The Stata Journal (2016)
16, Number 1, pp. 185–196

Features of the area under the receiver operating characteristic (ROC) curve. A good practice.

David Lora
Clinical Research Unit (imas12)
Hospital Universitario 12 de Octubre
and CIBER de Epidemiología y Salud Pública (CIBERESP)
Madrid, Spain
david@h12o.es

Israel Contador
Department of Basic Psychology, Psychobiology and Methodology of Behavioral Sciences
University of Salamanca
Salamanca, Spain

José F. Pérez-Regadera
Department of Radiation Oncology
Hospital Universitario 12 de Octubre
Madrid, Spain

Agustín Gómez de la Cámara
Clinical Research Unit (imas12)
Hospital Universitario 12 de Octubre
and CIBER de Epidemiología y Salud Pública (CIBERESP)
Madrid, Spain

Abstract. The area under the receiver operating characteristic (ROC) curve is a measure of discrimination ability used in diagnostic and prognostic research. The ROC plot is usually represented without additional information about decision thresholds used to generate the graph. In our article, we show that adding at least one or more informative cutoff points on the ROC graph facilitates the characterization of the test and the evaluation of the discriminatory capacities, which can result in more informed medical decisions. We use the `rocreg` and `rocregplot` commands.

Keywords: st0430, receiver operating characteristic (ROC) curve, area under the ROC curve, cervix cancer, diagnostic test, discrimination, prognostic models, rocreg, rocregplot

1 Introduction

The receiver operating characteristic (ROC) area represents the probability that in a specific diagnostic test or prognostic model, a randomly chosen diseased subject is ranked with greater suspicion than a randomly chosen nondiseased subject (Hanley and McNeil 1982). The area under the ROC curve is a measure of discrimination ability used in diagnostic tests and prognostic models. The discriminatory capacity corresponds to the area below the curve. The ROC area can be graphed by varying the cutoff points used to determine which values of the clinical procedure will be considered abnormal and then plotting the resulting true-positive rate (sensitivity) against the corresponding false-positive rate (1 − specificity) (DeLong, DeLong, and Clarke-Pearson 1988). However, decision thresholds are usually not displayed on the ROC plot, though they are known and used to generate the graph (Zweig and Campbell 1993). In this article, we show that the addition of cutoff point information enables one to 1) locate the decision threshold with respect to its sensitivity and specificity and 2) show its importance regarding the other thresholds. This information, which is useful in making medical decisions (Royston, Altman, and Sauerbrei 2006), facilitates the characterization of the test as well as the evaluation of the discriminatory capacity when using the ROC plot (Moons et al. 2015).

