

Receiver Operating Characteristic (ROC) Curve Analysis: A Tutorial Using Body Mass Index (BMI) as a Measure of Obesity

Peter D. Hart*

Health Promotion Program, Montana State University - Northern, & Health Demographics, Havre, MT 59501, USA

*Corresponding author: peter.hart@msun.edu

Abstract The purpose of this tutorial-based lecture is to show the usefulness of performing a receiver operating characteristic (ROC) curve analysis. The lecture will explain the background terminology associated with a ROC curve analysis, show the necessary SAS software coding to run a ROC curve analysis, and finally help interpret the analysis output in order to make informed research decisions. An example evaluating body mass index (BMI) cutoff scores indicating obesity is used in the tutorial.

Keywords: Receiver operating characteristic (ROC), sensitivity, specificity, Body Mass Index (BMI), obesity

Cite This Article: Peter D. Hart, "Receiver Operating Characteristic (ROC) Curve Analysis: A Tutorial Using Body Mass Index (BMI) as a Measure of Obesity." *Journal of Physical Activity Research*, vol. 1, no. 1 (2016): 5-8. doi: 10.12691/jpar-1-1-2.

1. Introduction

In physical activity research, it is often of interest to dichotomize a continuous variable to subsequently use as either an outcome or predictor in an analysis. In clinical situations, these variables may be part of an evaluation process and lead to specific programming or prescription. All too often, the cutoff value for creating such a variable is arbitrary or based on outdated standards. There does exist, however, a diagnostic statistical procedure to help evaluate these cutoff values and in some instances provide justification for changing them. The receiver operating characteristic (ROC) curve is a procedure that can produce both tabular and graphical output to aid in the assessment of a cutoff value used to create a dichotomous variable. This brief tutorial is the first in a series of lectures on applied statistics and measurement and will rely on the free SAS University Edition online statistical program [1].

2. Terminology

Before beginning this tutorial, it is important to first define some terminology associated with a ROC curve analysis. When evaluating the effectiveness of a dichotomous variable, the question usually posed is how accurate is the measure in classifying observations? If, for example, a dichotomous measure is used to diagnose a disease, then the question becomes how accurate is it in classifying those who have the disease and those who do not? This attribute of a measure can also be considered its ability to discriminate. A 2x2 table helps see the possible outcomes to this question.

Table 1. Possible outcomes from using a new measure as compared to a true outcome

Measure	True Outcome	
	Positive	Negative
Positive	TP (a)	FP (b)
Negative	FN (c)	TN (d)

Note. TP is true positive, FP is false positive, FN is false negative, and TN is true negative. Sensitivity is $a / (a + c)$. Specificity is $d / (b + d)$.

Two cells in Table 1 are emphasized in a ROC curve analysis. A *true positive* (TP) is the number of observations that our measure identifies as positive and actually are positive. Alternatively, a *true negative* (TN) is the number of observations our measure identifies as negative and actually are negative. The values of TP and TN lead us to two more descriptive measures. *Sensitivity* refers to the proportion (probability) of positive observations that our measure accurately identifies as positive (i.e., TP rate). Sensitivity can easily be calculated from our cells: $\text{sensitivity} = a / (a + c)$. *Specificity* refers to the proportion (probability) of negative observations that our measure accurately identifies as negative (i.e., TN rate). Specificity can also be calculated from our cells: $\text{specificity} = d / (b + d)$.

Ideally, a measure would possess both high sensitivity and high specificity. However, in reality, these two values often have an inverse relationship [2]. Specific situations will dictate which is more significant. For example, it may be more important to rule out a disease that a researcher may sacrifice overestimating the disease. In other words, in this case, our measure allows for more positive outcomes. And although more positive outcomes can increase TP outcomes they can also increase FP outcomes. This would imply the measure has high sensitivity but low specificity. A negative test result here gives us greater

confidence that the disease is not present. A different situation may require more emphasis on ruling in a disease that the researcher sacrifices underestimating the disease. Here, our measure would yield more negative outcomes with increased TN outcomes but also increased FN outcomes. This would imply the measure has high specificity but low sensitivity. Alternatively, a positive test result here gives us greater confidence that the disease is present.

The ultimate decision in research which leads to these opposing values of sensitivity and specificity is the cutoff value used to create the dichotomous variable. The cutoff value is what converts a continuous measure to a dichotomous measure which in turn is used to classify observations. A ROC curve analysis produces a statistical report for the researcher to evaluate and then use to select or modify a cutoff score [3]. The following example will illustrate a ROC curve analysis.

