

Creating an ROC curve and identify the optimal threshold value for a detection method

[demoROC.xls](#)

ROC curve

This method was first developed during World War II to develop effective means of detecting Japanese aircrafts. It was then applied more generally to signal detection and medicine where it is now widely used.

The problem is as follows: we study a phenomenon, often binary (for example, the presence or absence of a disease) and we want to develop a test to detect effectively the occurrence of a precise event (for example, the presence of the disease).

If the test is quantitative (possibly ordinal), for example, a concentration of a molecule, we will try to determine from what concentration can a patient be considered as ill. The ROC curves and indices calculated here will help us make the best decision.

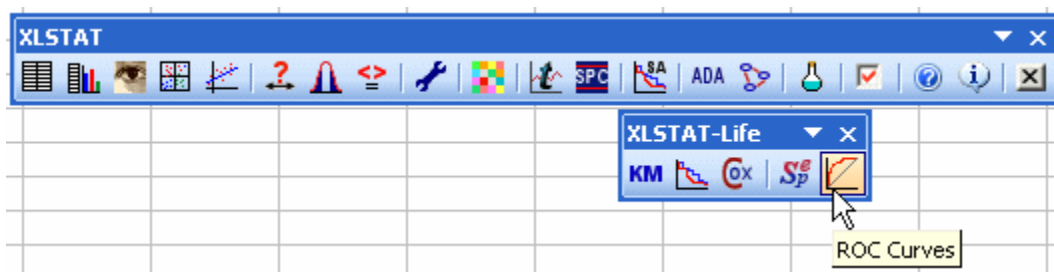
Dataset to generate a ROC curve

An Excel sheet with both the data and results used in this tutorial can be downloaded by clicking [here](#).

The data correspond to a medical experiment during which 50 patients, among which 20 are sick, are submitted to a screening test where the concentration of a viral molecule is being measured.

Setting up of a ROC curve

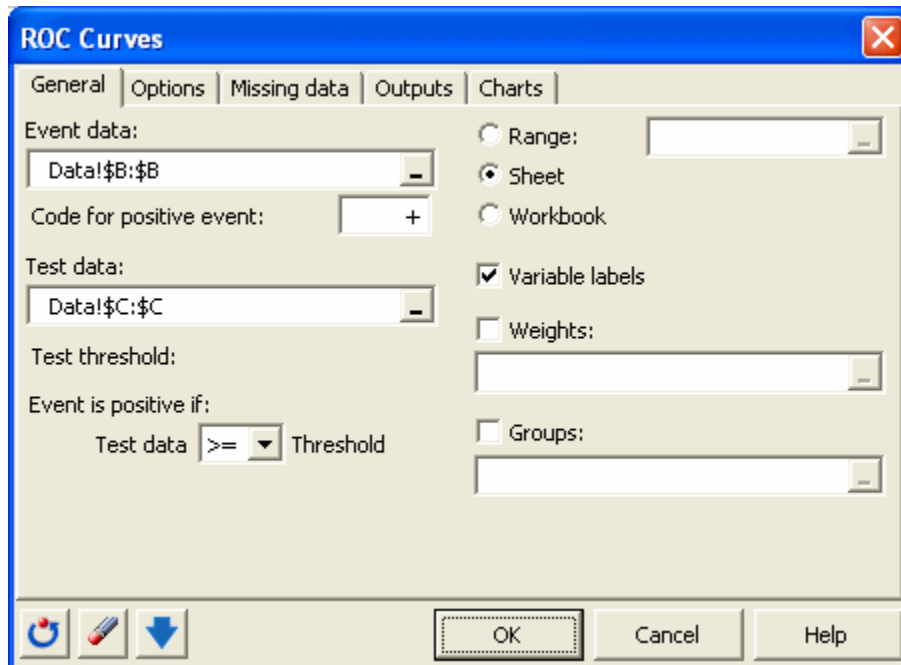
Once XLSTAT has been started, select the **XLSTAT-Life / ROC Curves** command, or click on the corresponding button of the **XLSTAT-Life** toolbar (see below).



When you click on the button, a dialog box appears. Select the data that correspond to the **event data** and enter the code that is associated to positive cases.

Then select the data that correspond to the diagnostic test that we are evaluating and specify what type of rule should be used to identify the threshold value below or above which the test should be considered positive.