2 Example

To demonstrate the importance of the description of the area under the ROC curve, we generated a random dataset of 200 patients in the early stages of cervical cancer with three variables: the lymph node metastasis (LNM) status (40 out of 200) and the predicted probabilities of two preoperative prognostic models for the identification of LNM in early cervical cancer. The first model incorporated age, tumor size by magnetic resonance imaging (MRI), and LNM assessed by positron emission tomography and computed tomography variables; the second model incorporated age, tumor size by MRI, and LNM assessed by MRI and squamous cell carcinoma antigen. The sample was formed by comparing the patients' data with the information given in Kim et al. (2014).

```
. set obs 200
number of observations (_N) was 0, now 200
. set seed 1556
. generate LNM=1 if _n<=40
(160 missing values generated)
. generate predprob=0.01*runiform() if _n<=1
(199 missing values generated)
. replace predprob=0.01+(0.05-0.01)*runiform() if _n<=3  & _n>1
(2 real changes made)
. replace predprob=0.05+(0.25-0.05)*runiform() if _n<=10 & _n>3
(7 real changes made)
. replace predprob=0.25+(0.50-0.25)*runiform() if _n<=18 & _n>10
(8 real changes made)
. replace predprob=0.50+(0.75-0.50)*runiform() if _n<=28 & _n>18
(10 real changes made)
```

```
. replace predprob=0.75+(1.00-0.75)*runiform() if _n<=40 & _n>28
(12 real changes made)
. generate predprob2=0.15*runiform() if _n<=1
(199 missing values generated)
. replace predprob2=0.15+(0.25-0.15)*runiform() if _n<=3  & _n>1
(2 real changes made)
. replace predprob2=0.25+(0.45-0.25)*runiform() if _n<=10 & _n>3
(7 real changes made)
. replace predprob2=0.45+(0.65-0.45)*runiform() if _n<=18 & _n>10
(8 real changes made)
. replace predprob2=0.65+(0.85-0.65)*runiform() if _n<=28 & _n>18
(10 real changes made)
. replace predprob2=0.85+(1.00-0.85)*runiform() if _n<=40 & _n>28
(12 real changes made)
. replace LNM=0 if _n> 40
(160 real changes made)
. replace predprob=0.01*runiform() if _n<=86 & _n>40
(46 real changes made)
. replace predprob=0.01+(0.05-0.01)*runiform() if _n<=124 & _n>86
(38 real changes made)
. replace predprob=0.05+(0.25-0.05)*runiform() if _n<=161 & _n>124
(37 real changes made)
. replace predprob=0.25+(0.50-0.25)*runiform() if _n<=180 & _n>161
(19 real changes made)
. replace predprob=0.50+(0.75-0.50)*runiform() if _n<=195 & _n>180
(15 real changes made)
. replace predprob=0.75+(1.00-0.75)*runiform() if _n<=200 & _n>195
(5 real changes made)
. replace predprob2=0.15*runiform() if _n<=86 & _n>40
(46 real changes made)
. replace predprob2=0.15+(0.25-0.15)*runiform() if _n<=124 & _n>86
(38 real changes made)
. replace predprob2=0.25+(0.45-0.25)*runiform() if _n<=161 & _n>124
(37 real changes made)
. replace predprob2=0.45+(0.65-0.45)*runiform() if _n<=180 & _n>161
(19 real changes made)
. replace predprob2=0.65+(0.85-0.65)*runiform() if _n<=195 & _n>180
(15 real changes made)
. replace predprob2=0.85+(1.00-0.85)*runiform() if _n<=200 & _n>195
(5 real changes made)
```

Using the `rocreg` command, we estimate the area under the ROC curve for the classifier based on model 1 with 95% confidence intervals (CI) and store the true-positive rate (sensitivity) and false-positive rate (1 − specificity) values in variables for each classifier point, `_roc_predprob` and `_fpr_predprob`, respectively.

```
. rocreg LNM predprob, nodots
Bootstrap results                                 Number of obs      =        200
                                                  Replications       =      1,000

Nonparametric ROC estimation
Control standardization: empirical
ROC method             : empirical
Area under the ROC curve
   Status    : LNM
   Classifier: predprob
```

AUC	Observed Coef.	Bias	Bootstrap Std. Err.	[95% Conf.	Interval]	
	.8335938	.0012301	.0341697	.7666225	.900565	(N)
				.7586384	.8939733	(P)
				.7467949	.8849932	(BC)

We use the `rocregplot` command to draw the ROC curve for the first model (figure 1).

```
. rocregplot, plot1opts(msymbol(none) lcolor(black))
>         scheme(s1color) rlopts(lcolor(black))
>           ylabel(0(0.25)1, angle(0) format(%3.2f)) ytitle("Sensitivity")
>           xlabel(0(0.25)1, format(%3.2f)) xtitle("1 - Specificity")
>           legend( col(1) order(1) label(1 "ROC AUC= 0.83 [0.77;0.90]"))
```

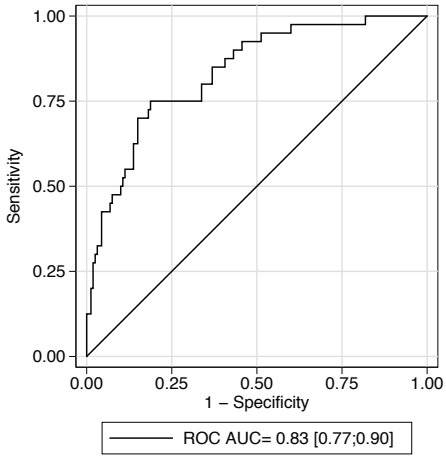

Figure 1. ROC curve for model 1

This curve is depicted by consecutive values from the test (that is, predicted probabilities from the prognostic model) to classify an individual as positive (LNM ill = value is above the threshold) or negative (free of LNM = values are low or under the selected threshold) in comparison with a gold-standard criterion. An increase of the cutoff point

