

# Chapter 46

## Comparing the Performance of Diagnostic Tests (650 and 588 Patients)

### 1 General Purpose

Both logistic regression and c-statistics can be used to evaluate the performance of novel diagnostic tests (see also Machine learning in medicine part two, Chap. 6, pp 45–52, Springer Heidelberg Germany, 2013, from the same authors). This chapter is to assess whether one method can outperform the other.

### 2 Schematic Overview of Type of Data Files

Outcome	predictor
binary	
.	.
.	.
.	.
.	.
.	.
.	.
.	.
.	.

---

This chapter was previously partly published in “Machine learning in medicine a complete overview”, Chap. 41, Springer Heidelberg Germany, 2015, from the same authors.

### 3 Primary Scientific Question

Is logistic regression with the odds of disease as outcome and test scores as covariate a better alternative to concordance (c)-statistics using the area under the curve of ROC (receiver operated characteristic) curves.

### 4 Data Sample One

In 650 patients with peripheral vascular disease a noninvasive vascular lab test was performed. The results of the first ten patients are underneath.

Presence of peripheral vascular disease (0 = no, 1 = yes)	Test score
.00	1,00
.00	1,00
.00	2,00
.00	2,00
.00	3,00
.00	3,00
.00	3,00
.00	4,00
.00	4,00
.00	4,00

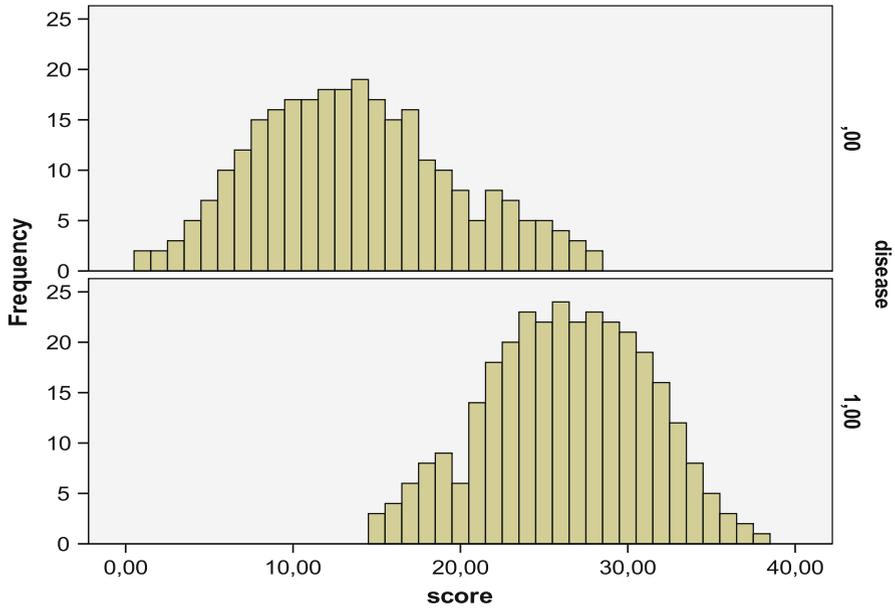
The entire data file is in [extras.springer.com](http://extras.springer.com), and is entitled “chapter46-1performancediagnostictest”. Start by opening the data file in SPSS.

### 5 Data Histogram Graph from Sample One

Then Command:

Graphs....Legacy Dialogs....Histogram....Variable(s): enter “score”....Row(s): enter “disease”....click OK.

The underneath figure shows the output sheet. On the x-axis we have the vascular lab scores, on the y-axis “how often”. The scores in patients with (1) and without (0) the presence of disease according to the gold standard (angiography) are respectively in the lower and upper graph.



## 6 Data Sample Two

The second data file is obtained from a parallel-group population of 588 patients after the noninvasive vascular test has been improved. The first ten patients are underneath.

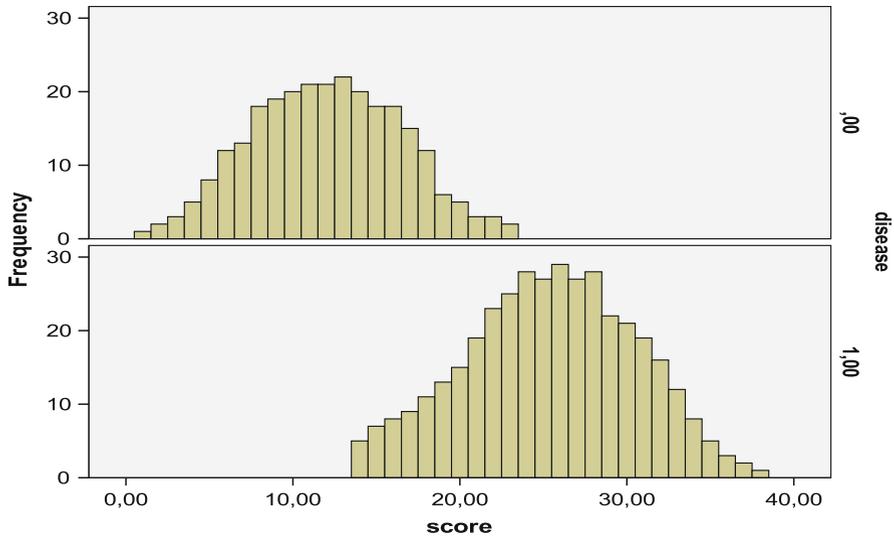
Presence of peripheral vascular disease (0 = no, 1 = yes)	Test score
,00	1,00
,00	2,00
,00	2,00
,00	3,00
,00	3,00
,00	3,00
,00	4,00
,00	4,00
,00	4,00
,00	4,00

