b]	.	.	0	.	.	.	.
LCC	Intercept	-1.587	.259	37.634	1	.000			
	[gene=0]	2.193	.442	24.576	1	.000	8.963	3.766	21.331
	[gene=1]	0[b]	.	.	0	.	.	.	.

a. The reference category is: Control.

b. This parameter is set to zero because it is redundant.

有基因突變比沒有基因突變，
發生風險分別為腺癌有 **14.667** 倍、鱗狀細胞癌有 **3.667** 倍、
大細胞癌有 **8.963** 倍且皆達統計差異。

From log odds to probability

Here's the equation of a logistic regression model with 1 predictor X:

a.k.a. **Log Odds** or **Logit**

Intercept

$$\log\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 X$$

Where **P is the probability** of having the outcome and **P / (1-P) is the odds** of the outcome.

When X = 0, the intercept β_0 is the log of the odds of having the outcome

From log odds to probability

$$\ln\frac{p}{1-p} = \beta_0 + \beta_1 X \quad \Longrightarrow \quad \frac{P}{1-P} = e^{\beta_0 + \beta_1 X}$$

$$P = e^{\beta_0 + \beta_1 X} - Pe^{\beta_0 + \beta_1 X}$$

$$P(1 + e^{\beta_0 + \beta_1 X}) = e^{\beta_0 + \beta_1 X}$$

$$\boxed{P = \frac{e^{\beta_0 + \beta_1 X}}{1 + e^{\beta_0 + \beta_1 X}}}$$

# SPSS dataset probability
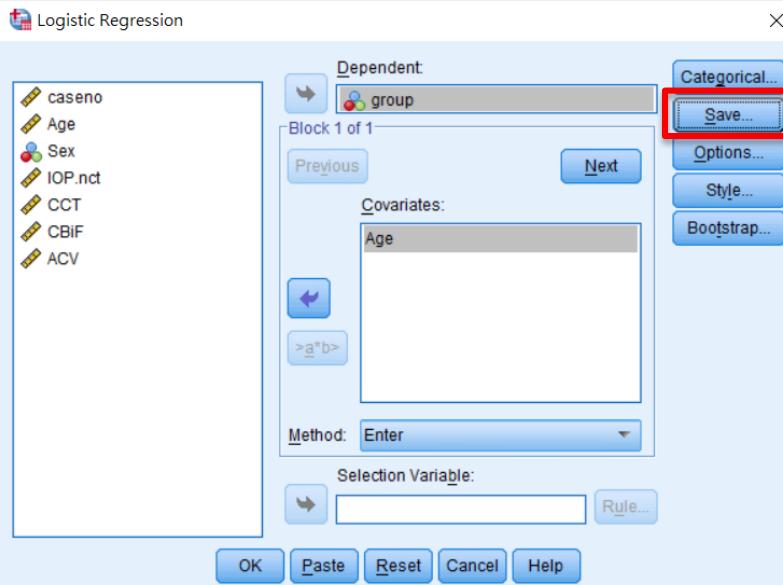
➤ SPSS dataset probability

| caseno	group	Age	PRE_1
1	0	82.0	.84802
2	1	77.0	.75962
3	1	75.0	.71569
4	1	76.0	.73825
5	0	72.0	.64155
6	0	73.0	.66725
7	1	69.0	.55995
8	0	79.0	.79867
9	1	68.0	.53177
10	1	68.0	.53177
11	1	78.0	.77977
12	0	62.0	.36471
13	0	70.0	.58775
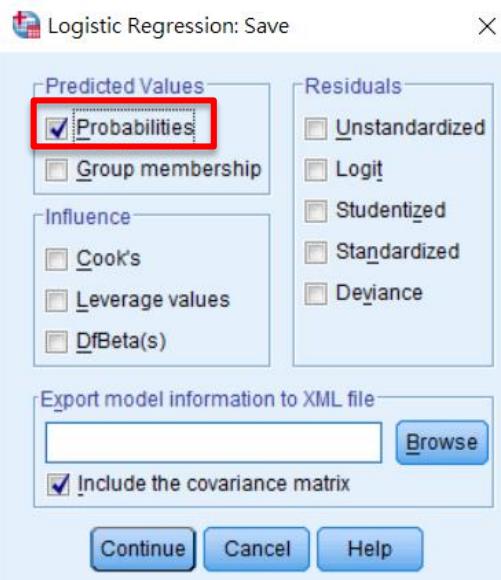
ROC curve