would decrease sensitivity and increase specificity. We have added the model's discriminatory capacity—0.83 (95% CI [0.77; 0.90])—to the plot using the `legend()` option. We can also characterize and enhance the model by displaying cutoff point information on the plot. To add the predicted probabilities, we first list the observations with predicted probabilities close to 1%, 5%, and 10%. The observations with predicted probabilities closest to 1%, 5%, and 10% are indicated with an arrow.

```
. sort predprob
. list if (predprob > 0.009 & predprob < 0.013)
```

	LNM	predprob	predpr~2	_roc_p~b	_fpr_p~b	
45.	0	.0095514	.0902184	.975	.73125	
46.	0	.0095826	.0040212	.975	.725	
47.	0	.0096395	.1032082	.975	.71875	<--
48.	0	.0120601	.2063762	.975	.7125	
49.	0	.0128457	.184004	.975	.70625	

```
. list if (predprob > 0.049 & predprob < 0.06)
```

	LNM	predprob	predpr~2	_roc_p~b	_fpr_p~b	
85.	0	.0490097	.2001811	.925	.49375	
86.	0	.0498729	.228905	.925	.4875	
87.	0	.0499744	.2237535	.925	.48125	<--
88.	0	.0567622	.3069373	.925	.475	

```
. list if (predprob > 0.09 & predprob < 0.11)
```

	LNM	predprob	predpr~2	_roc_p~b	_fpr_p~b	
102.	0	.0936247	.3390107	.875	.40625	<--
103.	0	.1064184	.2700449	.85	.4	

Now, we use `rocregplot` to redraw the ROC curve. We add captions at the sensitivity and $1 -$ specificity of the indicated observations to the plot using the `text()` option (figure 2).

```
. rocregplot, plot1opts(msymbol(none) lcolor(black))
>        scheme(s1color) rlopts(lcolor(black))
>          ylabel(0(0.25)1, angle(0) format(%3.2f)) ytitle("Sensitivity")
>          xlabel(0(0.25)1, format(%3.2f)) xtitle("1 - Specificity")
>           text(0.99 0.69 "1%", size(small) placement(n))
>           text(0.94 0.44 "5%", size(small) placement(n))
>           text(0.89 0.37 "10%", size(small) placement(n))
>           text(0.98 0.72 "x", size(large) placement(c))
>           text(0.93 0.48 "x", size(large) placement(c))
>           text(0.88 0.41 "x", size(large) placement(c))
>          legend( col(1) order(1) label(1 "ROC AUC= 0.83 [0.77;0.90]"))
```

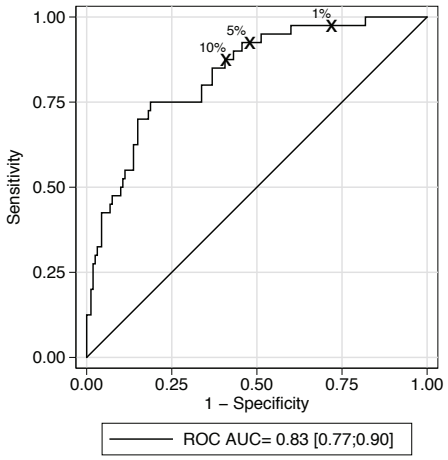

Figure 2. ROC curve for model 1 with predicted probabilities closest to 1%, 5%, and 10%

The result is that we show the predicted probabilities of developing LNM with the associated sensitivity and specificity values in the ROC plot. Further, we can describe the sensitivity and 95% CI with respect to its 1 − specificity using the `roc()` option (figure 3) and describe the false-positive rate and 95% CI for given true-positive values using the `invroc()` option in the `rocreg` command (figure 4):

```
. rocreg LNM predprob, nodots roc(0.71875 0.48125 0.40625)
Bootstrap results                                 Number of obs     =        200
                                                  Replications      =      1,000

Nonparametric ROC estimation

Control standardization: empirical
ROC method             : empirical

ROC curve

   Status    : LNM
   Classifier: predprob
```