3. Example

Body mass index (BMI) is a common and useful measure of body composition in physical activity research [4]. Despite a wide agreed upon cut point for obesity ($\text{BMI} \geq 30$), it may be beneficial to ensure our classification is accurate in the population we are investigating. The data analyzed in this tutorial include BMI measurements on 252 males [5]. The dataset also contains an obesity status variable (OBESE) derived from percent body fat (PBF). The obesity status variable serves as a gold-standard measure, which is needed for any ROC curve analysis. The OBESE variable is coded '1' for obese males and '0' for non-obese males. A PBF of 25 and greater indicated obesity [6]. Table 2 shows the SAS University Edition (SAS Studio) code required to produce a list of the dataset (Table 3). The OBS=5 option is used to print only the first five observations.

Table 2. SAS code for printing the initial dataset

```
proc print data=roc1 (obs=5);
  var age height weight bmi pbf obese;
run;
```

Note. ROC1 is the SAS dataset containing the prepared data. The OBS=5 option limits the output to the first five observations, for space concern only. The VAR statement specifies which variables to print.

Table 3. Partial list output of the dataset

Obs	age	height	weight	bmi	pbf	obese
1	23	67.75	154.25	23.7	12.3	0
2	22	72.25	173.25	23.4	6.1	0
3	22	66.25	154.00	24.7	25.3	1
4	26	72.25	184.75	24.9	10.4	0
5	24	71.25	184.25	25.6	28.7	1

Note. The above list output is from the PROC PRINT procedure. The only variables needed for a ROC curve analysis are BMI and OBESE.

A ROC curve analysis can easily be performed using the PROC LOGISTIC procedure, as seen in Table 4 [7]. The MODEL statement declares the OBESE variable as the dependent variable (left side of equal sign). SAS will model the probability of $\text{OBESE}=0$ by default. Therefore an EVENT='1' option is necessary since obesity is defined as '1' in the dataset. BMI is the continuous

independent variable (right side of equal sign) from which the cutoff score will be established to indicate obesity. The OUTROC= option creates a second dataset containing sensitivity and specificity data which here is called ROCDATA. This dataset will be referred to later in the tutorial. The ROC statement produces a ROC curve and the ROCCONTRAST statement produces a significance test for the ROC curve.

Table 4. SAS code for the ROC curve analysis

```
proc logistic data=roc1;
  model obese(event='1')=bmi/outroc=rocd;
  roc; roccontrast;
run;
```

Note. OBESE is the dependent variable. The EVENT='1' informs SAS that obesity is coded '1'. BMI is the independent variable. OUTROC= creates a second dataset. ROC produces a ROC curve. ROCCONTRAST produces a significance test.

Table 5 shows partial output from the PROC LOGISTIC procedure. This part of the output is not a main concern of a ROC analysis. However, the regression coefficients will be used later in the tutorial and the significance of BMI ($p < .0001$) is necessary for a useful ROC curve.

Table 5. Partial logistic regression output

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-15.7009	2.0395	59.2650	<.0001
bmi	1	0.5584	0.0761	53.8483	<.0001

Note. The model's intercept (-15.7009) and slope (0.5584) are needed to compute the BMI cutoff scores.

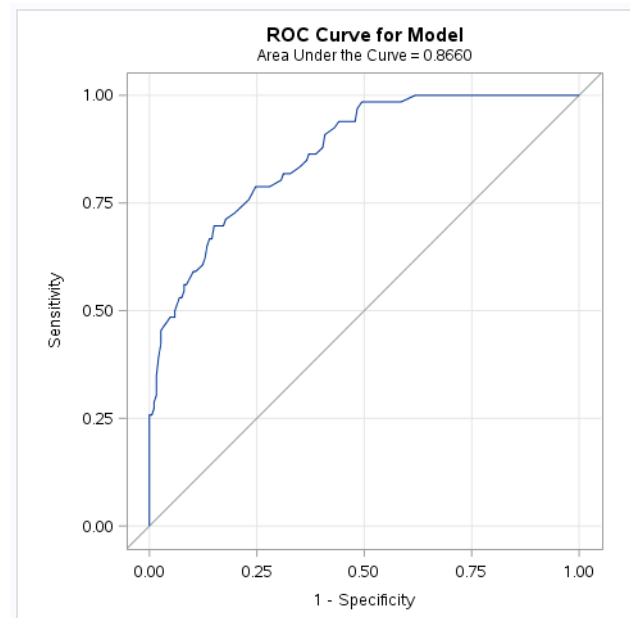


Figure 1. ROC curve and area under the curve (AUC)

Note. The farther the curve is to the upper left corner of the graph, the better BMI is at classifying obesity status.