- 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)
    - 比較不同Biomarker (Aβ42/t-tau) 預測認知功能有無下降
- 目的
  - 將連續數值決定臨界值
  - 用來比較不同工具的好壞
- 最適切點 (Youden Index = Sensitivity + Specificity -1)
  - Sensitivity + Specificity 最大的點
  - 0-1 之間，越接近 1 越好

AUC	Discrimination
0.5	沒有鑑別力
0.7+	可接受的鑑別力
0.8+	好的鑑別力
0.9+	非常好的鑑別力

Sensitivity: 78.0
Specificity: 70.8
Criterion : >382.68

100-Specificity

21

Area under the ROC Curve

AUC = 0.9

AUC = 0.8

AUC = 0.7

AUC = 0.5
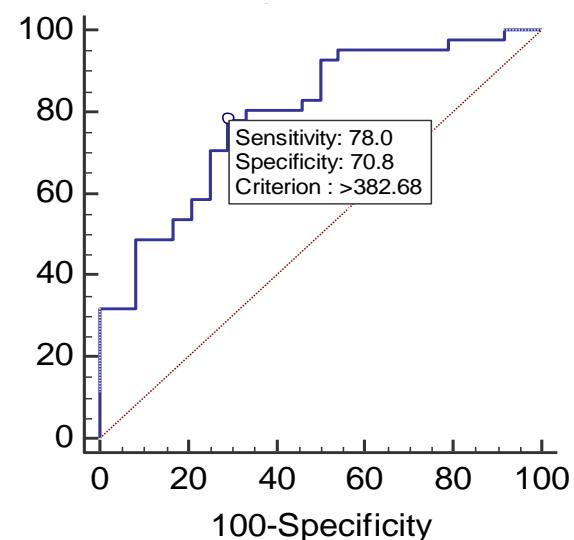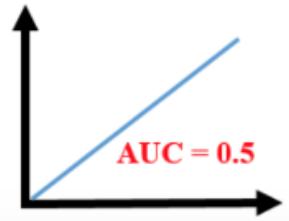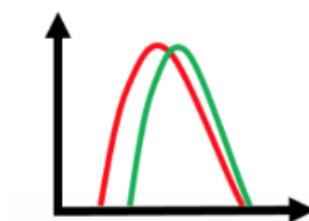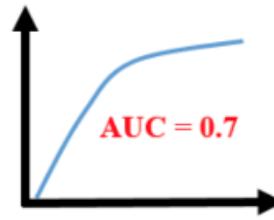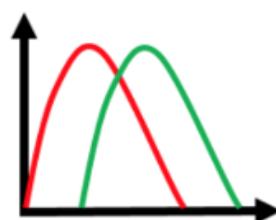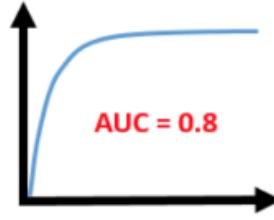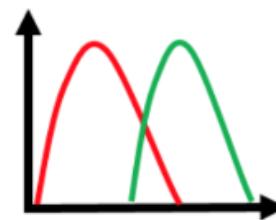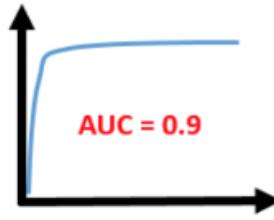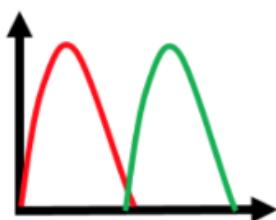
診斷工具指標

	Gold Standard Test		Total
	Disease	Control	
Screen test			
Positive	a	b	a+b
Negative	c	d	c+d
Total	a+c	b+d	a+b+c+d

Ref: https://www.medcalc.org/manual/roc-curves.php

$$Sensitivity = \frac{a}{a+c}$$

$$Specificity = \frac{d}{b+d}$$
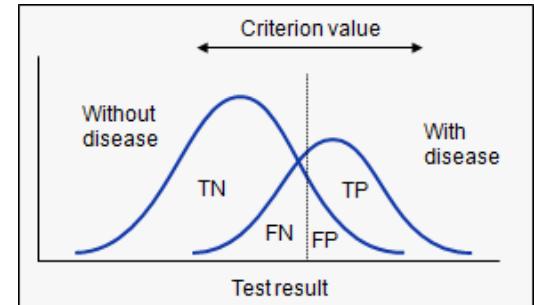
$$False\ positive = \frac{b}{b+d}$$