ROC	Observed Coef.	Bias	Bootstrap Std. Err.	[95% Conf.	Interval]	
.71875	.975	-.0002608	.0249877	.9260249	1.023975	(N)
				.9166667	1	(P)
				.9090909	1	(BC)
.48125	.925	-.0044098	.0487009	.829548	1.020452	(N)
				.8101673	1	(P)
				.8	1	(BC)
.40625	.875	-.0159269	.0697107	.7383696	1.01163	(N)
				.7105263	.9725976	(P)
				.7333333	.9767442	(BC)

```
. rocregplot, btype(p) plot1opts(msymbol(none) lcolor(black))
>      scheme(s1color) rlopts(lcolor(black))
>      ylabel(0(0.25)1, angle(0) format(%3.2f)) ytitle("Sensitivity")
>      xlabel(0(0.25)1, format(%3.2f)) xtitle("1 - Specificity")
>      legend( col(1) order(1) label(1 "ROC AUC= 0.83 [0.77;0.90]"))
```

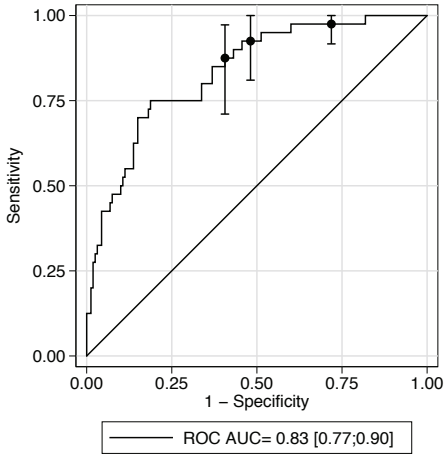

Figure 3. ROC curve for model 1 showing sensitivity and 95% CI with respect to its 1 − specificity

```
. rocreg LNM predprob, nodots invroc(0.875 0.925 0.975)
Bootstrap results                                 Number of obs     =        200
                                                  Replications      =      1,000

Nonparametric ROC estimation

Control standardization: empirical
ROC method             : empirical

False-positive rate

   Status    : LNM
   Classifier: predprob
```

invROC	Observed Coef.	Bias	Bootstrap Std. Err.	[95% Conf.	Interval]	
.875	.40625	.0025914	.0776321	.254094	.558406	(N)
				.1911568	.5642543	(P)
				.1875	.5616438	(BC)
.925	.45625	.0346563	.1047699	.2509048	.6615952	(N)
				.3393398	.8106061	(P)
				.3125	.64375	(BC)
.975	.6	.0644339	.1507743	.3044879	.8955121	(N)
				.4086708	.8700964	(P)
				.3806452	.8543046	(BC)

```
. rocregplot, plot1opts(msymbol(none) lcolor(black))
>      scheme(s1color) rlopts(lcolor(black))
>      ylabel(0(0.25)1, angle(0) format(%3.2f)) ytitle("Sensitivity")
>      xlabel(0(0.25)1, format(%3.2f)) xtitle("1 - Specificity")
>      legend( col(1) order(1) label(1 "ROC AUC= 0.83 [0.77;0.90]"))
```

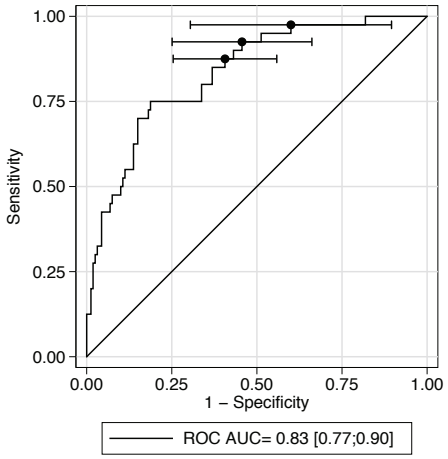

Figure 4. ROC curve for model 1 showing false-positive rate and 95% CI for given true-positive values