Figure 1 displays the ROC curve from our analysis. The graph is constructed with sensitivity on the vertical axis and 1-specificity on the horizontal. This should indicate that the area in the upper left region provides the most useful discrimination in terms of a cutoff score [8]. The diagonal line, from (0,0) to (1,1), is indicative of an independent variable that discriminates no different from

guessing (50/50 chance). However, our ROC curve is well above this line. In fact, the area under our curve (AUC) is 0.866 as compared to that of the diagonal line which is always 0.500 (half the graph). The AUC of 0.866 more specifically means that of all possible obese/non-obese male pairs created from our sample, the model with BMI assigned (in 86.6% of the pairs) a higher predicted probability of being obese to the obese male than it did to the non-obese male. A little simpler, BMI correctly

classifies randomly drawn pairs of obese/non-obese men 86.6% of the time. Table 6 displays significance tests for our AUC. Since the 95% confidence interval (0.819, 0.913) does not contain 0.500, we can conclude that our AUC is significantly better than chance. A Chi-square test is also in the output and provides a p -value ($p < .0001$) associated with the null hypothesis that our AUC equals 0.500. This test is also significant.

Table 6. ROC curve AUC significance tests

ROC Association Statistics						
ROC Model	Mann-Whitney			Somers' D	Gamma	Tau-a
	Area	Standard Error	95% Wald Confidence Limits			
Model	0.8660	0.0240	0.8190 0.9131	0.7321	0.7369	0.2842
ROC1	0.5000	0	0.5000 0.5000	0	.	0

ROC Contrast Test Results			
Contrast	DF	Chi-Square	Pr > ChiSq
Reference = Model	1	232.2239	<.0001

Note. If 0.500 lies outside the 95% confidence interval, we can conclude that our ROC curve is significant.

Now that we know our AUC is significant, we can determine an optimal BMI cutoff value for obesity. In theory, this can be accomplished by inspecting points on our curve that fall closest to the upper left corner of the graph. In reality, since these points represent probabilities and not BMI values, we must first make a few calculations.

Table 7. SAS code for ROC curve analysis variables

```
data roc2; set rocdata;
  logit=log(_prob_/(1-_prob_));
  Cutoff=(logit+15.7009)/0.5584;
  Sensitivity=_sensit_;
  Specificity=1-_lmspec_;
  YJ=Sensitivity+Specificity-1;
run;
```

Note. The cutoff variable is formed by re-arranging any logistic regression model to solve for X. The model is: $\text{logit} = \text{intercept} + \text{slope}(X)$.

Table 7 shows the creation of a new SAS dataset (ROC2) from the dataset that was created with the OUTROC= option earlier. A PROC PRINT of this dataset (PROC PRINT DATA=ROCDATA;) may help understand the variable names that SAS assigned. But is omitted from this tutorial. The second line creates a new variable called LOGIT. A logit is the log of the odds in favor of an event. The LOG() function will compute the log while the odds by definition is the probability of success (_prob_) over the probability of failure (1-_prob_). The predicted probability (_prob_) variable was in the ROCDATA dataset and will now also be in the ROC2 dataset. The third line creates the CUTOFF variable which contains different BMI cut points indicating obesity. The 15.7009 and 0.5584 are the intercept and slope (respectively) from our logistic regression analysis. This line of code comes from the mathematical rearrangement of the regression equation: $\text{logit} = -15.7009 + 0.5584(\text{BMI})$. The idea is that we want to solve for BMI for each probability cut point in the ROCDATA dataset by using our regression equation. The fourth line just gives sensitivity a better working name by creating a duplicate

variable. The fifth line yields specificity in its normal form. And the sixth line computes *Youden's J* (YJ). YJ is the sum of sensitivity and specificity minus 1 and is often used as criteria for selecting an optimal cutoff value [9].

Table 8 shows the code needed to print the new variables. It first may be an advantage to sort the new ROC2 dataset by YJ. Using the DESCENDING option will list the output in descending order of YJ. This is useful because larger values of YJ are optimal. Only the first twenty-five observations are printed.

Table 8. SAS code for printing analysis variables

```
proc sort data=roc2;
  by descending YJ;
run;

proc print data=roc2 (obs=25) noobs;
  var Cutoff Sensitivity Specificity YJ;
run;
```

Note. Sorting the new dataset by YJ can help identify cutoff scores more easily.

If the only criteria used to select a BMI cutoff score is YJ, then we would choose the very first row of the output which has the highest YJ statistic of 0.54643. Since YJ represents the TP rate minus the FP rate, larger values indicate a more accurate measure. The associated BMI cutoff value of 26.7996 is then the optimal cutoff in classifying individuals into obese and non-obese categories. However, YJ is often used as criteria when the researcher has no preference over sensitivity or specificity. If the researcher does have such a preference, each sensitivity and specificity combination in the output should be considered. As mentioned earlier, some researchers may suggest that it is more important to maximize the TP outcomes and others may prefer to maximize the TN outcomes. But whatever your position, this tutorial provides the methods for making such decisions. It is clear from this analysis, however, that when considering both sensitivity and specificity our traditional BMI cutoff of 30 (indicating obesity) may be too high.