$$False\ negative = \frac{c}{a+c}$$

$$Positive\ Predictive\ Value = \frac{a}{a+b}$$

$$Negative\ Predictive\ Value = \frac{d}{c+d}$$

$$Likelihood\ ratio\ positive = \frac{Sensitivity}{1-specificity}$$

$$Likelihood\ ratio\ negative = \frac{1-sensitivity}{specificity}$$

Likelihood ratio

概似比的臨床意義	
Likelihood ratio	Interpretation
>10	Strong evidence to rule in disease
5~10	Moderate evidence to rule in disease
2~5	Weak evidence to rule in disease
0.5~2.0	No signification change in the likelihood of disease
0.2~0.5	Weak evidence to rule out disease
0.1~0.2	Moderate evidence to rule out disease
<0.1	Strong evidence to rule out disease

Source: Sackett, Richardson, Rosenberg, Haynes. Evidence-Based Medicine: How to Practice and Teach EBM. Churchill Livingstone, London, 1997.

ROC example

- 利用高光譜影像評估皮膚狀況是否預測硬皮症

TABLE 3 ROC analysis of skin assessments in the diagnosis of scleroderma

Variables	AUC (95% CI)	P	Optimal cutoff	Sensitivity, %	Specificity, %	Accuracy, %	PPV, %	NPV, %
Skin scores[a]	0.712 (0.658, 0.763)	<0.001	>0	42.47	100.00	64.33	100.00	51.58
Strain ratio	0.522 (0.464, 0.580)	0.510	≤0.88	21.51	92.11	48.33	81.63	41.83
Thickness	0.585 (0.527, 0.642)	0.009	>1.31	38.17	82.46	55.00	78.02	44.98
SAM	0.812 (0.763, 0.854)	<0.001	>0.025	61.83	84.21	70.33	86.47	57.49

FIG. 3 Comparisons of scleroderma diagnostic accuracy among tests
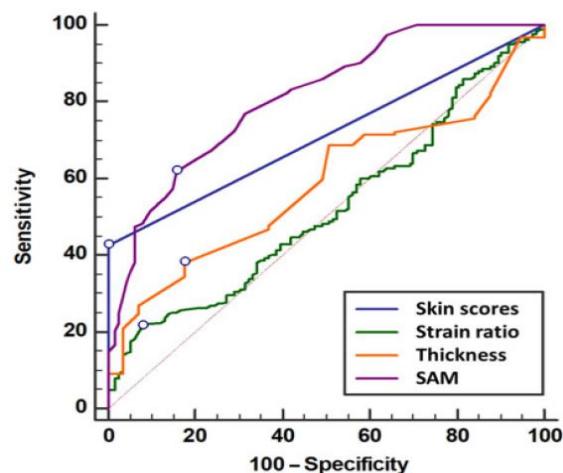
*Rheumatology 59.11 (2020): 3201-3210.*

ROC SPSS dataset

|	caseno	scleroderma	Skin_scores	thickness_mean	SAM	Strainratio
1	1	1.00	.00	.175	.016	1.02
2	2	1.00	.00	.129	.014	5.07
3	3	1.00	.00	.142	.022	4.53
4	4	1.00	.00	.127	.011	1.12
5	5	1.00	.00	.131	.015	1.45
6	6	1.00	.00	.148	.014	7.93
7	7	1.00	2.00	.255	.041	.19
8	8	1.00	2.00	.253	.038	3.16
9	9	1.00	2.00	.216	.038	6.15
10	10	1.00	2.00	.216	.044	1.91
11	11	1.00	2.00	.203	.040	6.77
12	12	1.00	2.00	.161	.012	5.81