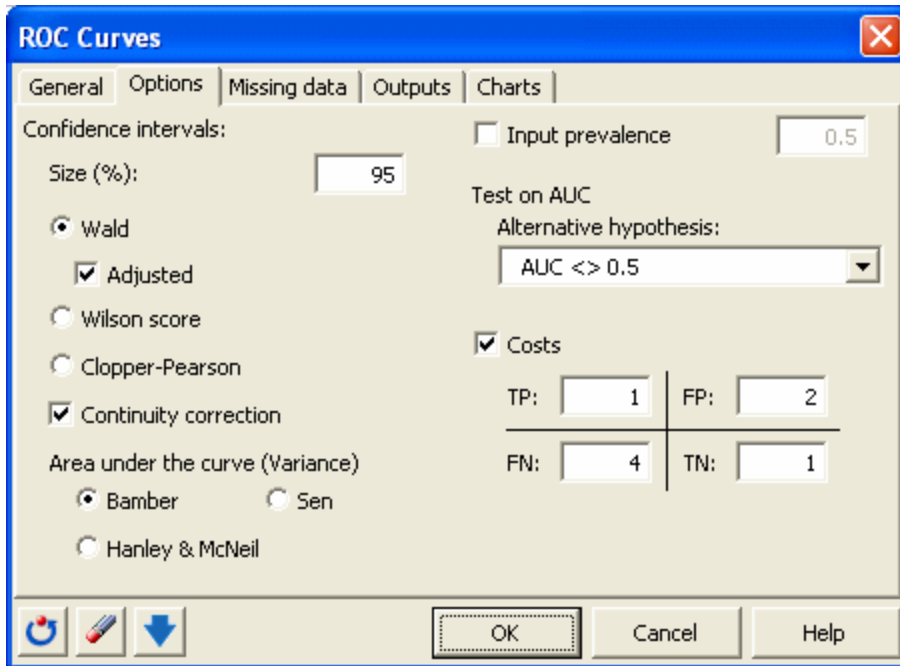
We choose here to consider that the **test is positive if the concentration is greater than or equal to a value to be determined.**



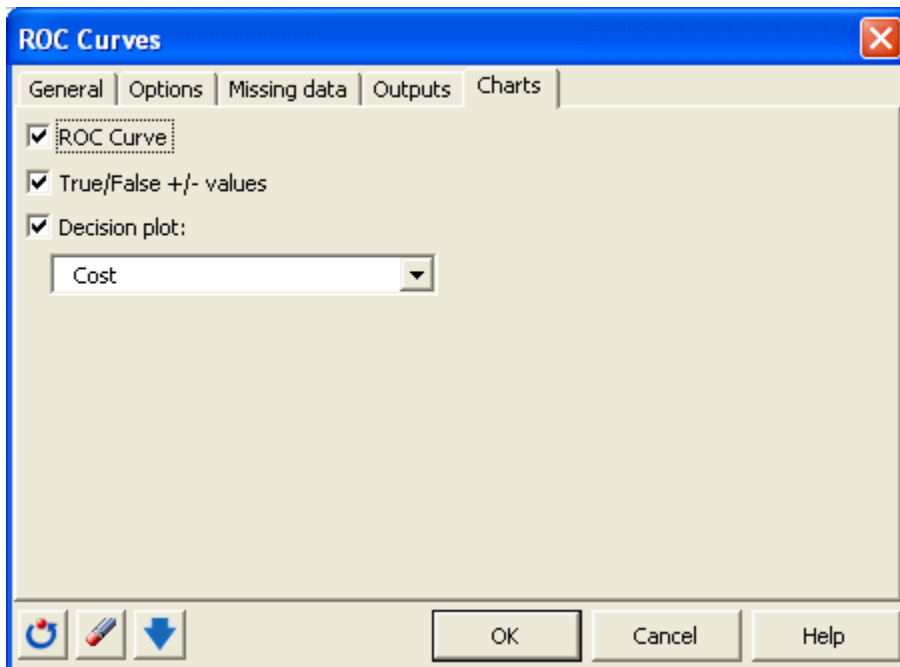
In the **Options** tab, you can specify the method for calculating the confidence intervals.

XLSTAT is the software offering the widest choice. The defaults are those most recommended.

In this tab, you can also assign a cost to the various cases. We wish to strongly penalize diagnostic errors and more particularly the case where the sick patients are not detected.



In the **Charts** tab, we choose to display a decision plot based on costs.