Table 9. Partial output of BMI cutoff scores with statistics

Cutoff	Sensitivity	Specificity	YJ
26.7996	0.69697	0.84946	0.54643
25.9996	0.78788	0.75269	0.54057
26.6996	0.69697	0.83871	0.53568
26.3996	0.71212	0.82258	0.53470
26.5996	0.69697	0.83333	0.53030
26.2996	0.72727	0.80108	0.52835
26.1996	0.74242	0.78495	0.52737
26.9996	0.66667	0.86022	0.52688
26.0996	0.75758	0.76882	0.52639
26.4996	0.69697	0.82796	0.52493
25.8996	0.78788	0.73656	0.52444
26.8996	0.66667	0.85484	0.52151
27.1996	0.65152	0.86559	0.51711
25.7996	0.78788	0.72043	0.50831
25.4996	0.81818	0.68817	0.50635
24.6997	0.90909	0.59140	0.50049
24.4997	0.93939	0.55914	0.49853
25.5996	0.80303	0.69355	0.49658
24.5997	0.92424	0.56989	0.49413
25.0996	0.86364	0.62903	0.49267
27.2996	0.62121	0.87097	0.49218
25.3996	0.81818	0.67204	0.49022
23.9997	0.98485	0.50538	0.49022
27.5996	0.59091	0.89785	0.48876
24.9997	0.86364	0.62366	0.48729

Note. A large YJ may be one criteria for deciding an appropriate cutoff score.

A final situation that may be useful to the physical activity researcher is the comparison of two or more ROC curves. For example, let's say a second independent variable such as waist circumference (WC) was also available in our dataset and likewise provided a significant AUC in classifying obese men. A follow-up question would be, is WC a significantly better measure than BMI? This can easily be evaluated in SAS by including separate ROC statements for each independent variable being tested [10]. Table 10 shows this SAS software code. The output in this situation (not shown) will include a separate ROC curve analysis for each variable, a combined ROC

curve graph showing curve comparisons, as well as significance tests comparing AUC pairs.

Table 10. SAS code for the comparison of two ROC curves

```
proc logistic data=roc1;
  model obese(event='1')=bmi wc/outroc=rocdata;
  roc 'WC' wc; roc 'BMI' bmi;
  roccontrast reference('BMI')/estimate;
run;
```

Note. BMI and WC are both independent variables in the model statement. The 'WC' and 'BMI' options give the graphs names. Including the REFERENCE('BMI')/ESTIMATE to the ROCCONTRAST statement sets BMI as a reference to compare WC to in the significance test.

In summary, this tutorial-based lecture provides rationale, terminology, and evaluation criteria for using a ROC curve analysis in judging cutoff scores that create dichotomous variables. As well, ROC curve analysis code for the free SAS University Edition online statistical program was provided. This paper marks the first of a series of lectures on applied statistics and measurement presented by the journal. Stay tuned for our next lecture!

References

- [1] SAS Institute Inc. 2015. SAS® University Edition: Installation Guide for Windows. Cary, NC: SAS Institute Inc.
- [2] Bewick, Viv, Liz Cheek, and Jonathan Ball. "Statistics review 13: receiver operating characteristic curves." *Critical care* 8.6 (2004): 508.
- [3] Wood, Terry M., and Weimo Zhu. *Measurement theory and practice in kinesiology*. Human Kinetics, 2006.
- [4] Hart, Peter. D. Body mass index (BMI) and sedentary time in adults. *Obesity and Control Open Access*. 2 (2015): 106.
- [5] Johnson, Roger. Fitting percentage of body fat to simple body measurements", *Journal of Statistics Education*, 4, 1 (1996): 265.
- [6] American Council on Exercise (ACE). What is my ideal body fat percentage? Retrieved on Sept 15, 2015: <http://www.acefitness.org/>.
- [7] SAS Institute Inc. 2014. SAS/STAT® 13.2 User's Guide. The Logistic Procedure. Cary, NC: SAS Institute Inc.
- [8] Streiner, David L., and John Cairney. "What's under the ROC? An introduction to receiver operating characteristics curves." *Canadian Journal of Psychiatry* 52.2 (2007): 121.
- [9] Bangdiwala, Shrikant I., et al. "The agreement chart as an alternative to the receiver-operating characteristic curve for diagnostic tests." *Journal of Clinical Epidemiology* 61.9 (2008): 866-874.
- [10] Downer, Robert G., and P. J. Richardson. "Illustrative Logistic Regression Examples using PROC LOGISTIC: New Features in SAS/STAT® 9.2." *Paper SP03-2009* (2009).