The entire data file is in [extras.springer.com](http://extras.springer.com), and is entitled “chapter46-2performancediagnostictest”. Start by opening the data file in SPSS.

## 7 Data Histogram Graph from Sample Two

Command:

Graphs....Legacy Dialogs....Histogram....Variable(s): enter “score”....Row(s): enter “disease”....click OK.



The above figure is in the output sheet. The first test (upper figure) seems to perform less well than the second test (lower figure), because there may be more risk of false positives (the 0 disease curve is more skewed to the right in the upper than in the lower figure).

## 8 Performance Assessment with Binary Logistic Regression

For analysis the statistical model Binary Logistic Regression in the module Regression is required.

Binary logistic regression is used for assessing this question. The following reasoning is used. If we move the threshold for a positive test to the right, then the proportion of false positive will decrease. The steeper the logistic regression line the faster this will happen. In contrast, if we move the threshold to the left, the proportion of false negatives will decrease. Again, the steeper the logistic

regression line, the faster it will happen. And so, the steeper the logistic regression line, the fewer false negatives and false positives, and, thus, the better the diagnostic test.

For both data files the above analysis is performed, using the model Binary Logistic in the module Regression.

Command:

Analyze... Regression... Binary Logistic... Dependent variable: disease... Covariate: score...OK.

The output sheets show the best fit regression equations.

Variables in the equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 <sup>a</sup>	VAR00001	,398	,032	155,804	1	,000	1,488
	Constant	-8,003	,671	142,414	1	,000	,000

<sup>a</sup>Variable(s) entered on step 1: VAR00001

Variables in the equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 <sup>a</sup>	VAR00001	,581	,051	130,715	1	,000	1,789
	Constant	-10,297	,915	126,604	1	,000	,000

<sup>a</sup>Variable(s) entered on step 1: VAR00001

Data file 1: log odds of having the disease = -8.003 + 0.398 times the score

Data file 2: log odds of having the disease = -10.297 + 0.581 times the score.

The regression coefficient of data file 2 is much steeper than that of data file 1, 0.581 and 0.398.

Both regression equations produce highly significant regression coefficients with standard errors of respectively 0.032 and 0.051 and p-values of < 0.0001. The two regression coefficients are tested for significance of difference using the z – test (the z-test is in Chap. 2 of Statistics on a Pocket Calculator part 2, pp 3–5, Springer Heidelberg Germany, 2012, from the same authors):

$$z = (0.398 - 0.581) / \sqrt{(0.032^2 + 0.051^2)} = -0.183 / 0.060 = -3.05, \text{ which corresponds with a p-value of } < 0.01.$$

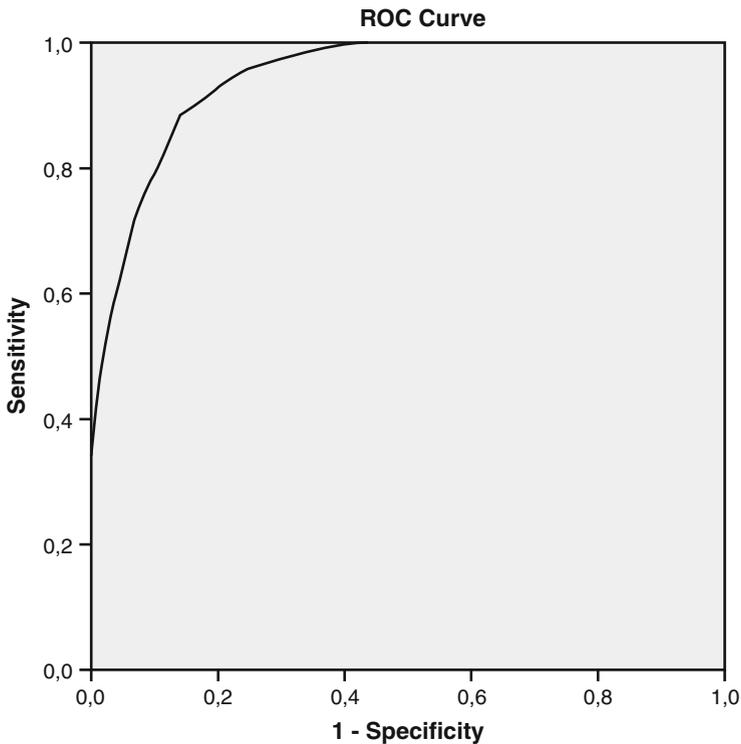
Obviously, test 2 produces a significantly steeper regression model, which means that it is a better predictor of the risk of disease than test 1. We can, additionally, calculate the odds ratios of successfully testing with test 2 versus test 1. The odds of disease with test 1 equals  $e^{0.398} = 1.488$ , and with test 2 it equals  $e^{0.581} = 1.789$ . The odds ratio =  $1.789 / 1.488 = 1.202$ , meaning that the second test produces a 1.202 times better chance of rightly predicting the disease than test 1 does.