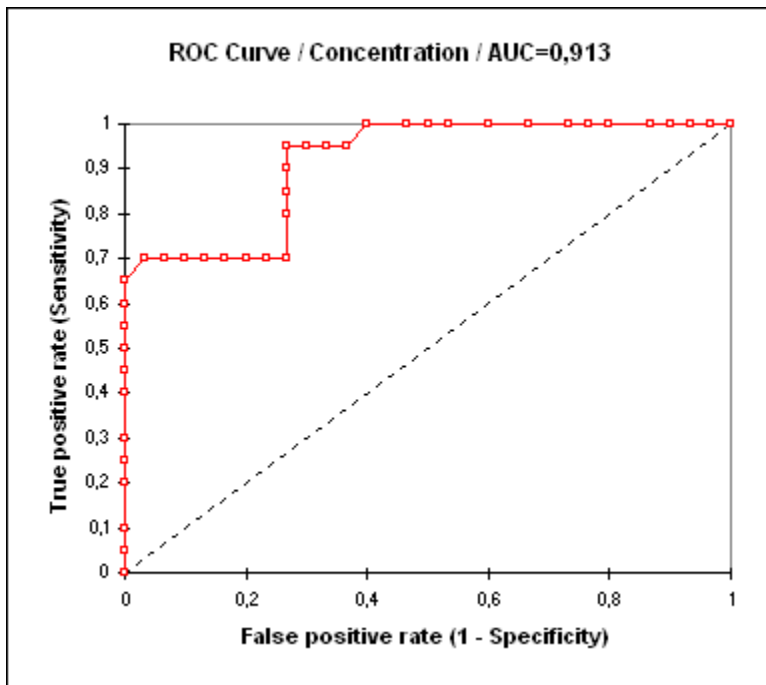
When you click **OK**, the computations are done and the results are displayed.

Interpretation of a ROC curve

The first table displays the descriptive statistics for the test variable, here the concentration, followed by the statistics for the event variable, here the disease. The prevalence observed on the sample is equal to 0.4.

Summary statistics:							
Variable	Observation	with miss	without miss	Minimum	Maximum	Mean	Std. deviation
Concentration	50	0	50	0,520	1,500	1,029	0,313
Disease	Frequency	%					
+	20	40%					
-	30	60%					
Prevalence	0,400	40%					

The ROC is then displayed. To each small square corresponds an observation.

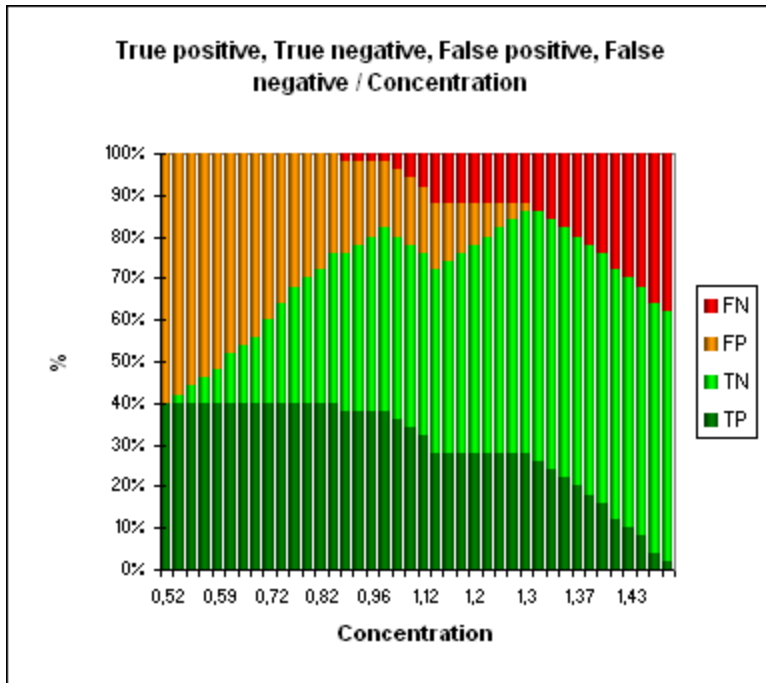


The "ROC Analysis" that follows, displays for each possible threshold value, the value of the various performance indices. For example, if we decide to declare a patient sick when the concentration is greater than or equal to 0.98, the sensitivity is 0.95, the specificity of 0.733 and the cost is 61. For more details on the various indices, you can check the tutorial on [sensitivity and specificity analysis](#).