The risk thresholds of the prognostic model allow us to see the effects of selecting different cutoff points on medical decisions (Royston, Altman, and Sauerbrei 2006). Further, we can display the ROC curves of different classifiers in the same figure, but we should remember that even when two classifiers agree in sensitivity and specificity, they will not necessarily have the same cutoff point values. Now, we show an example with the classifiers for the first and second models. We use `rocreg` to calculate the sensitivity and $1-$specificity values for both models and then list the predicted probabilities where the sensitivity and $1-$specificity values agree.

```
. rocreg LNM predprob predprob2, nodots
Bootstrap results                                 Number of obs     =        200
                                                  Replications      =      1,000

Nonparametric ROC estimation

Control standardization: empirical
ROC method             : empirical

Area under the ROC curve

   Status    : LNM
   Classifier: predprob
```

AUC	Observed Coef.	Bias	Bootstrap Std. Err.	[95% Conf.	Interval]
	.8335938	-.0010747	.0368254	.7614172	.9057703 (N)
				.7531488	.8963632 (P)
				.7459636	.8940851 (BC)

```
   Status    : LNM
   Classifier: predprob2
```

AUC	Observed Coef.	Bias	Bootstrap Std. Err.	[95% Conf.	Interval]
	.8339062	-.0006906	.0357851	.7637688	.9040437 (N)
				.7543176	.8938994 (P)
				.752333	.891875 (BC)

```
Ho: All classifiers have equal AUC values.
Ha: At least one classifier has a different AUC value.

P-value:     .9741677                  Test based on bootstrap (N) assumptions.

. list if _roc_predprob==_roc_predprob2 & _fpr_predprob==_fpr_predprob2
```

	LNM	predprob	predpr~2	_roc_p~b	_fpr_p~b	_roc_p~2	_fpr_p~2
12.	0	.002632	.0291194	1	.93125	1	.93125
133.	0	.2548581	.4659569	.75	.2375	.75	.2375

```
. rocreg LNM predprob predprob2, nodots roc(0.2375)
Bootstrap results                                 Number of obs     =        200
                                                  Replications      =      1,000

Nonparametric ROC estimation

Control standardization: empirical
ROC method             : empirical

ROC curve

   Status    : LNM
   Classifier: predprob
```

ROC	Observed Coef.	Bias	Bootstrap Std. Err.	[95% Conf.	Interval]
.2375	.75	-.0050337	.0736913	.6055677	.8944323 (N)
				.6	.8787879 (P)
				.6136364	.8857143 (BC)

```
Status    : LNM
Classifier: predprob2
```

ROC	Observed Coef.	Bias	Bootstrap Std. Err.	[95% Conf.	Interval]	
.2375	.75	-.0151284	.087794	.5779269	.9220731	(N)
				.5357346	.8837209	(P)
				.5833333	.90625	(BC)

```
Ho: All classifiers have equal ROC values.
Ha: At least one classifier has a different ROC value.
Test based on bootstrap (N) assumptions.
```

ROC	P-value
.2375	1

At a sensitivity of 0.75 and 1−specificity of approximately 0.25, the classifiers agree. We use `rocregplot` to plot the ROC curves of both classifiers and caption the differing cutoff point values (figure 5).

```
. rocregplot, scheme(s1color) rlopts(lcolor(black))
>         plot1opts(msymbol(none) lcolor(black))
>         plot2opts(msymbol(none) lcolor(black))
>         ylabel(0(0.25)1, angle(0) format(%3.2f)) ytitle("Sensitivity")
>         xlabel(0(0.25)1, format(%3.2f)) xtitle("1 - Specificity")
>           text(0.75 0.19 "25%", size(small) placement(n) color("dknavy"))
>           text(0.70 0.29 "47%", size(small) placement(n) color("maroon"))
>           legend( col(1) order(1 2)
>            label(1 "ROC AUC - Model 1= 0.83 [0.77;0.90]")
>            label(2 "ROC AUC - Model 2= 0.83 [0.77;0.90]"))
```