ROC SPSS analysis

分析>ROC曲線
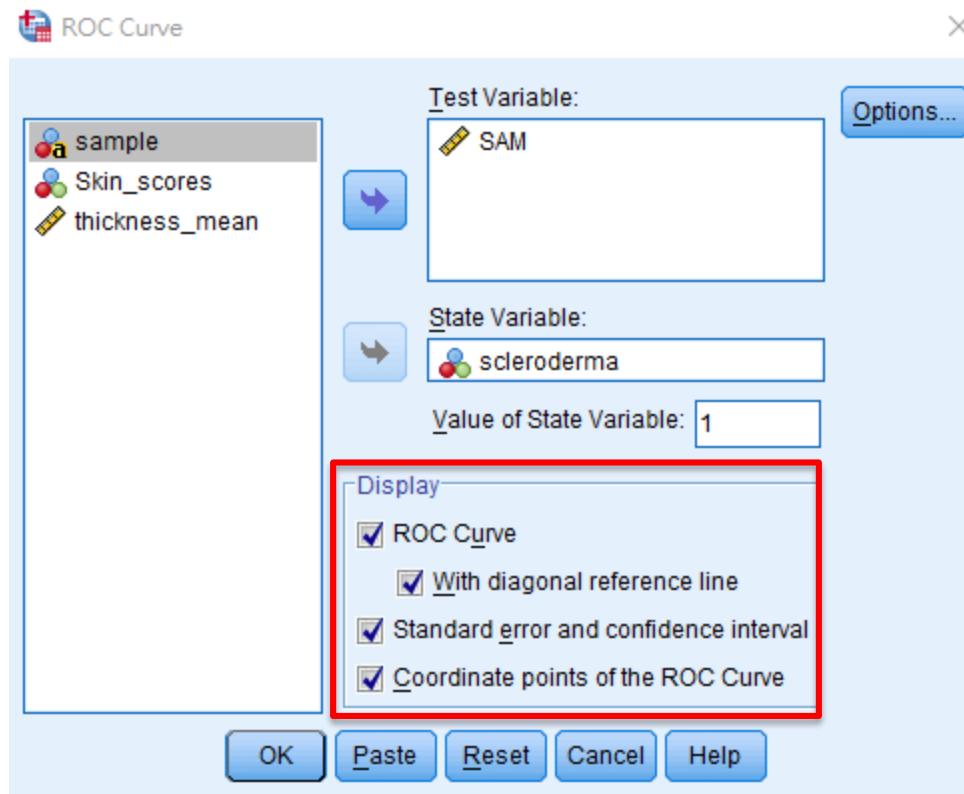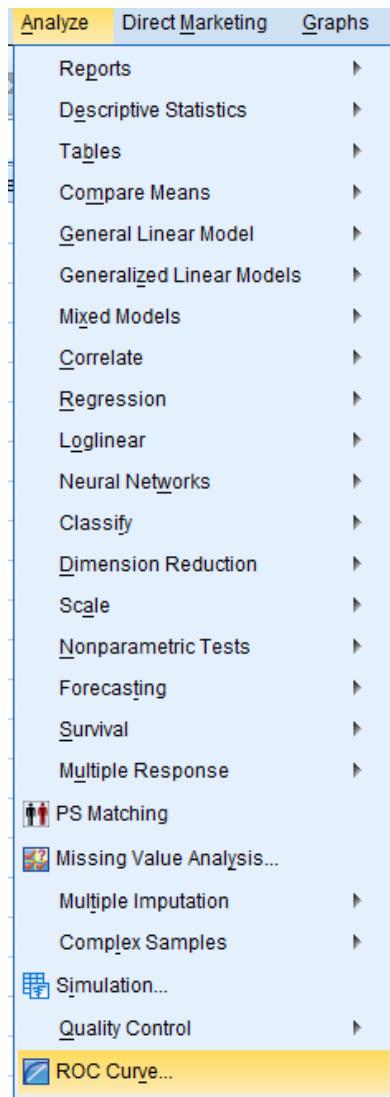
ROC SPSS output