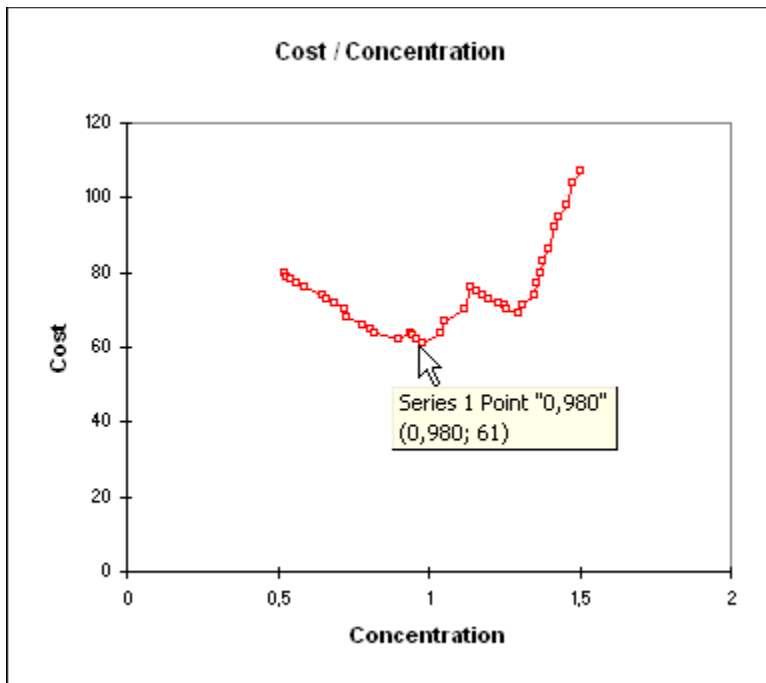
ROC analysis:															
Concentration	Sensitivity	Se bound	Se bound (3)	Specificity	Sp bound	Sp bound (3)	Cost	PPV	NPV	LR+	LR-	TP	TN	FP	FN
0,520	1,000	0,806	1,000	0,000	0,000	0,138	80	0,400		1,000		20	0	30	0
0,530	1,000	0,806	1,000	0,033	0,000	0,184	79	0,408	1,000	1,034	0,000	20	1	29	0
0,540	1,000	0,806	1,000	0,067	0,009	0,226	78	0,417	1,000	1,071	0,000	20	2	28	0
0,560	1,000	0,806	1,000	0,100	0,028	0,266	77	0,426	1,000	1,111	0,000	20	3	27	0
0,590	1,000	0,806	1,000	0,133	0,048	0,305	76	0,435	1,000	1,154	0,000	20	4	26	0
0,650	1,000	0,806	1,000	0,200	0,093	0,378	74	0,455	1,000	1,250	0,000	20	6	24	0
0,660	1,000	0,806	1,000	0,233	0,116	0,413	73	0,465	1,000	1,304	0,000	20	7	23	0
0,690	1,000	0,806	1,000	0,267	0,141	0,447	72	0,476	1,000	1,364	0,000	20	8	22	0
0,720	1,000	0,806	1,000	0,333	0,192	0,514	70	0,500	1,000	1,500	0,000	20	10	20	0
0,730	1,000	0,806	1,000	0,400	0,246	0,577	68	0,526	1,000	1,667	0,000	20	12	18	0
0,780	1,000	0,806	1,000	0,467	0,303	0,638	66	0,556	1,000	1,875	0,000	20	14	16	0
0,810	1,000	0,806	1,000	0,500	0,332	0,668	65	0,571	1,000	2,000	0,000	20	15	15	0
0,820	1,000	0,806	1,000	0,533	0,362	0,697	64	0,588	1,000	2,143	0,000	20	16	14	0
0,900	1,000	0,806	1,000	0,600	0,423	0,754	62	0,625	1,000	2,500	0,000	20	18	12	0
0,940	0,950	0,743	1,000	0,633	0,454	0,781	64	0,633	0,950	2,591	0,079	19	19	11	1
0,950	0,950	0,743	1,000	0,667	0,486	0,808	63	0,655	0,952	2,650	0,075	19	20	10	1
0,960	0,950	0,743	1,000	0,700	0,519	0,834	62	0,679	0,955	3,167	0,071	19	21	9	1
0,980	0,950	0,743	1,000	0,733	0,563	0,859	61	0,704	0,957	3,563	0,068	19	22	8	1
1,040	0,900	0,684	0,982	0,733	0,553	0,859	64	0,692	0,917	3,375	0,136	18	22	8	2
1,050	0,850	0,629	0,954	0,733	0,553	0,859	67	0,680	0,880	3,188	0,205	17	22	8	3
1,120	0,800	0,577	0,923	0,733	0,553	0,859	70	0,667	0,846	3,000	0,273	16	22	8	4
1,140	0,700	0,478	0,855	0,733	0,553	0,859	76	0,636	0,786	2,625	0,409	14	22	8	6
1,160	0,700	0,478	0,855	0,767	0,587	0,884	75	0,667	0,793	3,000	0,391	14	23	7	6
1,180	0,700	0,478	0,855	0,800	0,622	0,907	74	0,700	0,800	3,500	0,375	14	24	6	6
1,200	0,700	0,478	0,855	0,833	0,658	0,930	73	0,737	0,806	4,200	0,360	14	25	5	6
1,230	0,700	0,478	0,855	0,867	0,695	0,952	72	0,778	0,813	5,250	0,346	14	26	4	6
1,250	0,700	0,478	0,855	0,900	0,734	0,972	71	0,824	0,818	7,000	0,333	14	27	3	6
1,260	0,700	0,478	0,855	0,933	0,774	0,991	70	0,875	0,824	10,500	0,321	14	28	2	6
1,300	0,700	0,478	0,855	0,967	0,816	1,000	69	0,933	0,829	21,000	0,310	14	29	1	6
1,310	0,650	0,431	0,819	1,000	0,862	1,000	71	1,000	0,811	+Inf	0,350	13	30	0	7
1,350	0,600	0,386	0,781	1,000	0,862	1,000	74	1,000	0,789	+Inf	0,400	12	30	0	8
1,360	0,550	0,342	0,741	1,000	0,862	1,000	77	1,000	0,769	+Inf	0,450	11	30	0	9
1,370	0,500	0,300	0,700	1,000	0,862	1,000	80	1,000	0,750	+Inf	0,500	10	30	0	10
1,380	0,450	0,259	0,658	1,000	0,862	1,000	83	1,000	0,732	+Inf	0,550	9	30	0	11
1,400	0,400	0,219	0,614	1,000	0,862	1,000	86	1,000	0,714	+Inf	0,600	8	30	0	12
1,420	0,300	0,145	0,522	1,000	0,862	1,000	92	1,000	0,682	+Inf	0,700	6	30	0	14
1,430	0,250	0,110	0,474	1,000	0,862	1,000	95	1,000	0,667	+Inf	0,750	5	30	0	15
1,460	0,200	0,077	0,423	1,000	0,862	1,000	98	1,000	0,652	+Inf	0,800	4	30	0	16
1,480	0,100	0,018	0,316	1,000	0,862	1,000	104	1,000	0,625	+Inf	0,900	2	30	0	18
1,500	0,050	0,000	0,257	1,000	0,862	1,000	107	1,000	0,612	+Inf	0,950	1	30	0	19