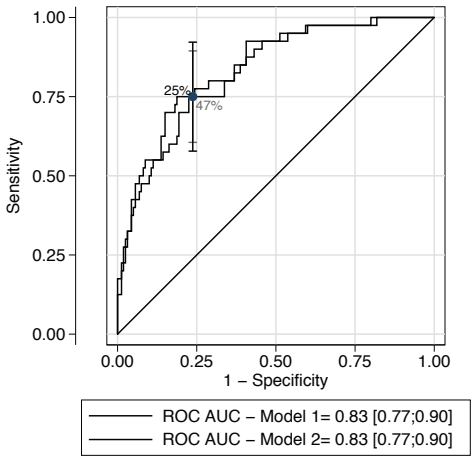

Figure 5. ROC curves for models 1 and 2

In this example, the layout of the thresholds allows us to define what predicted probability value is appropriate to define patients as "low risk" and "nonlow risk" of developing LNM. This will help us to evaluate, for example, the therapeutic value of lymphadenectomy in "nonlow-risk" patients in clinical trials, as in the role of sentinel lymph node biopsy and the prognostic value of metastatic nodal resection in this group and not in the entire population (Kim et al. 2014). That is, it will help us to identify those persons who may benefit from being included in a clinical trial, which lets us better understand disease mechanisms and select the best treatment options.

3 Conclusion

Adding at least one or more informative cutoff points on the ROC graph or textual description of a prognostic model enables one to better characterize the test as well as evaluate the strategic discriminatory capacities. This good practice will help improve our knowledge about the diagnostic and prognostic uses of clinical procedures.

4 References

DeLong, E. R., D. M. DeLong, and D. L. Clarke-Pearson. 1988. Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics* 44: 837–845.

Hanley, J. A., and B. J. McNeil. 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143: 29–36.

Kim, D.-Y., S.-H. Shim, S.-O. Kim, S.-W. Lee, J.-Y. Park, D.-S. Suh, J.-H. Kim, Y.-M. Kim, Y.-T. Kim, and J.-H. Nam. 2014. Preoperative nomogram for the identification of lymph node metastasis in early cervical cancer. *British Journal of Cancer* 110: 34–41.

Moons, K. G. M., D. G. Altman, J. B. Reitsma, J. P. A. Ioannidis, P. Macaskill, E. W. Steyerberg, A. J. Vickers, D. F. Ransohoff, and G. S. Collins. 2015. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): Explanation and elaboration. *Annals of Internal Medicine* 162: W1–W73.

Royston, P., D. G. Altman, and W. Sauerbrei. 2006. Dichotomizing continuous predictors in multiple regression: A bad idea. *Statistics in Medicine* 25: 127–141.

Zweig, M. H., and G. Campbell. 1993. Receiver-operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine. *Clinical Chemistry* 39: 561–577.

About the authors

David Lora is a biostatistician at the Department of Clinical Research, Clinical Trials, Clinical Epidemiology and Bioinformatics, Hospital Universitario 12 de Octubre, Madrid, Spain. He is also a researcher at the Spanish Epidemiology and Public Health Consortium. His research interests lie in the intersection of computer science and statistical methods with applications in the areas of prognosis and cancer research.

Israel Contador, PhD, is a professor of psychology at the Department of Basic Psychology, Psychobiology and Methodology of Behavioral Sciences, University of Salamanca, Spain. He is particularly interested in clinical neuropsychology, and his research covers topics such as the prognosis of mild cognitive impairment and the early detection of dementia.

José F. Pérez-Regadera, PhD, is the head of the Department of Radiation Oncology, Hospital Universitario 12 de Octubre, Madrid, Spain. He is an associate professor in the Department of Radiation Oncology, Complutense University, Madrid, Spain. His research focuses primarily on gynaecological cancer, neuro-oncology, and brachytherapy.

Agustín Gómez de la Cámara, PhD, is a senior investigator at and the head of the Department of Clinical Research, Clinical Trials, Clinical Epidemiology and Bioinformatics, Hospital Universitario 12 de Octubre, Madrid, Spain. He is a professor at the Department of Medicine at Madrid University. He is a member of the executive group of the Spanish Clinical Research Network and the European Clinical Research Network and a senior researcher at the Spanish Epidemiology and Public Health Consortium. His research focuses on the efficacy of medical and public health interventions and diagnostic and prognostic models in chronic diseases.