AUC
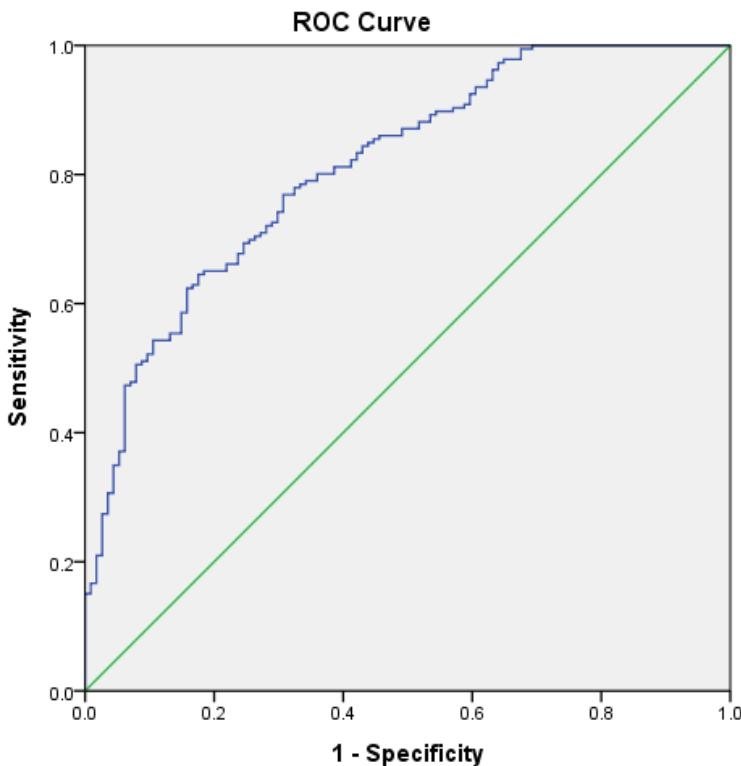
Area Under the Curve

Test Result Variable(s): SAM

Area	Std. Error[a]	Asymptotic Sig.[b]	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.812	.025	.000	.764	.861

a. Under the nonparametric assumption

b. Null hypothesis: true area = 0.5

Coordinates of the Curve

Test Result Variable(s): SAM

Positive if Greater Than or Equal To[a]	Sensitivity	1 - Specificity
.0000	1.000	1.000
.0045	1.000	.991
.0055	1.000	.974
.0065	1.000	.930
.0075	1.000	.886
.0085	1.000	.833
.0095	1.000	.789
.0105	1.000	.711
.0115	.989	.675
.0125	.973	.640
.0135	.935	.614
.0145	.903	.579
.0155	.892	.544
.0165	.860	.491
.0175	.833	.421
.0185	.823	.412
.0195	.769	.316
.0205	.726	.289
.0215	.715	.281

Youden Index

- Youden Index = sen + spe - 1
- 新增spe欄位, 公式為1 - (1-specificity)
- 將 sen + spe − 1 欄 由大至小排序

Coordinates of the Curve

Test Result \ SAM

Positive if Greater Than or Equal To[a]	Sensitivity	1 - Specificity	spe	sen+spe-1
.0255	.618	.158	0.842	0.460
.0195	.769	.316	0.684	0.453
.0245	.651	.202	0.798	0.449
.0235	.661	.219	0.781	0.442
.0205	.726	.289	0.711	0.436
.0225	.672	.237	0.763	0.435
.0215	.715	.281	0.719	0.434
.0265	.575	.149	0.851	0.426
.0285	.516	.096	0.904	0.420
.0275	.548	.132	0.868	0.417
.0175	.833	.421	0.579	0.412
.0305	.473	.061	0.939	0.412
.0185	.823	.412	0.588	0.410
.0295	.484	.079	0.921	0.405
.0315	.435	.061	0.939	0.374
.0165	.860	.491	0.509	0.369
.0325	.419	.061	0.939	0.358

Medcalc ROC-1

分析>**ROC curves**>**ROC curve analysis**   https://www.medcalc.org/download/

Medcalc ROC-2

Medcalc ROC-2

Variable	SAM
Classification variable	scleroderma

Sample size		300
Positive group :	scleroderma = 1	186
Negative group :	scleroderma = 0	114

Disease prevalence (%)	62

Area under the ROC curve (AUC)

Area under the ROC curve (AUC)	0.812
Standard Error[a]	0.0248
95% Confidence interval[b]	0.763 to 0.854
z statistic	12.560
Significance level P (Area=0.5)	<0.0001

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

Youden index

Youden index J	0.4604
Associated criterion	>0.025
Sensitivity	61.83
Specificity	84.21