## 9 Performance Assessment with C-statistics

C-statistics is used as a contrast test. Open the first data file again. For analysis the module ROC Curve must be used.

Command:

Analyze....ROC Curve....Test Variable: enter “score”....State Variable: enter “dis-  
ease”....Value of State Variable: type “1”....mark ROC Curve....mark Standard  
Error and Confidence Intervals....click OK.



Diagonal segments are produced by ties.

Area under the curve  
Test result variable(s): score

Area	Std. error <sup>a</sup>	Asymptotic sig. <sup>b</sup>	Asymptotic 95 % confidence interval	
			Lower bound	Upper bound
,945	,009	,000	,928	,961

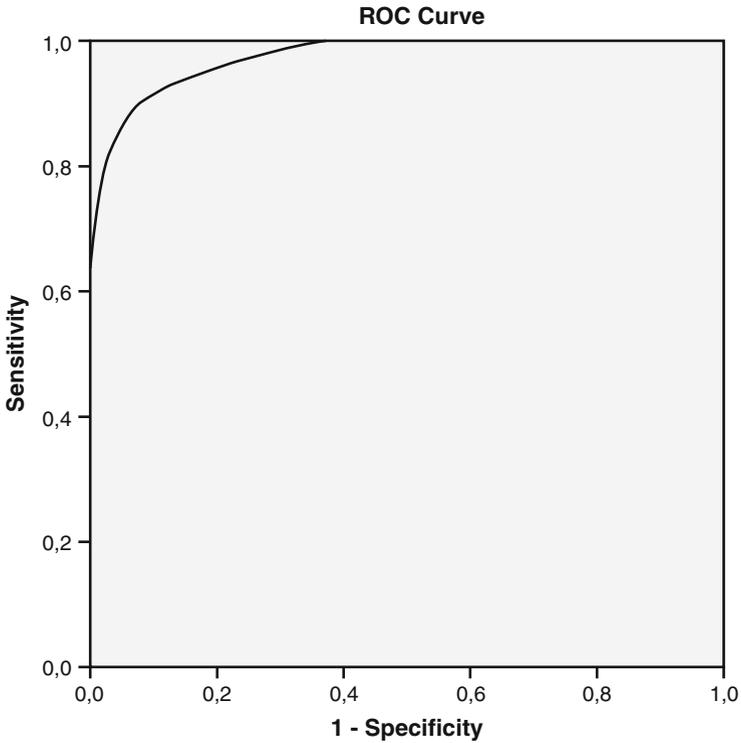
The test result variable(s): score has at least one tie between the positive actual state group and the negative actual state group

Statistics maybe biased

<sup>a</sup>Under the nonparametric assumption

<sup>b</sup>Null hypothesis: true area = 0.5

Subsequently the same procedure is followed for the second data file.



Diagonal segments are produced by ties.

Area under the curve  
 Test result variable(s): score

Area	Std. error <sup>a</sup>	Asymptotic sig. <sup>b</sup>	Asymptotic 95 % confidence interval	
			Lower bound	Upper bound
,974	,005	,000	,965	,983

The test result variable(s): score has at least one tie between the positive actual state group and the negative actual state group

Statistics maybe biased

<sup>a</sup>Under the nonparametric assumption

<sup>b</sup>Null hypothesis: true area = 0.5

The Area under curve of data file 2 is larger than that of data file 1. The test 2 seems to perform better. The z-test can again be used to test for significance of difference.

$$z = (0.974 - 0.945) / \sqrt{(0.009^2 + 0.005^2)} = 2.90$$
$$p < 0.01.$$

## 10 Conclusion

Both logistic regression with the presence of disease as outcome and test scores of as predictor and c-statistics can be used for comparing the performance of qualitative diagnostic tests. However, c-statistics may perform less well with very large areas under the curve, and it assesses relative risks while in practice absolute risk levels may be more important

## 11 Note

More background, theoretical and mathematical information of logistic regression and c-statistics is in *Machine learning in medicine part two*, Chap. 6, pp 45–52, Springer Heidelberg Germany, 2013, from the same authors.