Test is positive if Concentration >= threshold value

A graph constructed using this table is then displayed. He can see the evolution of counting TP (true positives), TN (true negative), FP (false positives) and FN (false negatives) depending on the chosen threshold value.



The decision plot allows you to choose the threshold value that minimizes the cost. To see what the minimum threshold value on the chart is, just leave your mouse over the corresponding point. This value corresponds to a concentration of 0.98, as we had identified earlier in the table analysis ROC.



The last series of results allows the study the area under the curve (AUC). The AUC and its confidence interval are calculated. The test that compares the AUC to 0.5 allows to check if the

diagnostic test is more powerful than just a random rule. In our case, the test is very powerful and the AUC is significantly different from 0.5.

Area under the curve (AUC):			
AUC	Standard error	Lower bound	Upper bound (95%)
0,913	0,032	0,851	0,976
Comparison of the AUC to 0.5:			
95% confidence interval on the difference between the AUC and 0.5 (Two-tailed test):			
] 0,351; 0,476 [
Difference	0,413		
z (Observed)	12,895		
z (Critical)	1,960		
p-value (Two-tailed)	< 0,0001		
alpha	0,05		
Test interpretation:			
H0: The AUC is not significantly different from 0.5.			
Ha: The AUC is significantly different from 0.5.			
As the computed p-value is lower than the significance level alpha=0,05, one should reject the null hypothesis H0, and accept the alternative hypothesis Ha.			
The risk to reject the null hypothesis H0 while it is true is lower than 0,01%.			

The comparison of the AUC is also a way to compare different diagnostic tests. XLSTAT can compare as many tests as you like.