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
>0.025	61.83	54.4 - 68.8	84.21	76.2 - 90.4	3.92	2.5 - 6.1	0.45	0.4 - 0.6	86.5	79.5 - 91.8	57.5	49.6 - 65.1
>0.026	57.53	50.1 - 64.7	85.09	77.2 - 91.1	3.86	2.4 - 6.1	0.50	0.4 - 0.6	86.3	79.0 - 91.8	55.1	47.4 - 62.6
>0.027	54.84	47.4 - 62.1	86.84	79.2 - 92.4	4.17	2.6 - 6.8	0.52	0.4 - 0.6	87.2	79.7 - 92.6	54.1	46.6 - 61.5
>0.028	51.61	44.2 - 59.0	90.35	83.4 - 95.1	5.35	3.0 - 9.5	0.54	0.5 - 0.6	89.7	82.3 - 94.8	53.4	46.1 - 60.6
>0.029	48.39	41.0 - 55.8	92.11	85.5 - 96.3	6.13	3.2 - 11.7	0.56	0.5 - 0.7	90.9	83.4 - 95.8	52.2	45.1 - 59.3
>0.03	47.31	40.0 - 54.7	93.86	87.8 - 97.5	7.71	3.7 - 16.0	0.56	0.5 - 0.6	92.6	85.4 - 97.0	52.2	45.1 - 59.2
>0.031	43.55	36.3 - 51.0	93.86	87.8 - 97.5	7.09	3.4 - 14.8	0.60	0.5 - 0.7	92.0	84.3 - 96.7	50.5	43.5 - 57.4
>0.032	41.94	34.8 - 49.4	93.86	87.8 - 97.5	6.83	3.3 - 14.3	0.62	0.5 - 0.7	91.8	83.8 - 96.6	49.8	42.9 - 56.6
>0.033	38.17	31.2 - 45.6	93.86	87.8 - 97.5	6.22	3.0 - 13.0	0.66	0.6 - 0.7	91.0	82.4 - 96.3	48.2	41.5 - 55.0
>0.034	36.02	29.1 - 43.4	94.74	88.9 - 98.0	6.84	3.1 - 15.3	0.68	0.6 - 0.8	91.8	83.0 - 96.9	47.6	40.9 - 54.3
>0.035	33.33	26.6 - 40.6	95.61	90.1 - 98.6	7.60	3.1 - 18.3	0.70	0.6 - 0.8	92.5	83.4 - 97.5	46.8	40.2 - 53.4
>0.036	29.03	22.6 - 36.1	96.49	91.3 - 99.0	8.27	3.1 - 22.2	0.74	0.7 - 0.8	93.1	83.3 - 98.1	45.5	39.1 - 52.0
>0.037	25.27	19.2 - 32.1	97.37	92.5 - 99.5	9.60	3.1 - 30.1	0.77	0.7 - 0.8	94.0	83.5 - 98.7	44.4	38.1 - 50.8
>0.038	22.58	16.8 - 29.3	97.37	92.5 - 99.5	8.58	2.7 - 27.0	0.80	0.7 - 0.9	93.3	81.7 - 98.6	43.5	37.4 - 49.9
>0.039	20.43	14.9 - 26.9	98.25	93.8 - 99.8	11.65	2.9 - 47.4	0.81	0.8 - 0.9	95.0	83.1 - 99.4	43.1	37.0 - 49.3
>0.04	17.74	12.5 - 24.0	98.25	93.8 - 99.8	10.11	2.5 - 41.3	0.84	0.8 - 0.9	94.3	80.8 - 99.3	42.3	36.2 - 48.5

Medcalc ROC-3

分析>ROC curves> Comparison of ROC curves

Medcalc ROC-4

Medcalc ROC-5

Pairwise comparison of ROC curves

Skin_scores ~ Strainratio	
Difference between areas	0.190
Standard Error[c]	0.0371
95% Confidence Interval	0.117 to 0.263
z statistic	5.126
Significance level	P < 0.0001
Skin_scores ~ thickness_mean	
Difference between areas	0.127
Standard Error[c]	0.0318
95% Confidence Interval	0.0647 to 0.189
z statistic	3.996
Significance level	P = 0.0001
Skin_scores ~ SAM	
Difference between areas	0.0997
Standard Error[c]	0.0283
95% Confidence Interval	0.0443 to 0.155
z statistic	3.527
Significance level	P = 0.0004
Strainratio ~ thickness_mean	
Difference between areas	0.0631
Standard Error[c]	0.0523
95% Confidence Interval	-0.0394 to 0.165
z statistic	1.207
Significance level	P = 0.2276
Strainratio ~ SAM	
Difference between areas	0.290
Standard Error[c]	0.0427
95% Confidence Interval	0.206 to 0.374
z statistic	6.785
Significance level	P < 0.0001
thickness_mean ~ SAM	
Difference between areas	0.227
Standard Error[c]	0.0399
95% Confidence Interval	0.149 to 0.305
z statistic	5.681
Significance level	P < 0.0001

[c] DeLong et al., 1988

臺中榮民總醫院
Taichung Veterans General Hospital

中榮官網　中榮FB

11/14問卷

感謝您的聆聽！
Thank you！

愛心 · 品質 · 創新 · 當責
Compassion　Quality　Innovation　Accountability